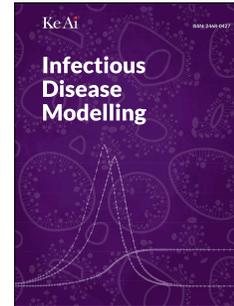


Accepted Manuscript

Evaluating the probability of silent circulation of polio in small populations using the silent circulation statistic

Celeste Vallejo, Carl A.B. Pearson, James Koopman, Thomas J. Hladish



PII: S2468-0427(19)30004-1

DOI: <https://doi.org/10.1016/j.idm.2019.06.003>

Reference: IDM 87

To appear in: *Infectious Disease Modelling*

Received Date: 26 January 2019

Revised Date: 3 June 2019

Accepted Date: 10 June 2019

Please cite this article as: Vallejo C., Pearson C.A.B., Koopman J. & Hladish T.J., Evaluating the probability of silent circulation of polio in small populations using the silent circulation statistic, *Infectious Disease Modelling* (2019), doi: <https://doi.org/10.1016/j.idm.2019.06.003>.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Evaluating the probability of silent circulation of polio in small populations using the silent circulation statistic

Celeste Vallejo^{a,*}, Carl A. B. Pearson^{b,c,d}, James Koopman^e, Thomas J. Hladish^f

^a*Mathematical Biosciences Institute, The Ohio State University*

^b*Department of Infectious Disease Epidemiology & Centre for Mathematical Modelling of Infectious Disease, London School of Hygiene & Tropical Medicine*

^c*South African Centre for Epidemiological Modeling and Analysis, Stellenbosch University*

^d*Very Good Research & Development, LLC*

^e*Department of Epidemiology, University of Michigan*

^f*Emerging Pathogens Institute, University of Florida; Department of Biology, University of Florida*

Abstract

As polio-endemic countries move towards elimination, infrequent first infections and incomplete surveillance make it difficult to determine when the virus has been eliminated from the population. Eichner and Dietz [*American Journal of Epidemiology*, 143, 8 (1996)] proposed a model to estimate the probability of silent polio circulation depending upon when the last paralytic case was detected. Using the same kind of stochastic model they did, we additionally model waning polio immunity in the context of isolated, small, and unvaccinated populations. We compare using the Eichner and Dietz assumption of an initial case at the start of the simulation to a more accurate determination that observes the first case. The former estimates a higher probability of silent circulation in small populations, but this effect diminishes with increasing model population. We also show that stopping the simulation after a specific time estimates a lower probability of silent circulation than when all replicates are run to extinction, though this has limited impact on small populations. Our extensions to the Eichner and Dietz work improve the basis for decisions concerning the proba-

*Corresponding author

¹ P_{SC} : probability of silent circulation, EE: endemic equilibrium, MS: mostly susceptible, ICA: initial case assumption, NICA: no initial case assumption

bility of silent circulation. Further model realism will be needed for accurate silent circulation risk assessment.

Keywords: silent circulation, silent circulation statistic, polio

1. Introduction

Poliovirus remains endemic in Pakistan, Nigeria, and Afghanistan Nnadi et al. (2017). There is evidence that unvaccinated sub-populations within well-vaccinated regions are enabling continued circulation in these areas Nnadi et al. (2017); Mbaeyi et al. (2017). As such regions progress towards eradication, these populations become increasingly hard to detect: intermittent paralytic cases (indicating infrequent first infections) coupled with poor surveillance Nnadi et al. (2017); Mbaeyi et al. (2017) diminishes our ability to reliably detect polio transmission. This potential transmission reservoir could defeat eradication efforts, and by definition defies observation, but model-based approaches can supplement our knowledge. Those efforts should focus on providing decision-oriented measures, most critically the risk of prolonged, undetected circulation. Such modeling work must be able to use the limited data available, such as detected paralytic cases, and be as insensitive as possible to missing data by accounting for censoring that may be introduced by the observation process.

The probability of silent circulation given the time since the last detected paralytic case has been widely considered using various models and methods: from models calibrated to specific regions McCarthy et al. (2016), to general statistical Famulare (2015) and transmission models Eichner & Dietz (1996); Houy (2015); Kalkowska et al. (2012). The transmission model and accompanying estimation of the probability of silent circulation (P_{SC}) as a function of time since the last detected paralytic case, hereafter referred to as the *silent circulation statistic* given in Eichner & Dietz (1996) provides the ground work for directly addressing operational requirements used by real public health organizations: as a part WHO's criteria to declare a region polio-free, a paralytic case can not have occurred for at least three years Henderson (1989). This decision

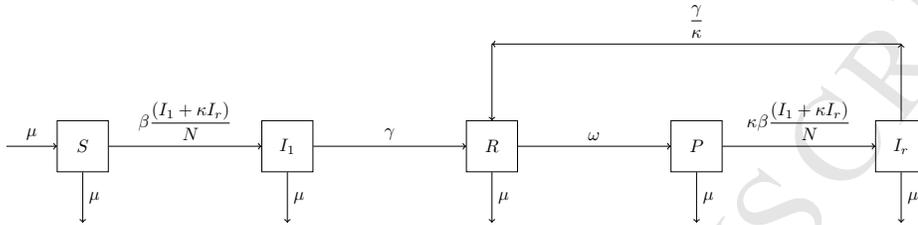
criterion's false positive rate (i.e., incorrectly declaring a region polio free) is P_{SC} at the relevant time. Considering different scenarios via simulation allows us to assess how that false positive rate responds to different situations, and to
 30 assess our confidence applying the criteria. A false positive rate that is sufficiently high would lend further support to using other methods of surveillance for polio circulation, such as environmental surveillance Brouwer et al. (2018), in a more sophisticated manner.

Estimating the silent circulation statistic from models, while tractable, comes
 35 with caveats: the statistic may be sensitive to model assumptions and parameters that in the real world vary with the population of interest. Furthermore, since the silent circulation statistic does not discern the validity of the model—and models may be difficult to validate because of the poor availability of polio data—it is important to analyze its sensitivity to varying model parameters and
 40 model forms. In this paper, we extend the results of Kalkowska et al. (2018) by examining the silent circulation statistic in detail, including assumptions that are inherent to its use. In Kalkowska et al. (2018), Kalkowska et al. correctly reproduced the results in Vallejo et al. (2017) as well as applied the silent circulation statistic to the model output. The goal of this analysis is to create
 45 a basis for better evaluation of the robustness of inferences about the state of polio circulation using the silent circulation statistic.

2. Methods

We calculate the silent circulation statistic for a discrete host, continuous-time stochastic model of polio transmission and vital dynamics, without vacci-
 50 nation, described in Vallejo et al. (2017) (with an adjustment described below) and summarized here. Individuals in the model transition through the following states: S (naive susceptible), I_1 (individuals experiencing their first infection), R (recovered and fully immune from infection), P (partially susceptible to infection), and I_r (individuals experiencing a repeat infection). The I_1 and I_r
 55 compartments are treated differently: individuals suffering a repeat infection

Figure 1: A schematic diagram of the model used in this paper. The compartments of the model are: S (naive susceptible), I_1 (individuals experiencing their first infection), R (recovered and fully immune from infection), P (partially susceptible to infection), and I_r (individuals experiencing a repeat infection). The diagram was reproduced from Vallejo et al. (2017).



have a reduced transmission rate (by a factor of $\kappa < 1$) and recover at a faster rate ($\frac{\gamma}{\kappa}$) than I_1 individuals. Additionally, P individuals have a reduced rate of becoming infected ($\kappa\beta < \beta$). When considering observed cases, we assume that repeat infections do not generate paralytic cases and, therefore, are not detectable Koopman et al. (2017). See Figure 1 for a diagram of the model and Table 1 for the model parameter values.

To maintain a constant population, the model we present here departs from Vallejo et al. (2017) by having a turnover event (combined birth and death), rather than independent births and deaths. We considered three waning immunity scenarios which are described in detail in Koopman et al. (2017). Briefly reviewing the shorthand labels for those scenarios: fast-shallow (immunity wanes quickly, but to a shallow depth [small κ and large ω]), slow-deep (immunity wanes slowly, but to a deep depth [large κ and small ω]), and intermediate (in between fast and slow both in speed and depth [moderate κ and ω]). However, previous work found that varying waning immunity scenarios had minimal impact on dynamics Vallejo et al. (2017). We confirmed these results (see Figs. 7 and 8). Because paralytic case reporting rate substantively influences observed inter-case periods, we also examine the sensitivity of the silent circulation statistic to reporting.

Table 1: These are the parameter values used in the model analysis.

Parameter	Parameter Value	Parameter Description
N	5,000; 10,000; 15,000; 20,000; 25,000	Total population
β	Varies	Infection rate (effective contacts/individual/year)
γ	13	Recovery rate ((year) ⁻¹)
ω	0.2 (0.02) [0.04]	Fast (slow) [intermediate] waning rate ((year) ⁻¹)
κ	0.4179 (0.8434) [0.6383]	Shallow (deep) [intermediate] waning depth
μ	0.02	Natural death rate ((year) ⁻¹)
PIR	0.005	Serotype 1 paralysis to infection ratio

75 When determining the silent circulation statistic, we initialized the population near the endemic equilibrium (EE), after Eichner & Dietz (1996), unless otherwise specified. This assumption is likely reasonable for populations with an extended history of polio transmission, but not for largely susceptible populations with a recent polio introduction. To determine the effect of initializing
 80 at the endemic equilibrium on the silent circulation statistic, we compared it to the probability of circulation when the population was initialized at a mostly susceptible (MS) state. Specifically, simulations begin with one individual in the I_1 compartment and the remaining individuals ($N - 1$) in the naive susceptible compartment (S). To reproduce the initialization conditions described in Eichner & Dietz (1996) (EE), we approximated an extended transmission history by
 85 calculating the endemic equilibrium of an infinite population, and used those compartment sizes to sample from a multinomial distribution to determine the starting conditions for each realization of the model.

After Eichner & Dietz (1996), we assumed that a paralytic case occurred
 90 at the onset of simulation (referred to as the initial case assumption), although we also investigate relaxing this assumption. The authors in Eichner & Dietz (1996) do not state a rationale for the initial case assumption, but a plausible justification is that the occurrence of cases is approximated by a Poisson process and, thus, memoryless. After Eichner & Dietz (1996), we ran each simulation
 95 for 10 years. To assess the error potential of the 10-year time horizon, we also ran all simulations until extinction.

During the simulation, we recorded the time intervals between detected paralytic cases and when polio, stochastically, was eliminated in the population. We define a **circulation interval** to be a time period during which polio is
 100 circulating without detected paralytic cases. Circulation intervals are further subdivided into **extinction intervals**, or intervals of time that end in extinction of the virus; and **intercase intervals**, which end in a paralytic case. By definition, all circulation intervals begin with a detected paralytic case, either explicitly simulated or assumed under the initial case assumption. We calculated the silent circulation statistic (P_{SC}) using the following formula. The
 105

probability of circulation after Δt years without a detected paralytic case is the fraction of intervals that had continued circulation of the virus after Δt years without a detected paralytic case. To construct the numerator of the fraction, we considered the number of intervals that had continued circulation during Δt . Continued circulation during the Δt interval can either occur during an interval that will end in extinction after Δt (extinction interval $\geq \Delta t$) or during an interval that will end in a detected paralytic case sometime after Δt (intercase interval $\geq \Delta t$). Thus, the number of intervals with continued circulation beyond Δt are the number of circulation intervals $\geq \Delta t$. The denominator of the fraction is composed of the number of intervals that have continued circulation beyond Δt and the number of intervals that have gone extinct during Δt (number of extinction intervals $< \Delta t$). Thus, the denominator is given by: number of extinction intervals $< \Delta t$ + number of extinction intervals $\geq \Delta t$ + number of intercase intervals $\geq \Delta t$. This simplifies to: number of extinction intervals + number of intercase intervals $\geq \Delta t$. This gives the probability of silent circulation after Δt years without a detected paralytic case ($P_{SC}(\Delta t)$) to be:

$$P_{SC}(\Delta t) = \frac{\text{number of circulation intervals } \geq \Delta t}{(\text{number of extinction intervals} + \text{number of intercase intervals } \geq \Delta t)} \quad (1)$$

The numerator represents how frequently silent circulation longer than Δt occurred in the model. The denominator represents the set of intervals that are consistent with not observing a case for Δt years. The probability of elimination after Δt years without a detected paralytic case is $P_E(\Delta t) = 1 - P_{SC}(\Delta t)$.

Evaluating the silent circulation statistic as is done in Eichner & Dietz (1996) implies several assumptions. We reviewed these in the context of the initial case assumption (abbreviated as ICA) and right censoring. First, the assumption that a paralytic case had occurred upon starting the simulation may not hold. Populations at the endemic equilibrium are less likely to produce paralytic cases than populations with an above-equilibrium number of infected individuals, either due to dynamics or chance fluctuations. We described the sensitivity of

135 the silent circulation statistic with and without the initial case assumption to
varying population sizes, paralytic case detection rates, β values, and waning
immunity scenarios. To investigate the ICA, we ran a first set of simulations
as described above and in Vallejo et al. (2017), and a second set of simulations
where intervals were only considered if they began with an explicitly simulated
140 paralytic case. Replicates that did not have at least one explicitly simulated
paralytic case were removed; in the real world, these circulation events would
be undetected. Simulations were run until there were 10,000 replicates for com-
parison in both cases.

Second, by truncating the observation window to ten years and not at elim-
145 ination, right censoring may occur if the virus was not eliminated from the
population within the ten year time horizon. To address the effect of right cen-
soring on the silent circulation statistic, we ran two sets of simulations, each
with 10,000 replicates: the first set was run for 10 years or until elimination,
whichever occurred first, and the second set was run until elimination. Elim-
150 ination is defined as the absence of infected individuals in the population. Ad-
ditionally, we show that left censoring has no impact on the silent circulation
statistic given the specific structure of the model under consideration because
the time between paralytic cases is exponentially distributed and is, thus, mem-
oryless.

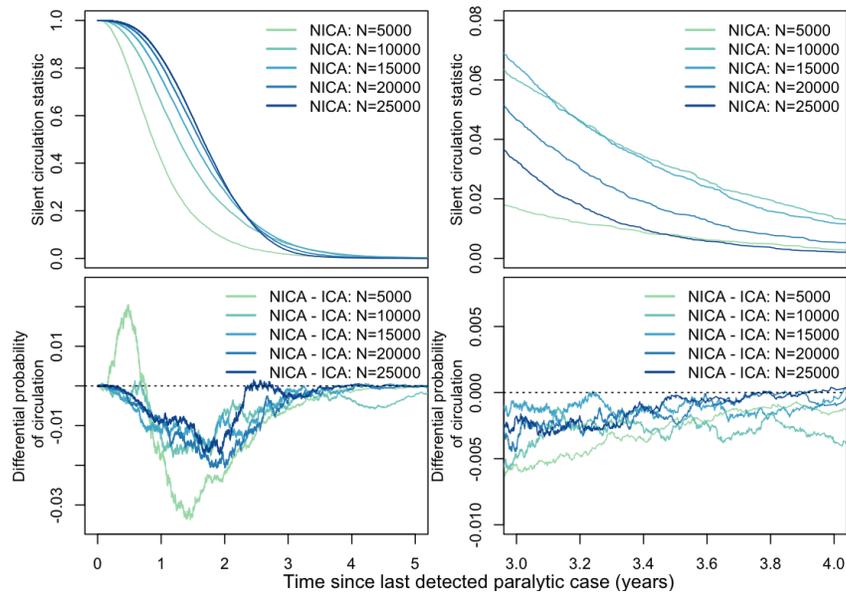
155 We will refer to the following simulation set-up throughout the paper. Unless
otherwise stated, all simulations were run to elimination for 10,000 replicates
assuming a fast-shallow waning immunity scenario with the contact rate (β)
fixed at 135 individuals/year, and a detection probability of 1. In most cases,
population sizes of 5,000 and 20,000 were used. This is to show how the effect
160 of the assumption under consideration varies with population size. All simula-
tions were initialized using a multinomial distribution with success probabilities
determined using the endemic equilibrium solution of the related differential
equations model. For succinctness, this initialization scheme will be referred to
as the endemic equilibrium regime or EE. In S5.1, we considered the effect of
165 initializing with a mostly susceptible (MS) regime.

3. Results

3.1. Initial Case Assumption

Under the initial case assumption (ICA) made in Eichner & Dietz (1996), a
 170 paralytic case is assumed to have occurred at the start of the simulation. As this
 may not accurately reflect the circumstances in the real world, we considered
 the effect of modifying this assumption on the probability of circulation with
 varying population size, paralytic case detection rates, contact rates, and waning
 immunity scenarios.

Figure 2: Effect of population size on the silent circulation statistic. The top panels are the
 silent circulation statistic under NICA (no initial case assumption). The silent circulation
 statistic under ICA (initial case assumption) is omitted for clarity. The bottom panels are
 the differential probability of circulation (NICA - ICA) for each population size considered.
 The right column contains the graphs from the left column restricted between years 3 and 4
 to highlight the behavior around the three year threshold. For all plots, the x-axis is the time
 since the last detected paralytic case (a Δt interval of time).



175 *3.1.1. Population Size*

The population distribution at the endemic equilibrium (EE) is not consistent with the expected distribution at the time a paralytic case has been detected (see Figures S3 and S4). Thus, the initial case assumption (ICA) is incongruous with the EE initialization scheme. To determine the effect of this assumption on the statistic, we explored relaxing it by simulating 10,000 replicates, all with explicitly simulated paralytic cases (NICA). To make comparisons, we used the differential risk of circulation, defined as the difference between NICA and ICA (NICA minus ICA). A positive differential risk of circulation implies $NICA > ICA$ (i.e. NICA predicts a higher probability of circulation) and a negative differential risk of circulation implies $ICA > NICA$ (i.e. ICA predicts a higher probability of circulation). We found that assuming a paralytic case happened at the start of the simulation (initial case assumption; ICA) affects the silent circulation statistic (Figure 2). In the smallest population size considered ($N = 5,000$), the differential risk of circulation (NICA minus ICA) was positive for short intervals (less than one year), but negative for longer intervals. In this case, in the short term, the NICA estimate was selecting against simulations that died out quickly by requiring that at least one paralytic case had occurred. As the population size increased up to 25,000, the ICA estimate generally predicted a higher probability of circulation (negative differential risk of circulation) as compared to the NICA estimate, particularly in the range of 1 to 3 years after a detected case. This was as a result of the dynamics associated with the occurrence of at least one paralytic case. By definition, NICA requires that at least one paralytic case occurs and is detected during the replicate. A paralytic case is more likely to occur with frequent infections. A sufficiently large short-term increase in infections results in a correspondingly large decrease in the susceptible population; the subsequent lull in transmission that follows is when stochastic elimination is most likely. If stochastic elimination does not occur, the period likely becomes a long interval of silent circulation, as low transmission in a population that has disproportionately few fully susceptible

205 people leads to cases being rare. The large discrepancies we observe between NICA and ICA within the first few years ultimately wane: the more time that passes, the less important the initial conditions are to the state of the system.

Since the simulations are initialized at the EE, and paralytic cases are more likely to occur when the population distribution has shifted away from the EE
210 and towards one with more first infections, it is not appropriate to make the initial case assumption. For sufficiently large populations, infections, and thus cases, occur frequently. Therefore, the ICA will not affect the statistic because a negligible (if not zero) proportion of these simulations will not have a detected paralytic case. However, for smaller populations that have a higher probability
215 of not having a detected paralytic case during the simulation, we have shown that the ICA affects the outcome of the silent circulation statistic. In these cases, one implication of using the ICA is that the silent circulation statistic predicts a higher probability of circulation for long intervals of time since a detected paralytic case. Thus, although there is a bias associated with this as-
220 sumption, it is one that produces more conservative elimination probabilities.

3.1.2. Paralytic Case Detection Rate

To consider the effect of the paralytic case detection rate on the silent circulation statistic (both with and without the initial case assumption), we set
225 up the simulations as described in Section 2 with the exception that the case detection probabilities used were 0.25, 0.5 and 1, to simulate the occurrence of under-reporting. These results are given in Figures 3 and 4 for population sizes of 5,000 and 20,000, respectively. Figures 3 and 4 show that decreasing the detection rate has the effect of increasing the probability of circulation with and
230 without the initial case assumption. Since the intervals used to calculate the probability of circulation rely on detecting cases, decreasing the rate at which these occur increases the length of the intervals, thus increasing the probability of circulation.

Furthermore, as the detection rate decreases, the discrepancy between the

235 ICA and NICA predictions generally increases. The increase in disparity is a
 result of dynamics similar to those that are responsible for the difference in
 circulation probability by population size. Since the NICA assumption requires
 that at least one paralytic case be observed in the simulation, these simulations
 had a higher occurrence of infections. Decreasing the detection rate further
 240 restricts the simulations that had at least one paralytic case detected such that
 those that were selected for had an even higher rate of infection. Detected cases
 in the NICA version were indicators of much higher levels of transmission than
 in ICA, and, thus, elimination was more likely to quickly follow. In the case of
 the larger population size (Figure 4), the infection rate is so high already that
 245 the ICA had less of an effect.

Figure 3: Effect of modifying the paralytic case detection rate on the silent circulation statistic for the $N = 5,000$ population size. The top panels are the silent circulation statistic under NICA. The silent circulation statistic under ICA is omitted for clarity. The bottom panels are the differential probability of circulation (NICA - ICA) for each detection rate considered. The right column contains the graphs from the left column restricted between years 3 and 4 to highlight the behavior around the three year threshold. For all plots, the x-axis is the time since the last detected paralytic case (a Δt interval of time).

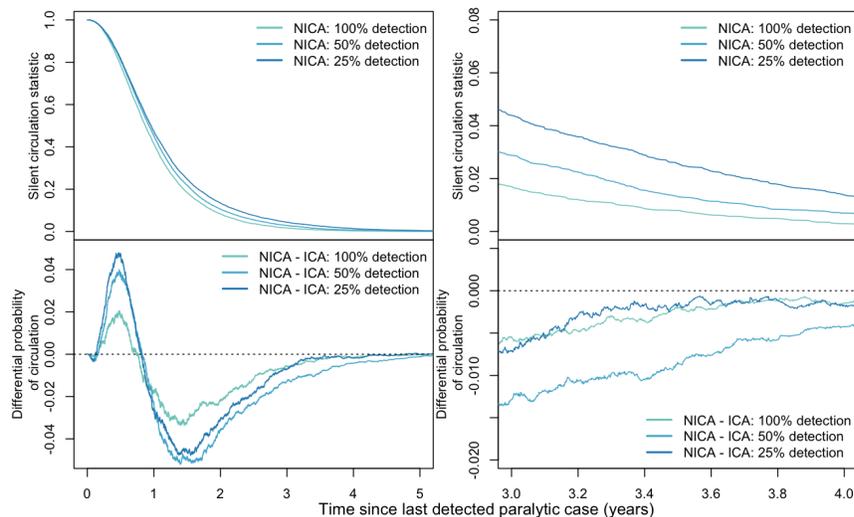
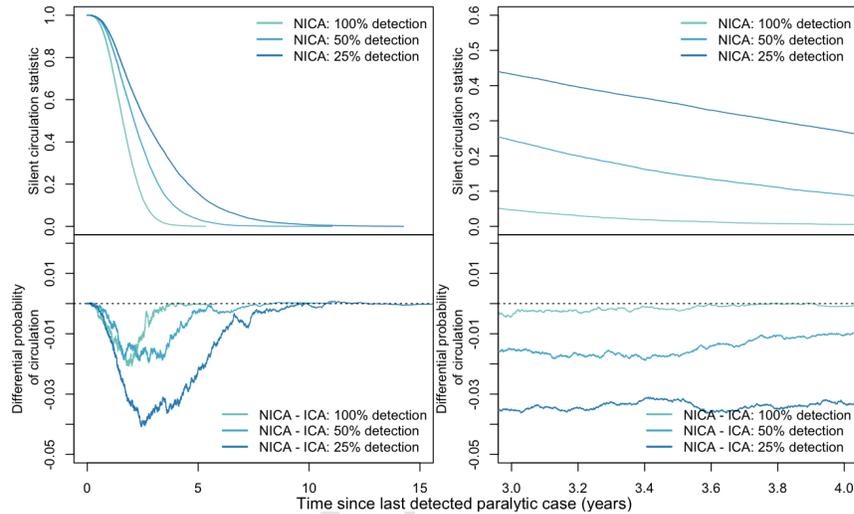


Figure 4: Effect of modifying the paralytic case detection rate on the silent circulation statistic for the $N = 20,000$ population size. The top panels are the silent circulation statistic under NICA. The silent circulation statistic under ICA is omitted for clarity. The bottom panels are the differential probability of circulation (NICA - ICA) for each detection rate considered. The right column contains the graphs from the left column restricted between years 3 and 4 to highlight the behavior around the three year threshold. For all plots, the x-axis is the time since the last detected paralytic case (a Δt interval of time).



3.1.3. Contact Rate and Waning Immunity Scenario

Figures 5 ($N = 5,000$) and 6 ($N = 20,000$) show the silent circulation statistic's sensitivity to varying β (and thus the force of infection) with and without the initial case assumption. Under either assumption, as β increased, the probability of silent circulation increased. A higher β means a higher force of infection, making stochastic extinction less likely since contacts between susceptible and infected individuals are more reliably made. For each β value, the NICA version of the statistic predicted a probability of circulation less than that predicted by the ICA version. This phenomenon is similar to ones described regarding varying population size and varying detection rate.

Figure 5: Effect of varying contact rates on the silent circulation statistic for the population size $N = 5,000$. The top panels are the silent circulation statistic under NICA. The silent circulation statistic under ICA is omitted for clarity. The bottom panels are the differential probability of circulation (NICA - ICA) for each contact rate considered. The right column contains the graphs from the left column restricted between years 3 and 4 to highlight the behavior around the three year threshold. For all plots, the x-axis is the time since the last detected paralytic case (a Δt interval of time).

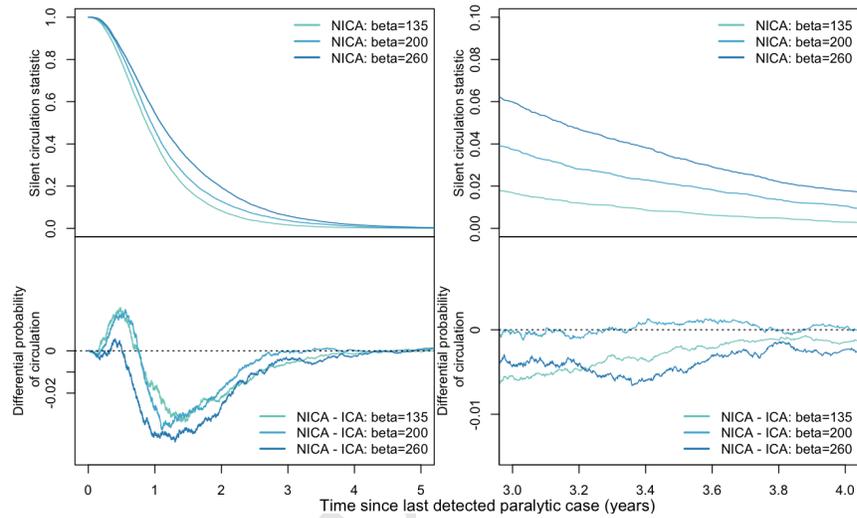
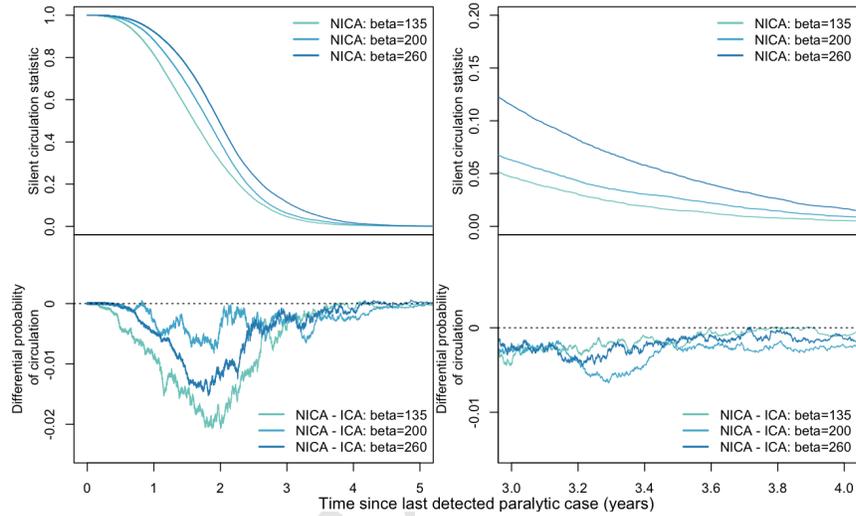


Figure 6: Effect of varying contact rates on the silent circulation statistic for the population size $N = 20,000$. The top panels are the silent circulation statistic under NICA. The silent circulation statistic under ICA is omitted for clarity. The bottom panels are the differential probability of circulation (NICA - ICA) for each contact rate considered. The right column contains the graphs from the left column restricted between years 3 and 4 to highlight the behavior around the three year threshold. For all plots, the x-axis is the time since the last detected paralytic case (a Δt interval of time).



Figures 7 ($N = 5,000$) and 8 ($N = 20,000$) demonstrate how the statistic varies across waning immunity scenarios. The probability of circulation is only slightly impacted by the assumed waning immunity structure. Similar results further confirming the minimal impact of waning immunity scenario selection are shown in Vallejo et al. (2017).

Figure 7: Effect of varying waning immunity scenarios on the silent circulation statistic for the population size $N = 5,000$. The top panels are the silent circulation statistic under NICA. The silent circulation statistic under ICA is omitted for clarity. The bottom panels are the differential probability of circulation (NICA - ICA) for each waning immunity scenario considered. The right column contains the graphs from the left column restricted between years 3 and 4 to highlight the behavior around the three year threshold. For all plots, the x-axis is the time since the last detected paralytic case (a Δt interval of time).

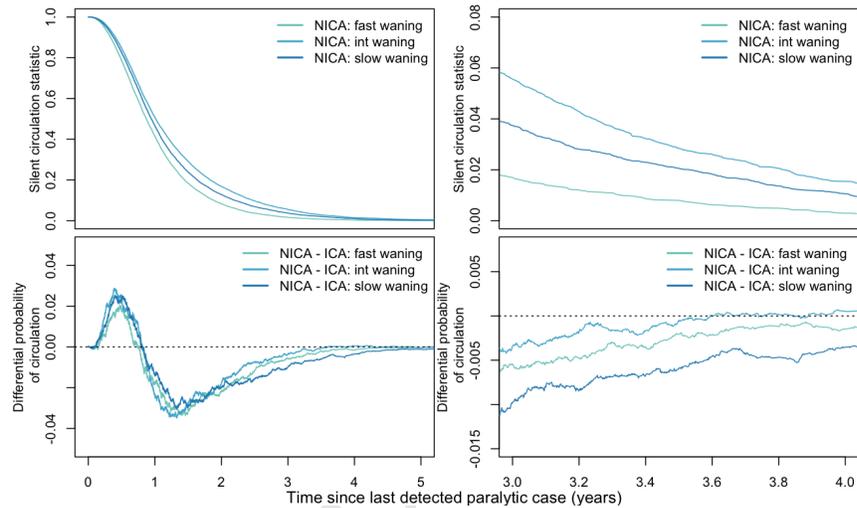
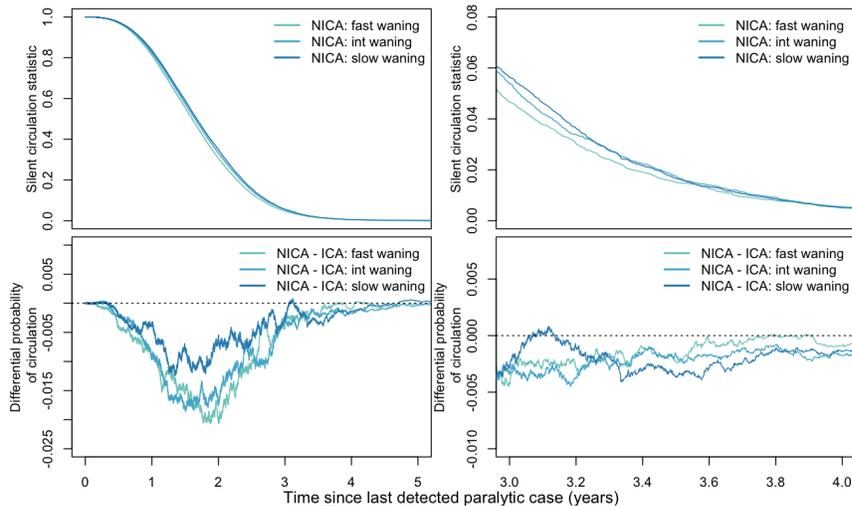


Figure 8: Effect of varying waning immunity scenarios on the silent circulation statistic for the population size $N = 20,000$. The top panels are the silent circulation statistic under NICA. The silent circulation statistic under ICA is omitted for clarity. The bottom panels are the differential probability of circulation (NICA - ICA) for each waning immunity scenario considered. The right column contains the graphs from the left column restricted between years 3 and 4 to highlight the behavior around the three year threshold. For all plots, the x-axis is the time since the last detected paralytic case (a Δt interval of time).



3.1.4. Left Censoring

265 The ICA could cause left censoring to occur if the time between detected paralytic cases is not approximated by an exponential distribution. In this case, neglecting the pre-history of the population before the simulation may cause the initial case-free period in each simulation to be truncated. However, we show that left censoring does not exist under the model conditions described in

270 Section 2. To confirm this, we constructed a quantile-quantile plot comparing the theoretical exponential distribution with mean the time between the start of the simulation and the first explicitly simulated paralytic case with the simulated first Eichner & Dietz intercase intervals (Figure S6). Additionally, in order to show that left censoring of the simulation intervals had no effect on the

275 silent circulation statistic, we modified the simulation data such that the first case-free period was replaced by a time drawn from an exponential distribution

with mean the time between the start of the simulation and the first simulated paralytic case. We compared the silent circulation statistic generated by the unmodified simulation output to the silent circulation statistic generated by the simulation output that was modified as described above (see Figure S7 for one example). We found that since the time between detected paralytic cases is well approximated by an exponential distribution, and the exponential distribution is characterized by the memorylessness property, the expected waiting time between paralytic cases is the same, regardless of when the simulation began. Thus, left censoring does not appreciably influence the analysis of the model.

3.1.5. Population Size-Dependent Effects of the Initial Case Assumption

We have shown that the impact of the ICA on the silent circulation statistic decreases with increasing population size. We hypothesize that this is due to two different effects. The first is interval type distribution. We have discussed three different types of intervals: the first Eichner & Dietz intercase interval (defined as the difference between the start of the simulation and the first simulated paralytic case), true intercase intervals (defined as the time between any consecutive explicitly simulated paralytic cases), and extinction intervals (defined as an interval of time that ends in extinction). The ICA affects the first Eichner & Dietz intercase interval. Therefore, the degree to which the ICA will have an impact on a simulation corresponds to the proportion of intervals in a simulation that are first Eichner & Dietz intercase intervals. This varies with population size. Table 2 shows a comparison of the distribution of intervals by population size. As population size increases, the proportion of intervals that are first Eichner & Dietz intercase intervals decreases, which is a result of an increase in the number of true intercase intervals. To see the effect of interval distribution on the silent circulation statistic, consider the 100% detection differential probability of circulation curve in both Figures 3 and 4. Since the ICA affected 24% of the intervals used to generate the silent circulation curve in Figure 3, the differential probability is larger (highest value of approximately

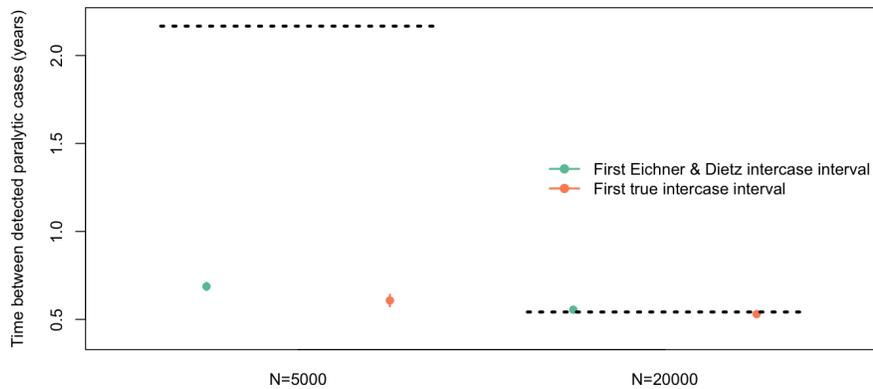
0.035 in absolute value) than the differential probability in Figure 4 (highest value of approximately 0.02 in absolute value) since only 7% of the intervals used to generate the silent circulation curve were affected by the ICA.

310 The second effect that determines the amount by which ICA affects the silent circulation statistic is how different the population disease-state distribution is from that given by the endemic equilibrium of the related differential equations model. At the endemic equilibrium, small perturbations in the population distribution have a minimal impact because, as long as the equilibrium is stable, 315 the system tends to stay near this point. The further the population distribution deviates, the more likely that fluctuations will have a consequential effect on long-term dynamics. To determine the magnitude by which the population distribution changed during the simulation from that of the endemic equilibrium, we compared the time between detected paralytic cases when measured from the 320 beginning of the simulation (first Eichner & Dietz intercase interval) and when measured from the first explicitly simulated case (first true intercase interval) to the time between detected paralytic cases expected at the endemic equilibrium (see Figure 9). The 5,000 population had a longer theoretical time between detected paralytic cases at the endemic equilibrium than actually observed in 325 the simulation, regardless of how it was measured. The shorter simulated time between cases is due to a transient increase in the force of infection: cases are more likely to occur when the force of infection fluctuates higher (Figure S3). As the population size increased, the time between detected paralytic cases more closely approximated that of the endemic equilibrium (black dashed line in Figure 9), regardless of which type of interval considered. Thus, the smaller the 330 population size, the more of an impact that ICA will have because, on average, the population distribution is further away from the distribution at the endemic equilibrium.

Table 2: Comparison across population size of interval distribution. All intervals recorded in the simulation fall into one of three categories: first Eichner & Dietz intercase interval, true intercase interval, and extinction interval. Note that the number of extinction intervals is constant across population size at 10,000.

Interval type	N = 5000	N = 20000
First Eichner & Dietz intercase intervals	24.0%	7.0%
All true intercase intervals	14.0%	86.0%
Extinction intervals	62.0%	7.0%

Figure 9: The dot and whisker plot displays the mean (dot) and 95% confidence intervals (whiskers) for the time between detected paralytic cases for varying population sizes for two points of measurement: measuring the time between paralytic cases that occur from the start of the simulation (first Eichner & Dietz intercase interval) and measuring the time between paralytic cases that occur from the first explicitly simulated case (first true intercase interval). The black dashed lines represent the expected time between detected paralytic cases at the endemic equilibrium of the related differential equations model for each population size considered, calculated under the assumption of an exponential distribution (i.e. the average time between cases is the reciprocal of the mean rate at which cases occur).



335 *3.2. Right Censoring*

As described in Eichner & Dietz (1996), the statistic was originally constructed such that all simulations were terminated after 10 years. Arbitrarily ending the simulations after 10 years has the effect of artificially shortening the last circulation interval for simulations that have not gone extinct within
340 10 years, causing the possibility of right censoring to occur. We explored this effect by comparing simulations that were run to the 10 year time horizon with those that were allowed to run to extinction. For small populations, elimination occurring after 10 years is very unlikely, and thus the effect is minimal (Figure S5). In the case of a population of size 10,000, for example, less than 1% of
345 the simulation replicates lasted beyond 10 years. Stochastic extinction happens more slowly for larger populations, however, and thus right censoring becomes more of a problem. As shown in Figures 10 ($N = 20,000$) and 11 ($N = 25,000$), right censoring had a substantial impact on the duration of circulation predicted by the silent circulation statistic for larger population sizes.

Figure 10: Effect of right censoring on the silent circulation statistic for the $N = 20,000$ population size. The top panel is the silent circulation with right censoring (truncates the simulation at 10 years or elimination, whichever happens first) and without right censoring (simulations end at elimination). The bottom panel presents the differential probability of circulation (the difference between the two curves in the top panel: no right censoring minus right censoring). For both plots, the x-axis is the time since the last detected paralytic case (a Δt interval of time).

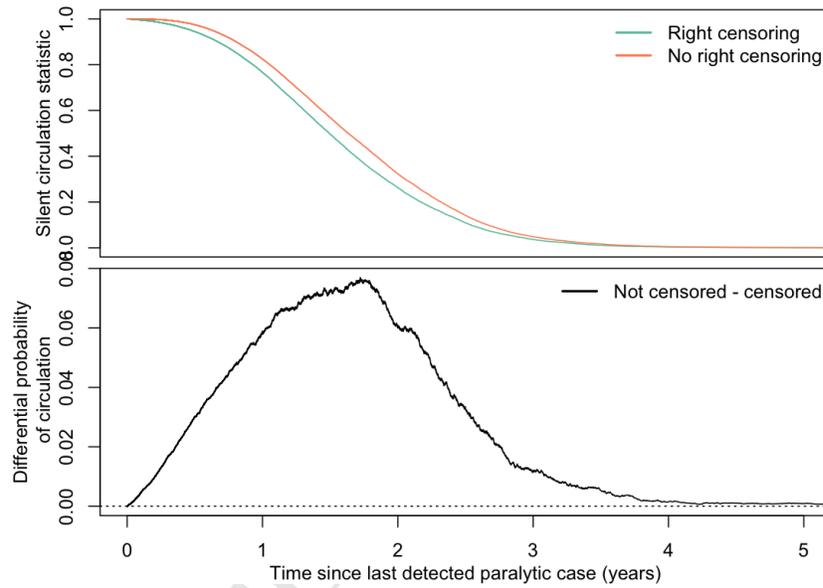
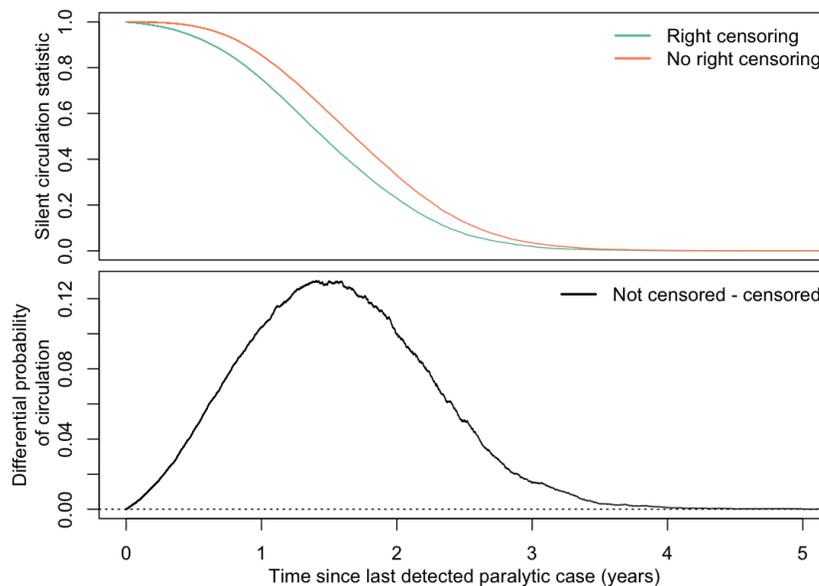


Figure 11: Effect of right censoring on the silent circulation statistic accounting for the $N = 25,000$ population size. The top panel is the silent circulation with right censoring (truncates the simulation at 10 years or elimination, whichever happens first) and without right censoring (simulations end at elimination). The bottom panel presents the differential probability of circulation (the difference between the two curves in the top panel: no right censoring minus right censoring). For both plots, the x-axis is the time since the last detected paralytic case (a Δt interval of time).



350 4. Discussion and Conclusion

In this paper, we presented a more general approach to the silent circulation statistic originally described in Eichner & Dietz (1996) and used in subsequent analyses in Kalkowska et al. (2018) and Kalkowska et al. (2012). This generalization highlights the sensitivity of estimating poliovirus silent circulation probability in different contexts, corresponding to different models. We demonstrated this sensitivity for a novel context relevant to elimination: unvaccinated, small populations. We also identified conditions where analytic artifacts impact the statistic. These included assumptions regarding the most recent case prior to simulation and truncation of simulation results.

360 Whether or not a paralytic case is assumed to have occurred at the start of
the simulation affects the estimated probability of silent circulation. We found
that under the model assumptions examined, assuming a case generally led to
estimating a higher likelihood of circulation. The exception to this was in very
small populations for relatively short intervals (less than a year) since the last
365 observed case. For the other scenarios we considered (i.e. high vs. low paralytic
case detection, high vs. low transmission, and fast vs. slow immunity waning),
assuming an initial case consistently led to estimating a higher probability of
ongoing silent circulation. Although it is unlikely for a paralytic case to have
occurred at the start of the simulation since the population is initialized at the
370 endemic equilibrium, the bias of the assumption is toward a higher probability
of circulation which can be a more conservative approach depending upon re-
source availability.

We demonstrated that the assumption of right censoring implicitly implied
by the construction of the statistic (ten-year time horizon) had very minimal
375 impact on the predicted duration of silent circulation for smaller populations,
but produced substantially different circulation probabilities for larger popula-
tions. To mitigate this effect, the silent circulation statistic should be applied to
simulations that are run to extinction, rather than to a specified time horizon.

Since the time between detected paralytic cases followed an exponential dis-
380 tribution, left censoring had no impact on the statistic. However, assuming the
time between detected paralytic cases is exponentially distributed is unrealistic,
and initial exploration of this issue indicates it could be problematic. Further
examination of this problem is needed.

In examining these small populations, we found that silent circulation per-
385 sisted beyond the three year time horizon used for elimination declaration by
the WHO. This suggests that the threshold for elimination declaration should
be revisited as it may not be appropriate for small populations in the absence
of vaccination.

It should be emphasized that these findings do not imply that in all scenar-
390 ios the silent circulation statistic under the ICA assumption would overestimate

the risk of silent circulation. Given the conditions examined in this paper of a single isolated population, in the absence of vaccination, and in which infection durations are exponentially distributed, the silent circulation statistic is biased towards a higher probability of circulation than under the NICA assumption.

395 This may not be the case for models that incorporate vaccination or seasonality. In areas with a history of vaccination, waning immunity might increase the risks of silent circulation, whereas in our model, waning immunity had minimal impact on transmission. Incorporating seasonality may have nuanced effects. If we consider seasonality that increases transmission during the high season

400 and decreases transmission during the low season (relative to a constant level of transmission), seasonality may have the effect of increasing the probability of extinction due to a more complete depletion of susceptibles during the high season, meaning transmission chains in the low season have both low transmission probability and fewer potential hosts. In contrast, seasonality may have the

405 effect of decreasing the probability of extinction if seasonality tends to curtail epidemics, keeping a higher population of susceptibles available for transmission chains. Moreover, inferences based on the silent circulation statistic could be affected by the complex dynamics arising from more realistic durations of infection than those modeled in this paper, or from metapopulation structure

410 where transmissions between metapopulations might sustain transmissions.

In Kalkowska et al. (2012), the issue of the ten-year time horizon was investigated using a model in which, among other differences such as the inclusion of vaccination and a focus on large populations, immunity was assumed to be permanent. They found that a five-year time horizon significantly reduced the

415 probability of extinction and that the ten-year time horizon had a minimal effect. Conversely, we showed that, given the structure of our model in which immunity is temporary and in the absence of vaccination, the ten-year time horizon becomes more problematic as the size of the population increases. Future work should examine the effects of vaccination, seasonality, and immunity

420 assumptions in order to better understand their effect on polio circulation in endemic regions.

In this paper, we constructed a generalized approach to the silent circulation statistic described in Eichner & Dietz (1996). We demonstrated how to account for its limitations when applied to models of small populations. The understanding of the nuances of the silent circulation statistic developed in this paper will be useful in helping to accurately predict when a region can be declared polio-free.

5. Supplementary Information

5.1. Initialization Assumption

Figure S1: Effect of different initialization schemes on the silent circulation statistic for the population size $N = 5,000$. The top panel presents the silent circulation statistic under four assumptions: initializing at the endemic equilibrium with the initial case assumption (EE: ICA), initializing at the endemic equilibrium without the initial case assumption (EE: NICA), initializing in a mostly susceptible state with the initial case assumption (MS: ICA), and initializing in a mostly susceptible state without the initial case assumption (MS: NICA). The bottom panel gives the differential probability of circulation, which measures the effect of the ICA for each initialization scenario. The differential probability of circulation is the difference between curves of the same color in the top panel, calculated by subtracting the P_{SC} for the silent circulation statistic with the initial case assumption (ICA) from the P_{SC} for the silent circulation statistic without the ICA (NICA). A positive differential probability of circulation implies $NICA > ICA$ (i.e. NICA predicts a higher probability of circulation) and a negative differential probability of circulation implies $ICA > NICA$ (i.e. ICA predicts a higher probability of circulation). For both plots, the x-axis is the time since the last detected paralytic case (a Δt interval of time).

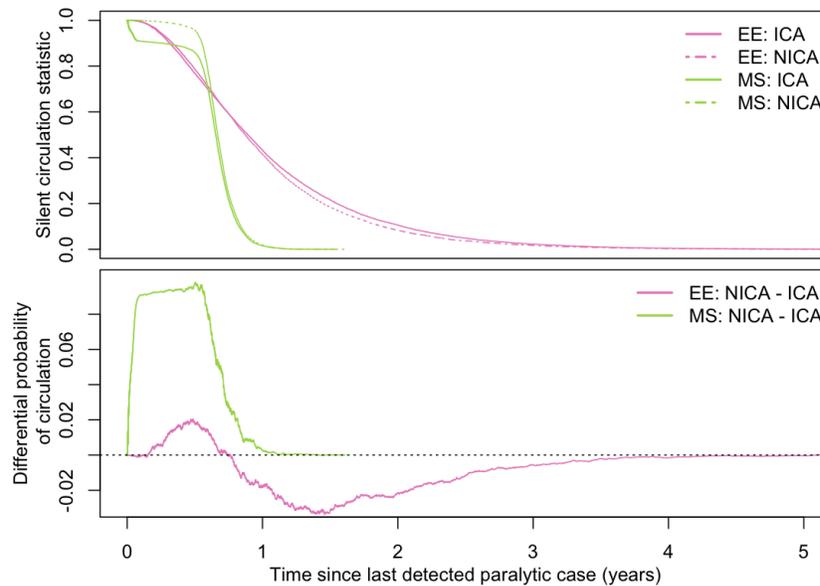
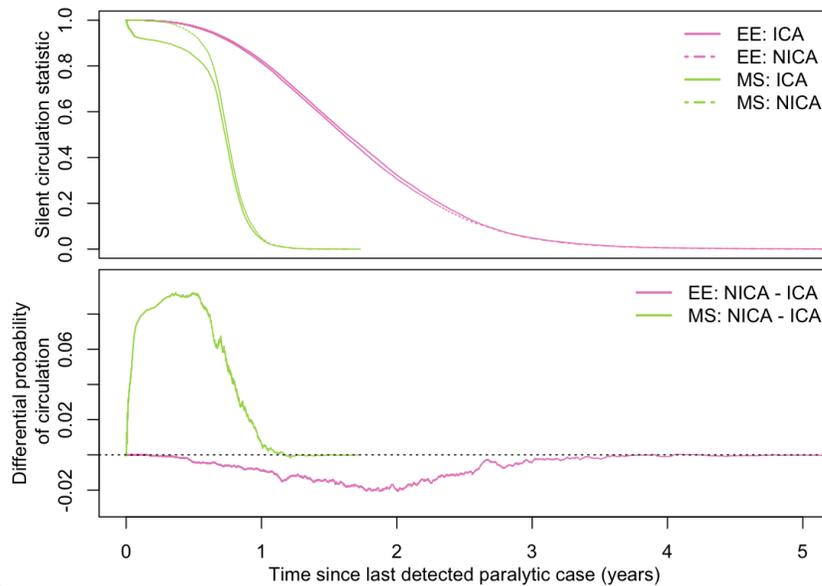


Figure S2: Effect of different initialization schemes on the silent circulation statistic for the population size $N = 20,000$. The top panel presents the silent circulation statistic under four assumptions: initializing at the endemic equilibrium with the initial case assumption (EE: ICA), initializing at the endemic equilibrium without the initial case assumption (EE: NICA), initializing in a mostly susceptible state with the initial case assumption (MS: ICA), and initializing in a mostly susceptible state without the initial case assumption (MS: NICA). The bottom panel gives the differential probability of circulation, which measures the effect of the ICA for each initialization scenario. The differential probability of circulation is the difference between curves of the same color in the top panel, calculated by subtracting the P_{SC} for the silent circulation statistic with the initial case assumption (ICA) from the P_{SC} for the silent circulation statistic without the ICA (NICA). A positive differential probability of circulation implies $NICA > ICA$ (i.e. NICA predicts a higher probability of circulation) and a negative differential probability of circulation implies $ICA > NICA$ (i.e. ICA predicts a higher probability of circulation). For both plots, the x-axis is the time since the last detected paralytic case (a Δt interval of time).



430

It is likely that populations with continued polio circulation are not at the endemic equilibrium. To examine the effect of this assumption on the silent circulation statistic, we considered two different initialization schemes: MS (one initial I_1 infected individual and the remaining individuals in the S compartment) and EE (compartments are initialized using a multinomial distribution

435 with compartment weights given by the endemic equilibrium of the related system of differential equations). Figures S1 and S2 demonstrate the effect of varying initialization paradigms for different population sizes ($N = 5,000$, Fig. S1; $N = 20,000$, Fig. S2). Due to the small number of initially infected individuals, rapid, stochastic elimination of the virus is more likely for the MS starting
440 condition than for EE. This is evident by the initial dip (approximately 0.2 years after a detected case) in the ICA curves for both population sizes. Since the NICA analysis conditions results on observing at least one paralytic case during the simulation, immediate die out is improbable, and thus the differential risk of circulation is positive. For the $N = 5,000$ population, the MS: NICA state
445 results in a higher P_{SC} for small Δt than assuming the population begins at the EE. As population size increases, the P_{SC} predicted by the EE initialization regime increases above that predicted by the MS regime. For those replicates that do not experience immediate die out and thus are more likely included among the MS: NICA replicates, the initially small epidemic takes more time to
450 run its course (relative to starting at EE) before stochastic elimination becomes likely due to susceptible depletion. Thus, the probability that a case has not been observed in a small interval of time (less than ≈ 0.6 years after a detected case) is less of an indicator of extinction for the MS regime than the EE regime. For the MS regime, the simulations tend to result in either rapid elimination or
455 a single early, large epidemic followed by rapid elimination.

5.2. Population Disease-State Distribution Comparison

Figure S3: Comparison of the population disease-state distribution at the endemic equilibrium with the population disease-state distribution at the first detected paralytic case. The dark black line represents the median of the box plot and the circles represent outliers. This data was generated for the $N = 5,000$ population size. The compartments are: S (naive susceptible), I_1 (first infected), R (recovered and fully immune), P (partially susceptible), and I_r (reinfectd). At the first detected paralytic case, the first infected force of infection ($FOI = \beta \frac{(I_1 + \kappa I_r)}{N}$) is on average higher than it is at the endemic equilibrium. This is driven by the higher number of I_1 and I_r individuals at the first detected paralytic case.

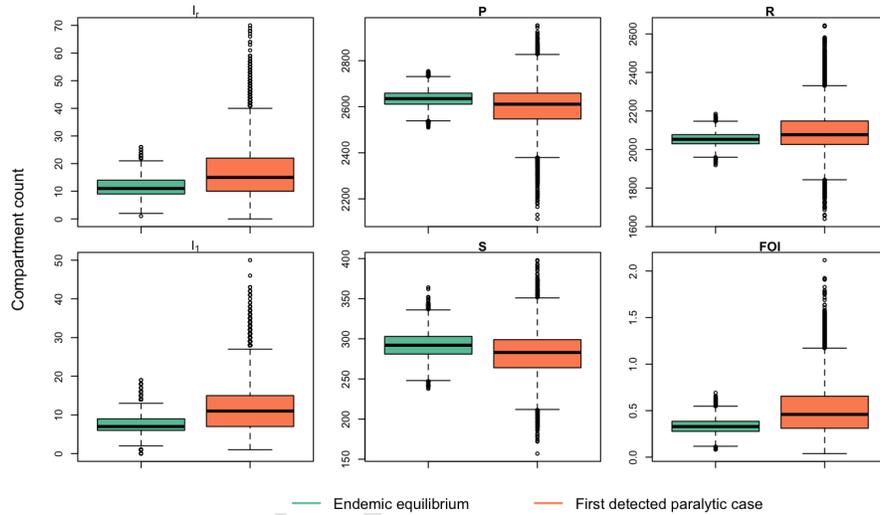
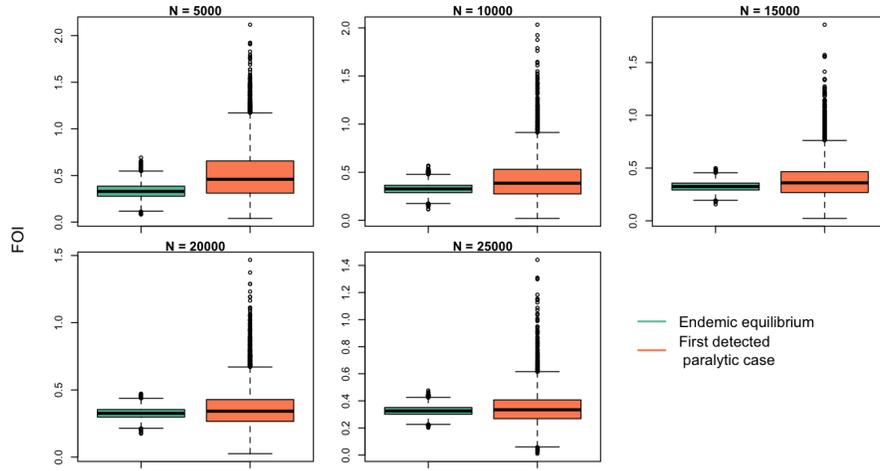
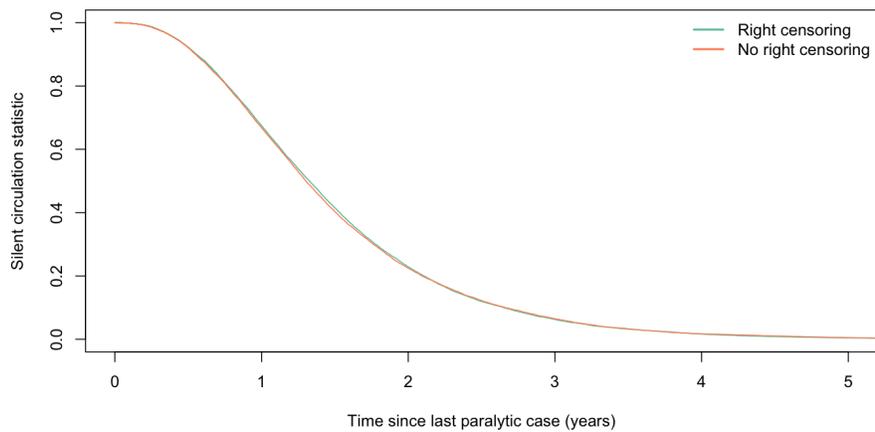


Figure S4: Comparison of the first infected force of infection ($FOI = \beta \frac{(I_1 + \kappa I_r)}{N}$) at the endemic equilibrium with the FOI at the first detected paralytic case across populations. As population size increases, the deviation of FOI at the first detected paralytic case decreases from the expected value at the endemic equilibrium.



5.3. Right Censoring

Figure S5: Effect of right censoring on the silent circulation statistic for the $N = 10,000$ population size. Simulations with right censoring end either at 10 years or with elimination, whichever comes first. Simulations without right censoring end with elimination. The x-axis is the time since the last detected paralytic case (a Δt interval of time).



5.4. Left Censoring

Figure S6: This quantile-quantile plot compares the quantiles of the simulated distribution of first Eichner & Dietz intercase intervals (the time between the start of the simulation and the first explicitly simulated paralytic case) with a theoretical exponential distribution with rate parameter the simulated mean detected paralytic case rate. The red line is $y = x$. The closer the points are to the red line, the more similar the theoretical and simulated distributions. This plot is for $N = 10,000$ population size.

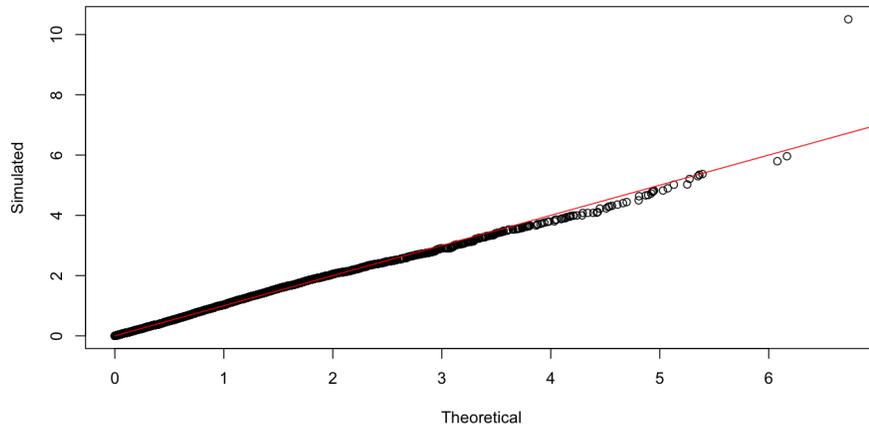
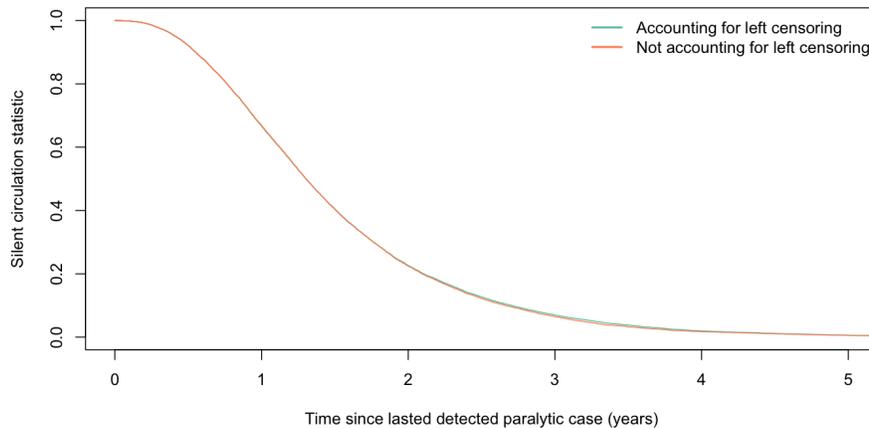


Figure S7: Effect of left censoring on the silent circulation statistic for $N = 10,000$ population size. The x-axis is the time since the last detected paralytic case (a Δt interval of time).



460 6. Acknowledgements

This research was supported in part by a National Institutes of Health/National Institute of General Medical Sciences Grant (<https://www.nih.gov/>; U54 GM111274) and in part by the Mathematical Biosciences Institute and the National Science Foundation under grant DMS 1440386.

465 References

References

- Brouwer, A. F., Eisenberg, J. N. S., Pomeroy, C. D., Shulman, L. M., Hindiyyeh, M., Manor, Y., Grotto, I., Koopman, J. S., & Eisenberg, M. C. (2018). Epidemiology of the silent polio outbreak in Rahat, Israel, based on modeling of environmental surveillance data. *PNAS*, *115*, E10625–E10633.
- 470 Eichner, M., & Dietz, K. (1996). Eradication of Poliomyelitis: When Can One Be Sure That Polio Virus Transmission Has Been Terminated? *American Journal of Epidemiology*, *143*, 816–822.

- Famulare, M. (2015). Has wild poliovirus been eliminated from Nigeria? *PLOS ONE*, *10*, 1–13. URL: <https://doi.org/10.1371/journal.pone.0135765>. doi:10.1371/journal.pone.0135765.
- Henderson, R. (1989). The World Health Organization's Plan of Action for Global Eradication of Poliomyelitis by the Year 2000. *Annals of the New York Academy of Sciences*, *569*, 69–85.
- Houy, N. (2015). The probability of undetected wild poliovirus circulation: Can we do better? *Journal of Theoretical Biology*, *382*, 272 – 278. URL: <http://www.sciencedirect.com/science/article/pii/S0022519315003252>. doi:<https://doi.org/10.1016/j.jtbi.2015.06.046>.
- Kalkowska, D. A., Tebbens, R. J. D., & Thompson, K. M. (2012). The Probability of Undetected Wild Poliovirus Circulation After Apparent Global Interruption of Transmission. *American Journal of Epidemiology*, *175*, 936–949.
- Kalkowska, D. A., Tebbens, R. J. D., & Thompson, K. M. (2018). Another look at silent circulation of poliovirus in small populations. *Infectious Disease Modelling*, *3*, 107–117.
- Koopman, J., Henry, C. J., Park, J. H., Eisenberg, M. C., Ionides, E. L., & Eisenberg, J. N. (2017). Dynamics affecting the risk of silent circulation when oral polio vaccination is stopped. *Epidemics*, *20*, 21–36. [Http://dx.doi.org/10.1016/j.epidem.2017.02.013](http://dx.doi.org/10.1016/j.epidem.2017.02.013).
- Mbaeyi, C., Ryan, M. J., Smith, P., Mahamud, A., Farag, N., Haithami, S., Sharaf, M., Jorba, J. C., & Ehrhardt, D. (2017). *Response to a Large Polio Outbreak in a Setting of Conflict – Middle East, 2013-2017*. Morbidity and Mortality Weekly Report Centers for Disease Control and Prevention.
- McCarthy, K. A., Chabot-Couture, G., & Shuaib, F. (2016). A spatial model of wild poliovirus type 1 in Kano State, Nigeria: calibration and assessment of elimination probability. *BMC Infectious Diseases*, *16*, 521. URL: <https://doi.org/10.1186/s12879-016-1817-3>. doi:10.1186/s12879-016-1817-3.

Nnadi, C., Damisa, E., Esapa, L., Braka, F., Waziri, N., Siddique, A., Jorba, J.,
wa Nganda, G., Oluabunwo, C., Bolu, O., Wiesen, E., & Adamu, U. (2017).
Continued Endemic Wild Poliovirus Transmission in Security-Compromised
505 *Areas – Nigeria, 2016*. Morbidity and Mortality Weekly Report Centers for
Disease Control and Prevention.

Vallejo, C., Keesling, J., Koopman, J., & Singer, B. (2017). Silent Circulation
of Poliovirus in Small Populations. *Infectious Disease Modelling*, 2, 431–440.