The Epidemiology of Scabies

Kylie Ainslie, Mariette Hooiveld, Jacco Wallinga

2024-09-11

# Introduction

Scabies is classified as a neglected tropical disease caused by infestation of the skin with a microscopic mite (Sarcoptes scabiei)1. Symptoms are characterized by itchiness and rash at the site of infestation. Scabies affects around 400 million people per year, and accounts for a large proportion of skin disease in many low- and middle-income countries1. A rise in scabies cases has been observed through out Europe in recent years2–7.

Despite the considerable burden scabies poses annually, little is known about the disease dynamics of scabies transmission, such as the generation time (time between infection of an index case and a secondary case), serial interval (the time from onset of symptoms in an index case to the time of symptom onset in a secondary case), growth rate, and reproduction number (the average number of secondary cases resulting from one index case). A series of studies in the 1940s by Kenneth Mellanby and colleagues8,9 form much of the basis of our current understanding of scabies transmission. Some important results from the experiments performed in the 1940s by Mellanby and colleagues include an estimation of the incubation period, or time to symptom onset, of 4-6 weeks; the change in parasite rate, defined as the number of mature female mites, over time, which was shown to increase rapidly, peak around 100 days, and then decline quickly; and the probability of transmission via a secondary object (e.g., bed sheets) is low and most transmission occurs from person-to-person contact8.

In this work, we aim to estimate epidemiological characteristics of scabies. It is critical to better understand the underlying disease dynamics of scabies to 1) assess current spread and 2) inform infection control policy. Some modelling work has been performed to study the potential impacts of intervention strategies, such as mass drug administration, on scabies transmission10–12; however, the values used to parameterise these models are based largely on Mellanby’s 1944 study and use information about the mite life-cycle to approximate quantities such as latent period and infectious period. However, this information doesn’t necessarily provide good estimates for transmission potential in humans. For example, Kinyanjui et al.10 assume a latent period of 7-14 days to allow for the time for the time it takes for a fertilised female mite to reproduce and incorporate it into a susceptible-exposed-infectious model of scabies transmission. However, this assumes an equal chance of infectiousness despite the number of mites inhabiting an infested individual. Mellanby8 shows that the probability of onward transmission of scabies is rare when the number of mites on an individual is low and increases as the mite population grows. A better approximation of the disease process can be obtained using the generation time or serial interval; however, no such estimates exist for scabies.

Additionally, the growth rate and reproduction number are quantities that describe how fast an infectious disease spreads. They are also used to determine appropriate control measures to reduce the number of secondary infections so that an epidemic dies out. However, in the case of scabies, the basic reproduction number, the average number of secondary infections resulting from one index case in a completely susceptible population, has never been described. This makes it immensely difficult to determine what level of control measures are needed to contain disease spread. As evidence continues to suggest that scabies is a growing problem in European countries and remains a large problem elsewhere, it is important to describe the epidemiological characteristics of scabies so that the required control efforts to curb scabies epidemics can be determined.

In this study, we use epidemic curves of scabies outbreaks from the literature to estimate the mean serial interval of scabies. We use also data on weekly scabies cases in the Netherlands to estimate time-varying reproduction number and annual growth rate. To our knowledge, this is the first study to estimate these quantities for scabies.

# Methods

## Data Sources

We used publicly available data on reported symptom onset dates of scabies cases from previously published studies of scabies outbreaks to estimate the serial interval of scabies. When the original study data was not provided, we reconstructed the data from the published epidemic curves. Additionally, we used data on the incidence of scabies over time in the Netherlands to estimate reproduction number and growth rate. These data have been described elsewhere2,13. Briefly, we obtained weekly incidence of scabies (per 1000 people) from 2011 to 2023 as diagnosed by general practitioners (GPs) that are stored in a nationally representative primary care database of GPs hosted by the Netherlands Institute for Health Services Research (Nivel)13. Scabies is not a notifiable disease in the Netherlands, and thus information on scabies cases can only be obtained from GP diagnoses. Individuals in institutions (e.g., care homes) are generally not taken into account in these registrations.

## Serial Interval

Using the dates of symptom onset for scabies cases from four different studies of scabies outbreaks (Table @ref(tab:si\_results\_tab)), we estimated the mean and standard deviation of the serial interval distribution using the method proposed by Vink et al.14. The method involves calculating the index case-to-case (ICC) interval for each person, where the person with the greatest value for number of days since symptom onset will be considered the index case. The rest of the individuals will have an ICC interval calculated as the number of days between their symptom onset and the index case. We assumed a Normal serial interval distribution. We performed a sensitivity analysis in which a Gamma distribution was assumed as the serial interval distribution (see (appendix)).

We also performed a Bayesian meta-analysis using the brms package in R15 to estimate the pooled mean serial interval. We used a Bayesian hierarchical random-effects model. We chose to use a random effects model because we assume that each study has its own “true” serial interval, which has been sampled from an overarching distribution of true serial intervals16. We specified a prior distribution N(100,50) for the true pooled mean and Cauchy(0,1) for the between-study heterogeneity. We performed sensitivity analyses on our choices of prior distributions (see \ref(appendix)).

## Growth rate

We estimated the annual growth rate of scabies cases by fitting an exponential growth model to annual incidence of scabies per 1000 people from 2011 to 2023 in the Netherlands17. We assumed normally distributed errors, with mean zero and constant variance. Using the fitted exponential growth model and the estimated growth rate, we then determined the projected incidence of scabies per 1000 people until 2033 if no interventions are implemented and the growth rate remains constant. We used bootstrapping with 1000 samples to obtain 95% confidence bounds for the projected incidence of scabies per 1000 people.

## Time-varying Reproduction Number

We obtained weekly reported cases of scabies per 1000 people from 2011 to 2023 in the Netherlands2,17. To estimate time-varying reproduction number, we first randomly assigned each reported case a date of symptom onset in the week in which the case was reported. Since scabies is very hard to diagnose prior to symptom onset9, and reported cases are captured as part of sentinel surveillance based on GP diagnosis17, we assume reporting date is consistent with date of symptom onset. Using the constructed daily time series of date of symptom onset, we applied the method proposed by Wallinga and Teunis18 to estimate the time-varying case reproduction number. The case reproduction number is defined as the average number of new infections that an individual who becomes infected, or symptomatic, at a particular time point will go on to cause19, and is useful in retrospective analyses such as the one presented here. The method of Wallinga and Teunis estimates the time-varying case reproduction number by determining the likelihood of an event occurring for every pair of time points20. The method requires the specification of the serial interval distribution. We assumed a Normal serial interval distribution with mean 123.24 days and standard deviation 31.55 days, as estimated previously. Due to unobserved onward cases at the end of the time series, we adjusted for right truncation21,22. To obtain 95% confidence intervals on the daily reproduction number, we generated 100 bootstrapped samples by resampling each case with replacement and reconstructing the incidence time series for each sample. Reproduction number estimates were then smoothed using a rolling average of 5 weeks.

[Do we need to do this?] We performed a sensitivity analysis in which we assumed the serial interval was Gamma distributed.

All analyses were performed in R 4.4.023. Additional R packages used in this work are cited in the Appendix.

# Results

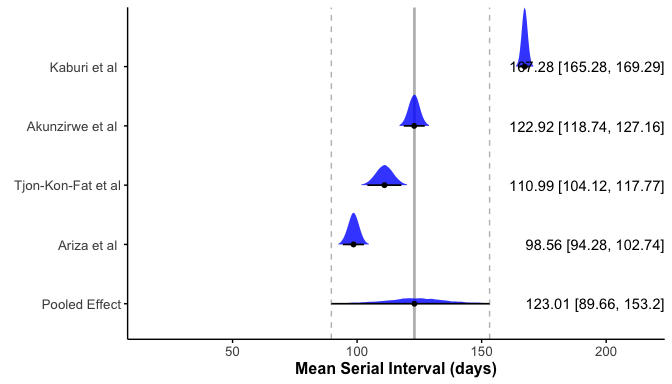
## Serial Interval

We estimated the mean and standard deviation of the serial interval using the epidemic curves from each study (Table @ref(tab:si\_results\_tab)). There was considerable variation between studies with the largest estimated mean serial interval of 167.34 days (SD = 9.72 days) from data from Kaburi et al.24 which describes an outbreak in a preschool in Ghana. The smallest estimated mean serial interval was 16.11 days (SD = 2.42 days), and was estimated from data from Larrosa et al.25 which describes an outbreak in a hospital in Spain. We performed a sensitivity analysis in which we estimated the mean and standard deviation of the serial interval assuming an underlying Gamma distribution (Table @ref(tab:tab\_si\_gam) and Figure @ref(fig:gam\_si\_plots)). We found that we obtained similar mean estimates as compared to assuming an underlying Normal distribution (albeit lower for most studies), but the estimated standard deviations were large. Model fit was assessed visually by plotting the estimated densities over the epidemic curves (Figure @ref(fig:gam\_si\_plot)) and we found that the Gamma distribution was not a good fit of the data.

Estimated mean and standard deviation (SD) of serial interval distribution, in days, for each study.

| **Study** | **Mean** | **Standard Deviation** |
| --- | --- | --- |
| Akunzirwe et al. | 122.9239 | 26.920354 |
| Ariza et al. | 98.4000 | 8.542332 |
| Kaburi et al. | 167.3444 | 9.717630 |
| Tjon-Kon-Fat et al | 110.7157 | 16.138793 |

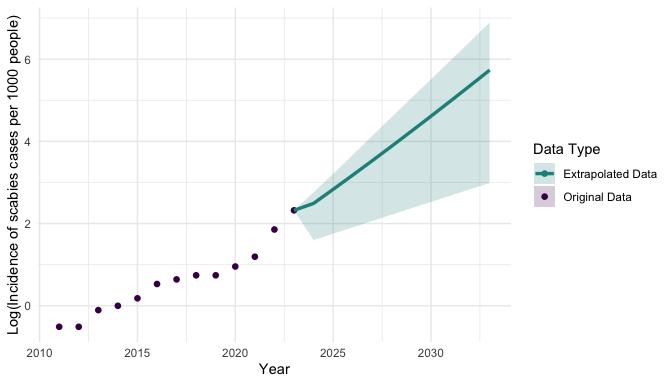
To obtain a pooled estimate of mean serial interval, we performed a Bayesian meta-analysis with random intercepts for each study. We excluded studies with study periods less than 100 days as it was unlikely to observe a primary-secondary infection in that time. We estimated a pooled mean serial interval of 123.24 days (95% credible interval: 91.44, 153.41) (Figure @ref(fig:forest\_plot). As we saw with the individual study estimates (Table @ref(tab:si\_results\_tab)), the meta-analysis provided further evidence of substantial heterogeneity among studies due to a large value of the standard deviation of the random intercepts (31.55). The large variation in the mean serial interval estimates can be visualised by plotting the estimated distributions of each study’s serial interval shown in Figure @ref(si\_dist\_plot). A normal distribution is assumed, and is parameterized by the estimated mean and standard deviation of serial interval for each study.



Forest plot of estimated mean serial intervals and pooled mean.

## Growth Rate and Basic Reproduction Number

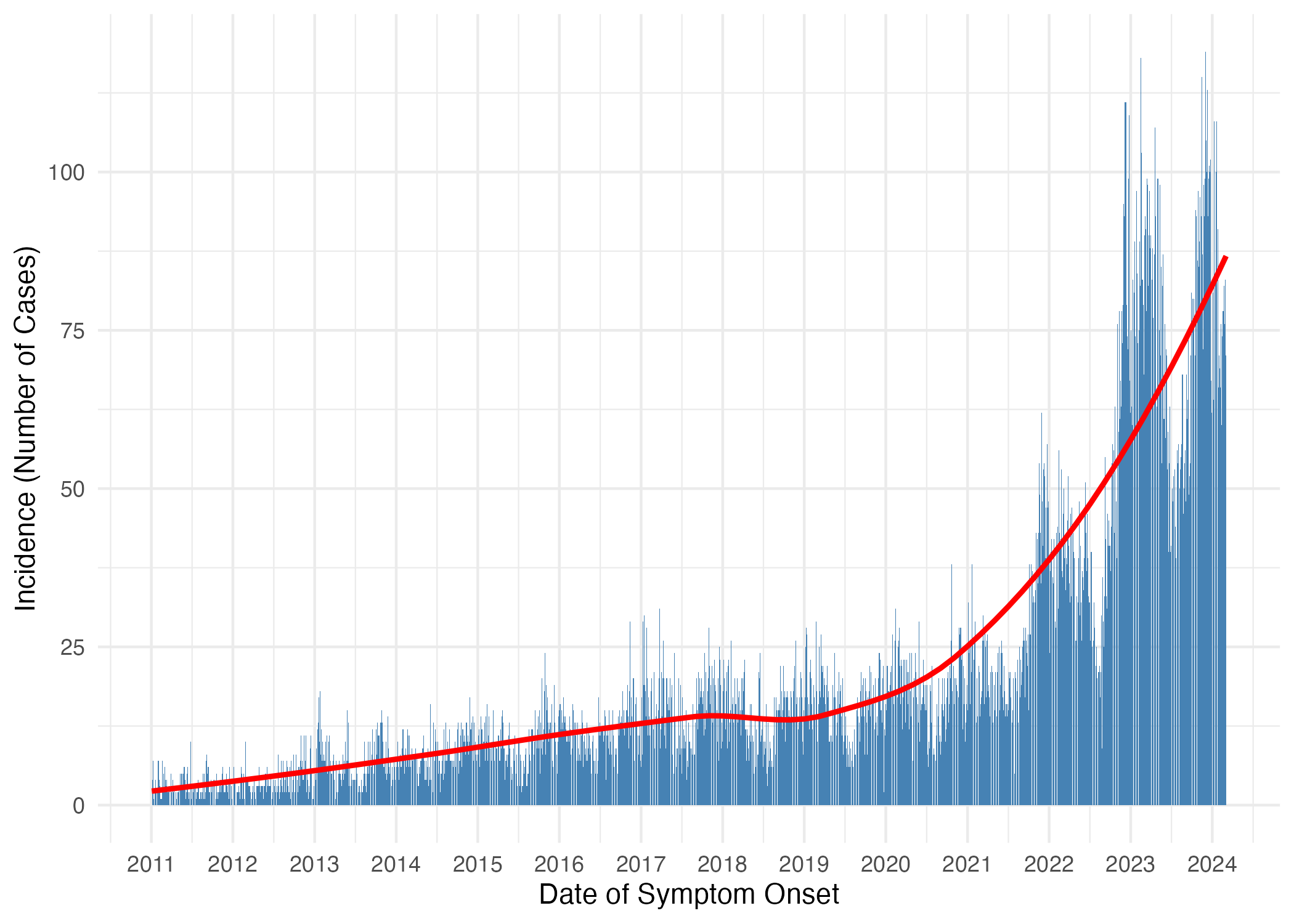
We estimated the annual growth rate of scabies cases by fitting an exponential growth model to annual incidence of scabies per 1000 people from 2011 to 2023 in the Netherlands17. We estimated an annual growth rate of 0.353 (95% CI: 0.152, 0.452). Using the estimated growth rate, we can estimate the basic reproduction number as R0 = 1 + (0.353 \* 1/b), where b is the generation time in days20. Using the previously estimated pooled mean serial interval (123.24 days) as a proxy for generation time, we estimate R0 = 1.002 (1.001, 1.004). Using the fitted exponential growth model and the estimated growth rate, we then determined the projected incidence of scabies per 1000 people until 2033 if no interventions are implemented. We found that there could be a substantial increase in scabies incidence in the next 10 years in the Netherlands if no measures are taken to mitigate scabies spread (Figure @ref(fig:growth\_rate\_fig)).



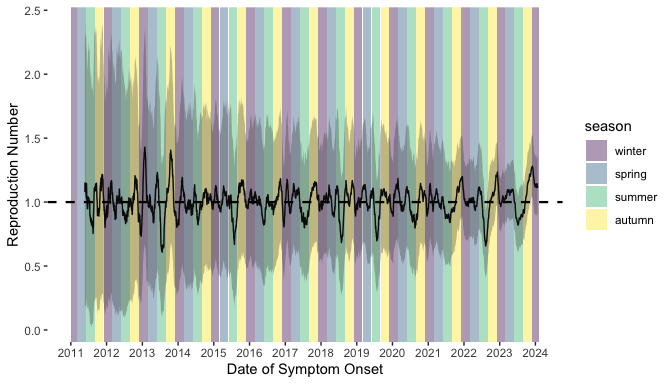
Scabies incidence per 1000 people from 2011 to 2023 (points) and then projected scabies indcidence per 1000 people for 2024 to 2033 (line) using an exponential growth model with annual growth rate = 0.353. Shaded regions represent 95% confidence intervals for projected scabies incidence. The y-axis is displayed on the log-scale and represents the natural log of scabies incidence per 1000 people.

## Time-varying Reproduction Number

To determine the pattern of scabies spread over time, particularly if the incidence of scabies infections display any sort of repeating temporal pattern (e.g., seasonal waves), we estimated time-varying case reproduction number of weekly scabies incidence in The Netherlands from 2011 to 2023 (Figure @ref(fig:rt\_plot)). We see yearly waves of incidence when looking at the incidence of scabies infections over time with the amplitude of incidence gradually increasing over time (Figure @ref(fig:epidemic\_curve)). When examining the time-varying reproduction number we see similar waves of incidence annually (Figure @ref(fig:rt\_plot)), with Rt oscillating around 1. To better determine if these waves of incidence occur within a certain season annually, we plotted time-varying reproduction number against Northern Hemisphere season (colored bars in Figure @ref(fig:rt\_plot)). We found that peaks in transmission occurred most often in autumn and winter season, while the lowest transmission occurred in the summer.



Incidence of scabies infections per 1,000 people in the Netherlands by date of symptom onset. The red line represents a LOESS smoothed trend line, illustrating the overall pattern in the incidence over time.



Time-varying reproduction number of scabies transmission. Colored bands denote season. Winter = December 1 – February 28 (or 29 on leap year); Spring = March 1 – May 31; Summer = June 1 – August. 31; Autumn = September 1 – November 31.

# Discussion

* Discuss time-series trends and population that it typically occurs in in NL (young adults/college students)
  + for policy implications, Rt oscillates around 1, so control efforts would not need to be extreme to control epidemics
  + discuss seasonal pattern in peaks in transmission. Offer possible explanation for why peak transmission tends to occur in late autumn/early winter. Perhaps, increased contact indoors [see Mimouni et al. Br J Dermatol. 2003]. Compare this finding with other findings on seasonality of scabies transmission -> this has been seen in other European countries [See van Deursen 2022 for more references]
* Discuss why estimates of mean serial interval across studies are so different (different populations and identification of symptom onset may be delayed).
* Discuss limitations of the approach and caveats for interpretation.
  + we assume each infection is a person’s first infection.
  + for estimates of Rt, we used the pooled mean and standard deviation from the meta-analysis.

# References

1. Scabies.

2. Deursen, B. van *et al.* Increasing incidence of reported scabies infestations in the netherlands, 2011-2021. *PLoS One* **17**, e0268865 (2022).

3. Reichert, F., Schulz, M., Mertens, E., Lachmann, R. & Aebischer, A. Reemergence of scabies driven by adolescents and young adults, germany, 2009-2018. *Emerg. Infect. Dis.* **27**, 1693–1696 (2021).

4. Redondo-Bravo, L. *et al.* Scabies in spain? A comprehensive epidemiological picture. *PLoS One* **16**, e0258780 (2021).

5. Lugović-Mihić, L. *et al.* An increasing scabies incidence in croatia: A call for coordinated action among dermatologists, physicians and epidemiologists. *Zdr Varst* **59**, 264–272 (2020).

6. Amato, E. *et al.* Increase of scabies infestations, norway, 2006 to 2018. *Euro Surveill.* **24**, (2019).

7. Donà, M. G. *et al.* Increasing trend in confirmed scabies cases in the only public dermatological institute of scientific research and care in italy. *Eur. J. Dermatol.* **33**, 709–710 (2023).

8. Mellanby, K. The development of symptoms, parasitic infection and immunity in human scabies. *Parasitology* **35**, 197–206 (1944).

9. Mellanby, K. *Scabies*. (EW Classey, Faringdon, England, 1972).

10. Kinyanjui, T. *et al.* Scabies in residential care homes: Modelling, inference and interventions for well-connected population sub-units. *PLoS Comput. Biol.* **14**, e1006046 (2018).

11. Lydeamore, M. J. *et al.* A biological model of scabies infection dynamics and treatment informs mass drug administration strategies to increase the likelihood of elimination. *Math. Biosci.* **309**, 163–173 (2019).

12. Tellioglu, N. *et al.* Modelling mass drug administration strategies for reducing scabies burden in monrovia, liberia. *Epidemiol. Infect.* **151**, e153 (2023).

13. Hasselaar, J. Nivel primary care database.

14. Vink, M. A., Bootsma, M. C. J. & Wallinga, J. Serial intervals of respiratory infectious diseases: A systematic review and analysis. *Am. J. Epidemiol.* **180**, 865–875 (2014).

15. Bürkner, P.-C. Brms: An R package for bayesian multilevel models using stan. *J. Stat. Softw.* **80**, 1–28 (2017).

16. Harrer, M., Cuijpers, P., Furukawa, T. A. & Ebert, D. D. Chapter 13 bayesian Meta-Analysis.

17. NIVEL. Cijfers ziekten per week in nederland.

18. Wallinga, J. & Teunis, P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am. J. Epidemiol.* **160**, 509–516 (2004).

19. Gostic, K. M. *et al.* Practical considerations for measuring the effective reproductive number, rt. *PLoS Comput. Biol.* **16**, e1008409 (2020).

20. Wallinga, J. & Lipsitch, M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. Biol. Sci.* **274**, 599–604 (2007).

21. Cauchemez, S. *et al.* Real-time estimates in early detection of SARS. *Emerg. Infect. Dis.* **12**, 110–113 (2006).

22. Cauchemez, S., Boëlle, P.-Y., Thomas, G. & Valleron, A.-J. Estimating in real time the efficacy of measures to control emerging communicable diseases. *Am. J. Epidemiol.* **164**, 591–597 (2006).

23. R Core Team. [*R: A Language and Environment for Statistical Computing*](https://www.R-project.org). (R Foundation for Statistical Computing, Vienna, Austria, 2019).

24. Kaburi, B. B. *et al.* Outbreak of scabies among preschool children, accra, ghana, 2017. *BMC Public Health* **19**, 746 (2019).

25. Larrosa, A. *et al.* Nosocomial outbreak of scabies in a hospital in spain. *Euro Surveill.* **8**, 199–203 (2003).

26. Klaus, B. & Strimmer., K. [*Fdrtool: Estimation of (Local) False Discovery Rates and Higher Criticism*](https://CRAN.R-project.org/package=fdrtool). (2021).

27. Wilke, C. O. [*Cowplot: Streamlined Plot Theme and Plot Annotations for ’Ggplot2’*](https://CRAN.R-project.org/package=cowplot). (2024).

28. Wickham, H., Hester, J., Chang, W. & Bryan, J. [*Devtools: Tools to Make Developing r Packages Easier*](https://CRAN.R-project.org/package=devtools). (2022).

29. Wickham, H., François, R., Henry, L., Müller, K. & Vaughan, D. [*Dplyr: A Grammar of Data Manipulation*](https://CRAN.R-project.org/package=dplyr). (2023).

30. Wickham, H., Bryan, J., Barrett, M. & Teucher, A. [*Usethis: Automate Package and Project Setup*](https://CRAN.R-project.org/package=usethis). (2024).

31. Wickham, H. [*Ggplot2: Elegant Graphics for Data Analysis*](https://ggplot2.tidyverse.org). (Springer-Verlag New York, 2016).

# Appendix

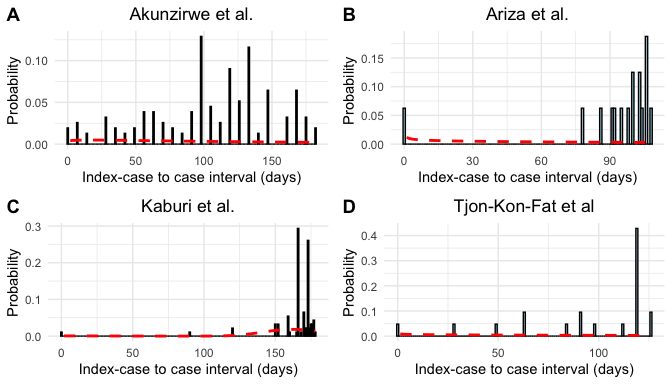
## Sensitivity Analyses

### Serial Interval

We performed a sensitivity analysis on the underlying distribution of serial interval. In the main analysis we assumed the serial interval was normally distributed. In the sensitivity analysis we assumed the serial interval was Gamma distributed. The estimated mean and standard deviation of serial interval for each study is shown in Table @ref(tab:si\_sa). For all studies, except Larrosa et al., the mean is serial interval is lower when estimated assuming an underlying Gamma distribution. When assuming an underlying Gamma distribution, the standard deviations were higher than when assuming an underlying Normal distribution. We see from Figure @ref(fig:gam\_si\_plots) that the Gamma distribution does not fit the data well. It is possible that the Gamma distribution fits scabies data poorly due to the long incubation period of scabies and the possibility of negative serial intervals.

Estimated mean and standard deviation of serial interval from different studies assuming a Normal distribution or Gamma distribution.

| **Study** | **Normal** | | **Gamma** | |
| --- | --- | --- | --- | --- |
|  | **Mean** | **SD** | **Mean** | **SD** |
| Akunzirwe et al. | 122.92385 | 26.920354 | 105.29254 | 92.61143 |
| Ariza et al. | 98.40000 | 8.542332 | 91.49484 | 113.66938 |
| Division of Public and Behavioural Health | 21.91776 | 15.236660 | 20.53631 | 31.98633 |
| Kaburi et al. | 167.34442 | 9.717630 | 164.68478 | 21.41681 |
| Larrosa et al. | 16.10625 | 2.421762 | 23.44372 | 32.53831 |
| Tjon-Kon-Fat et al | 110.71571 | 16.138793 | 94.30253 | 107.28602 |
| SD = standard deviation | | | | |
| **Study** | **Normal** | | **Gamma** | |
|  | **Mean** | **SD** | **Mean** | **SD** |
| Akunzirwe et al. | 122.92385 | 26.920354 | 105.29254 | 92.61143 |
| Ariza et al. | 98.40000 | 8.542332 | 91.49484 | 113.66938 |
| Division of Public and Behavioural Health | 21.91776 | 15.236660 | 20.53631 | 31.98633 |
| Kaburi et al. | 167.34442 | 9.717630 | 164.68478 | 21.41681 |
| Larrosa et al. | 16.10625 | 2.421762 | 23.44372 | 32.53831 |
| Tjon-Kon-Fat et al | 110.71571 | 16.138793 | 94.30253 | 107.28602 |
| SD = standard deviation | | | | |



Epidemic curves and estimated serial interval distributions from four scabies outbreaks. Red line indicates estimated serial interval density assuming an underlying gamma distribution.

We performed a sensitivity analysis in which we altered our choice of prior distribution for mean serial interval. In the main analysis we assumed a prior distribution of N(100,50). In the sensitivity analysis we assumed a prior distribution of N(50, 75) and N(150, 50). We obtained similar estimates of the pooled mean serial interval under the alternative prior distributions (Table @ref(tab:prior\_sa\_tab)).

Estimated pooled mean and standard deviation of serial interval under different prior distributions.

| **Prior** | **Mean** | **SD** |
| --- | --- | --- |
| N(100, 50) | 123.24 | 31.55 |
| N(50, 75) | 120.87 | 32.22 |
| N(150, 50) | 127.15 | 31.73 |

## Computing details

The additional R packages used in this work that have not previously been mentioned or cited in the main text are fdrtool26, cowplot27, devtools28, dplyr29, usethis30, ggplot231. [Packages to add: flextable, ftExtra, knitr, officer, rmarkdown, testthat]