See the writing-up section of the CID project handbook, page 51 of 81.

For the document style

• (fin.) Use A4 paper size (210 x 297mm). Set your electronic copy up as if it were going to be printed.

• (fin.) Use standard Arial 11-point font for your main text. You may wish to put specific

headings in larger font sizes; and could use different fonts for specific elements i.e.

quotations, which should also usually be indented and surrounded by quote marks.

• (fin.) Set line spacing at 1.5, and leave a one-line gap between separate paragraphs.

• (fin.) Use margins of 2.54cm (1 inch) all round the page.

• (fin.) Number all pages (in the footer). Page numbers may appear outside the 1-inch margin.

• Tables may be presented in an alternative font, of no less than 8-point size, and

single-line spaced – to help improve visual appearance or fit to the page

Title page

TODO: replace with the template provided via the programmer director.

MSc Project Report 2022-2023

A quantitative simulation-based evaluation of the early detection of poliovirus using environmental surveillance.

Candidate number: 221098

Word count: ????

Standard Project: A minimum of 7,000 words and a maximum of 10,000. All the main content of the project (from the Introduction to the Conclusion, including tables and footnotes) should be included in the word count or page count. Numbers in tables should be counted as corresponding to one word each, as per standard software packages.

Project length: Standard

Table of contents

[Abstract 4](#_Toc139957331)

[Acknowledgements 5](#_Toc139957332)

[Acknowledgement of academic support 5](#_Toc139957333)

[Acknowledgements of other support 5](#_Toc139957334)

[Introduction 6](#_Toc139957335)

[Planning Sections 6](#_Toc139957336)

[Global situation 6](#_Toc139957337)

[Environmental surveillance 6](#_Toc139957338)

[Aims and Objectives 7](#_Toc139957339)

[Aim 7](#_Toc139957340)

[Objectives 7](#_Toc139957341)

[Materials and Methods 7](#_Toc139957342)

[Results 18](#_Toc139957343)

[Discussion 32](#_Toc139957344)

[Recommendations 32](#_Toc139957345)

[Reference list 33](#_Toc139957346)

[Appendices 33](#_Toc139957347)

**List of Tables**

[Table 1. Outbreaks 7](#_Toc138412714)

[Table 2. Outbreaks of polio in outside African continent, Pakistan and Afghanistan. 7](#_Toc138412715)

**List of Figures**

[Figure 1. Test Figure. 7](#_Toc138412726)

# Abstract

TODO: Fill in the abstract in the final stage.

**Background:** Test

**Methods:** Method

**Results:** Reu

**Conclusions:** aaa

Word counts: ?? words

Not exceeding 300 words. Structured Abstract.

# Acknowledgements

## Acknowledgement of academic support

**Project development:** I express the large thanks for my supervisor, Dr Kath O’Reilly for organising and helping my research. I decided to conduct a modelling study in polio field, and Dr. Kath provided me with the aim of my research.

**Contact, input, and support:** I sought the help of feedback from Dr Kath on a monthly basis, and she kindly organised and planned the visit to South Africa.

**Main research work:** I originally conducted most of the literature searching, model formalisation and implementation of simulations. Dr Kath shared the several important literatures related to the environmental surveillance, steered me to focus on the low- and middle-income countries. She gave me the helpful comments on the improvements of the mathematical modelling and helped me interpret the results.

**Writing-up:** Dr Kath has read and left many comments on my master’s thesis.

## Acknowledgements of other support

**Practical assistance:** Dr Kath organised the travel plan for South Africa to conduct a scientific communication with modellers there.

**Permission you were granted:** My research did not need the ethical approval since my project only used secondary sources fully in the public domain.

**Assistance with finance and resources:** I obtained MSc project fund to cover the flight ticket to South Africa.

**Personal acknowledgements:** I would like to express my gratitude to Dr Akira Endo, Professor Sebastian Funk, and the members in the Centre for Mathematical Modelling and Infectious Disease. They all stimulated my academic curiosity and broadened my eyes about data interpretation, and modelling contributions.

I would also like to thank my friends, family and colleagues who supported me in several way. I am willing to give a great thanks to the Rotary Foundation to support me financially and making a community for Rotarians.

# Introduction

## Planning Section (will be deleted)

Most straight forward introduction

* Global spreading situation in polio
  + Mentioning the wild polio spread to Malawi, and cVDPV in UK, US, and other countries including Indonesia.
  + List surveillance ways.
  + (List control measures? Vaccination, SIA?) <- vaccination landscape is important to understand the situation of polio.
* The role of environmental surveillance
  + From the past (1980s Netherlands) to current.
  + Endemic areas, Pakistan, Afghanistan, and Egypt.
  + Crypt circulation in the UK, Israel,
* Past research on the environmental surveillance
  + Ranta and Kath’O Reilly.
  + COVID-19 for the early detection ability.
* State the need for the assessing the benefits of environmental surveillance.
  + Why South Africa?
  + Vaccination landscape?
* South Africa situation. (Methodological part).
  + Vaccine schedule.

Potential ideas

* Age distribution (methodological sections)

TODO: spell out the first acronym appearance.

## Global situation

The start of the main content of your project report should be presented as a formal introductory section – which might typically account for between 10% and 30% of the overall word count. The Introduction should finish by describing the gap in knowledge that your aims and objectives will address.

## Environmental surveillance

Current situation

Ranta1, Surveillance probability2

* Detailed disease surveillance including AFP surveillance and ES is described in [Dominika A Kalkowska 2015](https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-015-0791-5).
* The average time to detection, [Megan Auzenbergs 2023](https://gatesopenresearch.s3.eu-west-1.amazonaws.com/manuscripts/14920/7b375af8-4124-4905-b376-94a924dd9ce1_13272_-_megan_auzenbergs_v2.pdf) .

# Aims and Objectives

## Aim

See 200809\_cross-species-prediction for reference. Aims (the overall goal of the work) and Objectives (what you hoped to be achieved during the project work itself).T

The present study aims to quantitatively assess the early detection ability of environmental surveillance for poliovirus in polio-free countries with mathematical modelling.

## Objectives

The objectives in the present study are as follows:

1. To develop stochastic models to simulate the poliovirus dynamics considering clinical and environmental surveillance.

2. Identify key parameters by calculating the proportion of simulations in which poliovirus is detected more rapidly by environmental surveillance than clinical one, changing various assumptions.

# Materials and Methods

## Planning Section (will be deleted)

The structure for the Materials and Methods

* Study Area, Situation in South Africa.
  + Past history of polio introduction.
  + Date of last confirmed cases. Date of declaration of elimination from the South Africa.
  + Geographical characteristics, and population statistics.
  + Describe the vaccine schedule.
* Overview of model development <- beginning from here?
  + Summarise model description.
* Model components
  + Underlying disease transmission model
    - Data source -> mixed with the model components
      * Population data
      * Vaccination coverage data.
  + AFP development process and AFP surveillance
  + ES surveillance
  + Outcomes for the assessments.
  + Sensitivity analysis.
  + Parameter specification and simulating process
    - Code statements.
* Summarise assumptions.
* Data sharing statements, and ethical statements

## Model description

TODO: more descriptions about wild-type polio virus transmissions.

We constructed the stochastic spatio-temporal model of wild type polio virus assuming a single infection is introduced into polio-free countries among unvaccinated individuals (Figure a). Following the Ranta’s work ([Ranta 2001](https://onlinelibrary-wiley-com.ez.lshtm.ac.uk/doi/pdf/10.1111/0272-4332.t01-1-216174)), we separated the disease dynamics and observation processes since over 99% of polio cases are asymptomatic. The spatio-temporal SEIR model is adopted to simulate the underlying disease dynamics considering the spatial heterogeneity. Once individuals get infected, the become infectious after the incubation period, and recovered after some period. Everyone can spread disease to other areas by making contacts with them. In the following sections, we briefly explain settings and each model, and detailed explanations can be found in the Appendix.

A diagram of a model

Description automatically generated

**Figure a.** **Schematic representation of our model, which is decomposed into three parts: transmission model, ES model, and AFP surveillance model.** Each compartment represents as follows. S, susceptible population; E, latent population for infectiousness; I, infectious population; R, recovered population; A1 ~ A6, latent population for AFP; H, hospitalised population; W, virus concentration in wastewater. Subscript represents the location. Abbreviations: ES, environmental surveillance; AFP, acute flaccid paralysis. TODO: When finalised, export the picture with high resolution.

## Study settings

TODO: strengthen the aspects of importation of poliovirus into polio-free countries.

We focused on wild type polio circulation among unvaccinated children under 5 years old in South Africa for a case study of the assessment of ES compared to AFP surveillance in a polio-free country. For the population data, the children under 5 years old grid data was fetched from the WroldPop, which we aggregated to ~ km \* ~ km grid size ([Andrew J. Tatem 2017](https://www.nature.com/articles/sdata20174), TODO: fill in the final grid size). For simulation efficiency, the population less than ~ individuals in the grid are removed from the simulation. The actual population map used in the present study was shown in Figure b. The most populous area in South Africa is Cape town, ~~ District, following the City 2, District 2, (TODO: Fill the district name).

A map of the country

Description automatically generated

A map of south africa with different colored areas

Description automatically generated

**Figure b. Map of population size (A) and effective vaccination proportion (B) in South Africa.** Latitude diff: 21.31 km, Longitude diff: 19.74 km. TODO: Fetch the shape file, increase the resolution.

A graph of a number of blue bars

Description automatically generated

* Figure r12. Histogram of population size in each? km \*? km grid location.

We focused on the children under 5 years old since in the endemic countries, the majority of polio patients are this population ([Neal Nathanson 2010](https://pubmed.ncbi.nlm.nih.gov/20978089/), [Radboud J Duintjer Tebbens 2018](https://pubmed.ncbi.nlm.nih.gov/29314143/), [Bradley G. Wagner 2014](https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0113538&type=printable)). However, it is noted that individuals aged more than 5 years old were reported in a certain proportion in the imported outbreaks ([Bradley G. Wagner 2014](https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0113538&type=printable)).

South Africa has a extended immunisation programme for children to vaccinate the OPV at birth and ? months after birth and IPV three times until 6 months after birth. So most population receive both OPV and IPV vaccination.

South Africa conducted a survey on the vaccination coverage of OPV and IPV, and calculated the dropout rates at district level ([EXPANDED PROGRAMME ON IMMUNISATION (EPI) NATIONAL COVERAGE SURVEY REPORT 2020](https://www.health.gov.za/wp-content/uploads/2022/03/National-EPI-Coverage-Survey_Final-full-report-Dec-2020.pdf), page 48 – 54). Since OPV vaccination coverage was generally higher than the IPV vaccination coverage, and we took into account the booster vaccination effects of IPV in addition to OPV, we calculated the coverage of population immunised by both OPV and IPV from the vaccination coverage of hexavalent at each step, assuming one vaccination shot has 90% efficacy (TODO: Validate the immunity rates, ref.). Let *C*v,i be the vaccination coverage of *i* th hexavalent vaccine, and *V­E*be the vaccine efficacy of IPV. The effective vaccination coverage, *C*VE at each district is calculated by *C*v,4*V­E* 4 +(*C*v,3 - *C*v,4)*V­E* 3 + (*C*v,2 - *C*v,3)*V­E* 2 + (*C*v,1 - *C*v,2)*V­E*.

1 - (1 – VE)^3

## Transmission model

We adopt the stochastic SEIR model, and this transmission model assumes that the population is grouped into two: non-immunised population and immunised population by IPV and OPV, and we consider the one-zero vaccine effectiveness.

The rationale behind this is as follows. To assume the complete effectiveness, we have to consider the susceptibility, transmissibility, viral shedding, and AFP among vaccinated individuals. Regarding the susceptibility and transmissibility, OPV has shown to prevent further transmissions by developing mucosal immunity. Some evidence exists that OPV reduces the amount and length of viral shedding. Some studies showed that IPV have a large preventive effect against developing AFP, but scarce evidence is present for IPV against susceptibility, transmissibility and viral shedding. From the empirical evidence that in developed countries where IPV is only used for children, no large epidemic is observed, the moderate effectiveness against susceptibility and transmissibility is expected. In the end, since individuals in South Africa generally have OPV and IPV, those vaccinated are expected to have low susceptibility, low transmissibility, low viral shedding and low probability of developing AFP, resulting in the one-zero effectiveness assumption in our model. (TODO: Add references for those evidence.)

Therefore, the force of infection at location *i*, *λ*i, is expressed as

where β is a transmissibility, and α represents the proportion of travellers per day among location i. It is noted that in the initial value for *S*i is set to be *N*u,i, meaning that vaccinated individuals are assumed not to be infected at all, and contributes to reduction of transmissibility through the herd immunity.

*π*ij represents the proportion of travellers from j to i among all travellers moving outside location j. Since we do not have detailed mobilisation data in South Africa as detailed grid size as in the present study, we calculated *π*ij using the radiation model.

Radiation model is originally proposed to predict the commuting and mobility flows between different regions considering the population size ([Filippo Simini 2012](https://www.nature.com/articles/nature10856)), which is shown to work also well in the spread of infectious diseases, particularly in the rural areas ([Amy Wesolowski 2015](https://journals.plos.org/ploscompbiol/article/file?id=10.1371/journal.pcbi.1004267&type=printable)). The advantageous characteristic of this model is parameter-free, and the commuting probability is determined completely by the regional population size, which is expressed as

*N*i represents the population under 5 years old in location *i*, and *N*ijs is the total population in the circle of radius centred at *i* but excluding the source and destination population.

The alternative model to calculate the probability of commuting from j to i is the gravity model ([George Kingsley Zipf 1946](https://www.jstor.org/stable/2087063)), which was frequently used for the spatial spread of infectious diseases such as measles ([Yingcun Xia 2004](https://pubmed.ncbi.nlm.nih.gov/15278849/)) and influenza ([Stefano Merler, and Marco Ajelli 2010](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2842687/pdf/rspb20091605.pdf)). However, the one caveat of the gravity model is requisite of at least three parameter specification about the relationship between the population size and distance, which hinders the generalisability of obtained results. Recently, the promising parameter-free model using the analogy with electricity was proposed ([Kankoé Sallah 2017](https://ij-healthgeographics.biomedcentral.com/articles/10.1186/s12942-017-0115-7)).

We used the discrete Markov process model to simulate our compartment models. Each transition is based on the rate, and sampled from the Binomial process, and these transitions are expressed in the table a. The event time step is one day.

Table a. Discrete Markov process for the compartment model.

|  |  |  |
| --- | --- | --- |
| Event (location *i*) | Effect | Sampling way |
| Birth of susceptible individuals |  |  |
| Death of susceptible individuals |  |  |
| Infection of susceptible individuals |  |  |
| Individuals who will develop AFP |  |  |
| Death of latent individuals |  |  |
| Becoming infectious |  |  |
| Death of infected individuals |  |  |
| Recovered from infection |  |  |
| Death of recovered individuals |  |  |
| Progression of incubation period of AFP from stage k to k +1 |  |  |
|  |  |  |

## AFP development and observation process

Individuals infected with poliovirus will develop AFP with a probability of 1/200, *p*AFP, and those will seek for healthcare after the incubation period due to the severity of paralysis. It generally takes less than one day to seek for healthcare from onset of paralysis to seeking for healthcare ([Sarah Kidd 2020](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7454900/)). To be diagnosed as polio, patients visiting healthcare should be tested and this test result should be positive. Then the probability of AFP patient detected as polio at location *i* at time *t*, *O*t,i, is calculated by

where *P*H, *P*AFP,sample, and*P*AFP,test is the probability of seeking healthcare, taking stool samples, and testing to be positive (sensitivity), respectively. represents the number of individuals who develop AFP at time *t* at location *i*.

The incubation period was given to be 16.5 days and for the stochastic simulation, 6 compartments with 0.329 days-1 transition rates between compartments are assumed to match with the distribution of incubation period. These values were obtained from [Natalia A Molodecky 2023](https://pubmed.ncbi.nlm.nih.gov/34629206/), in which study these values were obtained by fitting an Erlang distribution to 36 independent data of intervals from poliovirus exposure to the onset of AFP ([ALBERT E. CASEY 1942](https://jamanetwork.com/journals/jama/article-abstract/258453), [NICHOLAS C. GRASSLY 2006](https://pubmed.ncbi.nlm.nih.gov/17110580/)).

* TODO: Check 6 compartments with H is right to express the distribution of AFP.
* TODO: Evidence for the effectiveness of IPV against developing the AFP.

## Environmental surveillance model

Main transmission routes for poliovirus is fecal-oral transmission and infectious individuals excreted virus particles to stools, which can be detected in wastewater. Then, the amount of virus in wastewater is governed by the virus shedding parameter, *g*, and pathogen decay rate, *ξ*. If qPCR is used for the poliovirus detection, this wastewater concentration, *W*, is detected after the virus is amplified to the detection threshold, τ. The cycle of amplification needed to detect the virus is called the cycle threshold (CT), denoted as *y* here. Then the relationship between them is expressed as τ = W·*2*y. By rearranging this formula according to *y*, we obtains *y* = log2(*τ/W*). Therefore, if threshold values are available for fitting, we expect *g* and *ξ*. However, the previous study ([Andrew F. Brouwer 2019](https://www.pnas.org/doi/10.1073/pnas.1808798115)) showed that in the framework of SEIR model, even though the pathogen decay rate, *ξ*, is theoretically identifiable given the threshold value data is available (validated by the differential algebra approach to structural identifiability), practically the pathogen decay rate cannot be estimated due to the flat likelihood curve for the large value. Therefore, if the compartment model is expressed as the ordinal differential equation as

and we choose the arbitrary large *ξ*, we can approximate *W*i as *I*i.

From this approximation, we express the probability of detecting virus per one time sampling at location *i* at time *t* as

where *g* is a scaling parameter for the sensitivity of ES detection. Let *n*t,i ­be the index variable of sampling at location i at time t (taking 0 or 1), and *P*ES,test be the test sensitivity for one collected samples. The number of polio-positive sample, *w*t,i is given by

## Outcomes for simulations and sensitivity analysis

We used the cumulative probability of detecting the first polio cases through AFP surveillance of ES over time. Among *N*sim simulations, let *tAFP,i* and *t*ES,i be the timing of the number of simulations detecting poliovirus through type i surveillance by the time of (i takes AFP surveillance or ES). We define the extinction of epidemic as no individuals are infected with poliovirus, meaning that individuals are only present in S, R or H compartments.

In some simulations, the epidemic dies out before any poliovirus is detected through either AFP surveillance or ES, we set these timings as null. We calculated the proportion of the detection patterns, and calculated the lead time of ES over AFP surveillance by *tAFP,I*  - *t*ES,I if poliovirus are detected through both cases.

The probability of detecting poliovirus through type s surveillance by t, *P(T*s*),* is defined as . Where the indicator function *Y*s takes 1 when the poliovirus is detected before t through the type i surveillance, and otherwise takes 0. Then, we also calculate the probability of *P(T*s*)* given the epidemic is not extinct, *Pc(T*s*),* by

where *U*j is the indicator function for *j* th simulation realisation and takes 1 if the simulation does not go extinct, and otherwise takes 0.

For the meta-population model, the definition for *tAFP,I* , *t*ES,I, *Y*i,j and *U*j is defined in the same way. This means even though we introduce the spatial heterogeneity into the model, we do not consider the site for the first detection of poliovirus. In the present study, the detection sites are not important since this triggers public health action to counter the polio outbreak, and spatial heterogeneity is introduced to model the diffusion process of polio and heterogeneity in the catchment area of ES.

Sensitivity analysis was conducted first on our most interest parameters of ES catchment area, sampling frequency, and ES detection sensitivity. Then, we investigated these characteristics of ES and AFP are maintained or not by varying the basic reproduction number, *R*0, and commuting rates between regions, α.

When conducting the sensitivity analysis, we first ran 10,000 simulations for transmission model, and based on these results, we run the stochastic process to detect obtain the first detection date of poliovirus through AFP surveillance and ES varying the parameters related to the surveillance process. For different values of *R*0 and α, the same procedure was performed. (TODO: Write the actual number of simulations.)

All codes for simulation and visualisation are written in Julia v1.8.3 and are available at <https://github.com/toshiakiasakura/polio_environmental_surveillance>. (TODO: is it ok to let my code and resources be publicly available?).

* [GeoBoundaries for acknowledgements](https://www.geoboundaries.org/index.html#getdata)

## Parameter specification

* TODO: Structure the description of parameter specification.

We list the parameters and their values used for the baseline scenario. In addition to the basic reproduction number, *R*0, we introduced the effective reproduction number at time 0, *R*e(0), to represent the potential transmissibility used in our simulation since we only focused on unvaccinated individuals. *R*e(0) is defined as *R*0*C*VE,wwhere *C*VE,w is the effective population weighted average of the effective vaccination coverage. The transmissibility, β, is given by *R*0γ2. We initially explored the transmission model based varying the value of *R*0 and commuting rate, α, and adjusted the value to match with the order of the number of AFP cases (around 1~25 cases for 3 years) and the spreading areas (around 1 ~ 100 grids). Then we arbitrarily chose one value for *R*0 and α.

* Effective immune proportion. [Dominika A Kalkowska 2015](https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-015-0791-5), [Dominika A. Kalkowska 2014](https://academic.oup.com/jid/article/210/suppl_1/S412/2194170), [Kathleen M O'Reilly 2012](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(12)60648-5/fulltext)

Table b. Parameters used in the baseline simulation.

|  |  |  |
| --- | --- | --- |
| Parameters | Values | Ref. |
| Basic reproduction number (R0) | 1.05 / 1.10 | From simulation results. |
| Population size (N),  should be under 15 years old | Grid level. | WorldPop |
| Latent period (γ1) | 1/4 days-1 | [John R. Paul WHO 1955, p14.](https://apps.who.int/iris/bitstream/handle/10665/41659/WHO_MONO_26.pdf?sequence=1&isAllowed=y) |
| Infectiousness period (γ2) | 1/15.02 days-1 | Fitted to [Radboud J. Duintjer Tebbens 2013](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7890644/pdf/nihms-1670167.pdf) |
| Transition rate of incubation period (σ) | 1/3.04 days-1 | [Natalia A Molodecky 2023](https://pubmed.ncbi.nlm.nih.gov/34629206/), [ALBERT E. CASEY 1942](https://jamanetwork.com/journals/jama/article-abstract/258453), [NICHOLAS C. GRASSLY 2006](https://pubmed.ncbi.nlm.nih.gov/17110580/) |
| Paralysis-to-infection ratio (pAFP) | 1/200 for WPV1, | [Neal Nathanson 2010](https://academic.oup.com/aje/article/172/11/1213/194806?login=false) |
| Vaccine effectiveness of IPV (based on seroconversion rate) | 0.63 | [Radboud J Duintjer Tebbens 2018](https://pubmed.ncbi.nlm.nih.gov/29314143/), [Kimberly M. Thompson 2013](https://pubmed.ncbi.nlm.nih.gov/23461599/), [Nicholas C. Grassly 2014](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4197908/) |
| Duration from the onset to the time of seeking healthcare (γ3). The incubation period of paralytics ranges from 7-14 days. Most patients seek medical care with 1 days, [Sarah Kidd 2020](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7454900/) | 7 days | Assumption  TODO: 36 days? From the old literature. |
| Function for virus shedding (g) to sewage (shedding of individuals and delay for sewage detection) | Expert opinion for fraction of the infected population shedding WPV1 | [Radboud J. Duintjer Tebbens 2013](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7890644/pdf/nihms-1670167.pdf) |
| *P*H , Probability of seeking healthcare. | 0.9 | [K. M. O'Reilly 2020](https://www.cambridge.org/core/journals/epidemiology-and-infection/article/surveillance-optimisation-to-detect-poliovirus-in-the-preeradication-era-a-modelling-study-of-england-and-wales/0FEFF0021B9516F067569ED252240295) |
| *P*AFP,sample , Probability of stool sampling | 0.8 | [K. M. O'Reilly 2020](https://www.cambridge.org/core/journals/epidemiology-and-infection/article/surveillance-optimisation-to-detect-poliovirus-in-the-preeradication-era-a-modelling-study-of-england-and-wales/0FEFF0021B9516F067569ED252240295) |
| *P*AFP,test , Sensitivity of polio virus detection when stool is tested. | 0.97 | [K. M. O'Reilly 2020](https://www.cambridge.org/core/journals/epidemiology-and-infection/article/surveillance-optimisation-to-detect-poliovirus-in-the-preeradication-era-a-modelling-study-of-england-and-wales/0FEFF0021B9516F067569ED252240295) |
| *PH*,test , Sensitivity of polio virus testing when environmental sampling is tested. | 0.97 | [K. M. O'Reilly 2020](https://www.cambridge.org/core/journals/epidemiology-and-infection/article/surveillance-optimisation-to-detect-poliovirus-in-the-preeradication-era-a-modelling-study-of-england-and-wales/0FEFF0021B9516F067569ED252240295) |

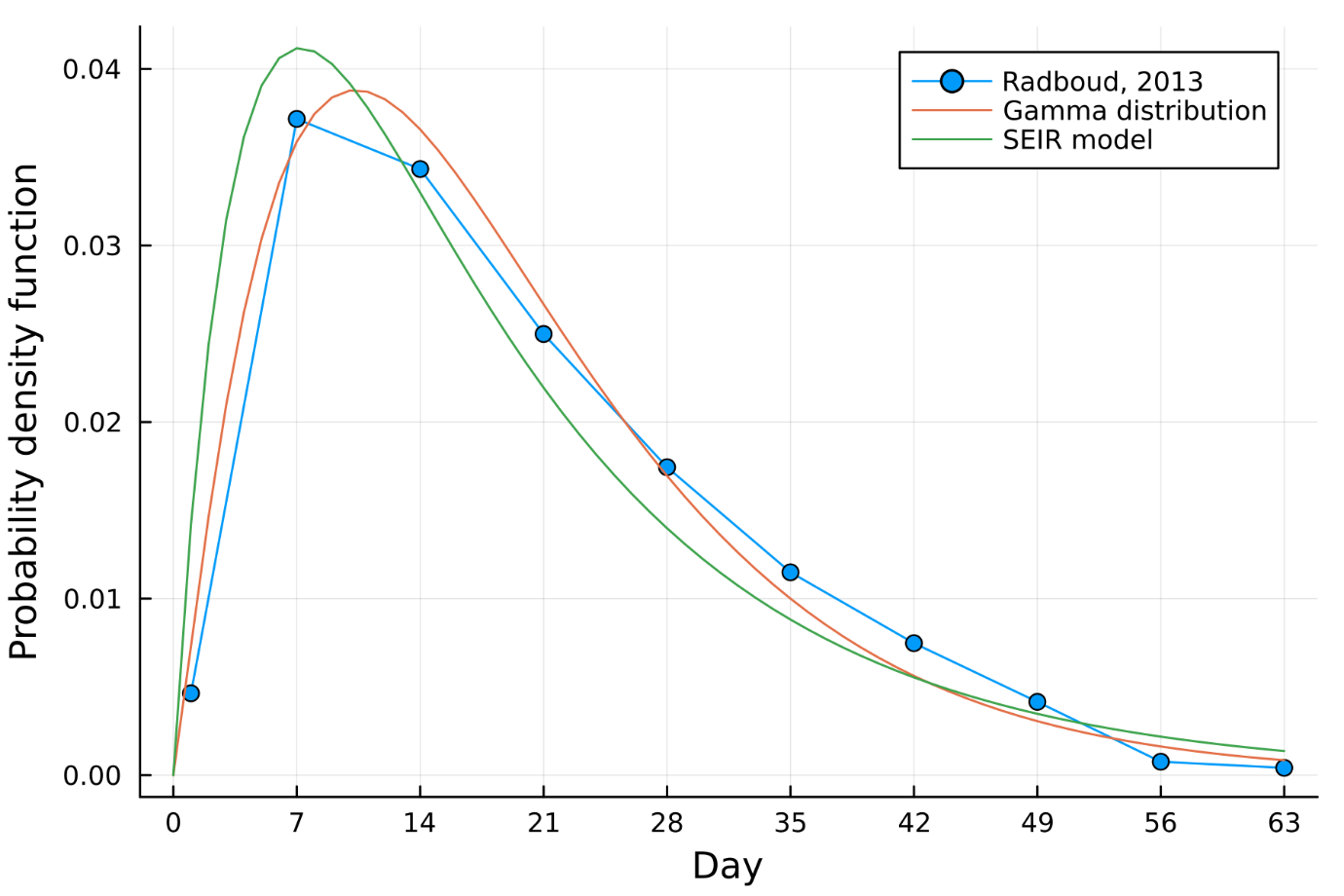
Regarding the recovery period, the prolonged feature of virus shedding is known, and the expert opinion draws the quantitative virus shedding from an infected individual over time. ([Radboud J. Duintjer Tebbens 2013](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7890644/pdf/nihms-1670167.pdf)). So, the simple SEIR model assumes the exponential decay of each compartment, and if we replace the infectious compartment with the n infectious compartment with a recovery period divided by n, the infectious compartment becomes gamma-distributed with the shape parameter of n. We assume the AFP reporting rates are 100% for any sites.

Then we first fit the gamma distribution to the expert opinion data and obtained ?? for the shape parameter. Then we subsequently fit the recovery period to the expert opinion data by minimising the Kullback-Leibler divergence.

We obtained the 2.16 of shape parameter and 8.88 of scale parameter, thus leading to adopt the single compartment for infectious compartment. Then, we fitted the convolution of two exponential distribution to the same distribution to estimate *γ*2 while we fixed the latent period, *γ*1, to be 1/4 day-1.

By fitting with this procedure, we obtain 15.02 days for the recovery period,1/*γ*2.

The fitting results for the gamma curve and this convoluted function is shown in Figure c.



**Figure c. The proportion of individuals excreting virus regardless of amount of virus shedding over time scaled to be the probability density function, and the gamma and convoluted exponential distribution both of which fitted to that distribution.** TODO: Remove the title and write the labels appropriately.

* TODO: Rational of using the proportion of individuals excreting virus regardless of amount of virus shedding.

A graph of a graph

Description automatically generatedA graph of a number of objects

Description automatically generated

Figure temp. Check incubation period of AFP

## Summary of assumptions

Here, we summarise the model assumption.

* We simulated the transmission dynamics among unvaccinated individuals.
* TODO: write the rest of assumptions of the model.

## Ethical statement

All the data used in this paper are publicly available, and the ethical approval is not required, which was confirmed from the MSc ethics committee.

## Mixin field

* The validation for the frequency-dependent model. [T.J. Hagenaars, N.M. Ferguson 2003](https://pubmed.ncbi.nlm.nih.gov/15234202/), [W. M. POST 1983](https://pdodds.w3.uvm.edu/files/papers/others/1983/post1983.pdf)
* Examples of a density-dependent model, [Natalia A Molodecky 2023](https://pubmed.ncbi.nlm.nih.gov/34629206/)
* [Kankoé Sallah 2017](https://ij-healthgeographics.biomedcentral.com/articles/10.1186/s12942-017-0115-7) Reference value for α.
* After introducing birth-death process in the model, we expect the epidemic sustains according to the critical community size. [M. J. KEELING AND B. T. GRENFEL 1997](https://pubmed.ncbi.nlm.nih.gov/8974392/)
* The early seminal work is done by [Ranta 2001](https://onlinelibrary-wiley-com.ez.lshtm.ac.uk/doi/pdf/10.1111/0272-4332.t01-1-216174), following the introductions of the detailed spatial heterogeneity by [K. M. O'Reilly 2020](https://www.cambridge.org/core/journals/epidemiology-and-infection/article/surveillance-optimisation-to-detect-poliovirus-in-the-preeradication-era-a-modelling-study-of-england-and-wales/0FEFF0021B9516F067569ED252240295)

# Results

## Planning result section (will be deleted)

Single cohort of individuals

* Figure r3. One single homogenous population. Cumulative probability for both surveillance, and box plots of lead time for a single homogenous population. (Write the number of extinctions for each pattern.)
* Figure r13. Histogram of final AFP cases.

Spatial characteristics

* Table r1. Vaccination coverage used for each region.
* Figure r12. Histogram of population size.
* Figure r1. Histogram and example map for the parameter πij. (for 4 places).
* Figure r2. Example epidemic trajectories for the baseline settings.
* Figure r4. (A) # of site at least one infection, (B) # of final size of AFP with inside figure excluding 0 cases. (C) 2d histogram combined with (A) and (B), (D) 2 examples of occurrence pattern of AFP over the time.
* Figure r5. Spatial baseline results. Cumulative probability for both surveillance, and box plots of lead time for a single homogenous population. (Write the number of extinctions for each pattern.)
* Figure r11. Summary simulation characteristics of simulations. (A) Histogram of locations at least one infection. (B) Histogram of final size of AFP. (C) Histogram of final size of AFP excluding 0. (D) Weekly number of AFP over time.

Sensitivity analysis.

* Figure r6. Prob of polio detection (A), (C), (E) and Lead time of ES detection (B), (D), (F), varying the following parameters: population coverage of catchment area for (A) and (B); sampling frequency for (C) and (D); environmental surveillance detection sensitivity (E) and (F).
* Figure r7. Scatter plot for the computation time and aggregation scale. (For the inability to compute the very fine scale of simulations).
* Figure r8. Effect of the basic reproduction number (the effective reproduction number among unvaccinated individuals) on the prob of polio detection (A), (C), (E) and lead time of ES (B), (D), (F) for each parameter: population coverage of catchment area for (A) and (B); sampling frequency for (C) and (D); environmental surveillance detection sensitivity (E) and (F).
* Figure r9. Effect of the contact rates of individuals between different regions on the prob of polio detection (A), (C), (E) and lead time of ES (B), (D), (F) for each parameter: population coverage of catchment area for (A) and (B); sampling frequency for (C) and (D); environmental surveillance detection sensitivity (E) and (F).

## AFP and ES surveillance quality for a single site

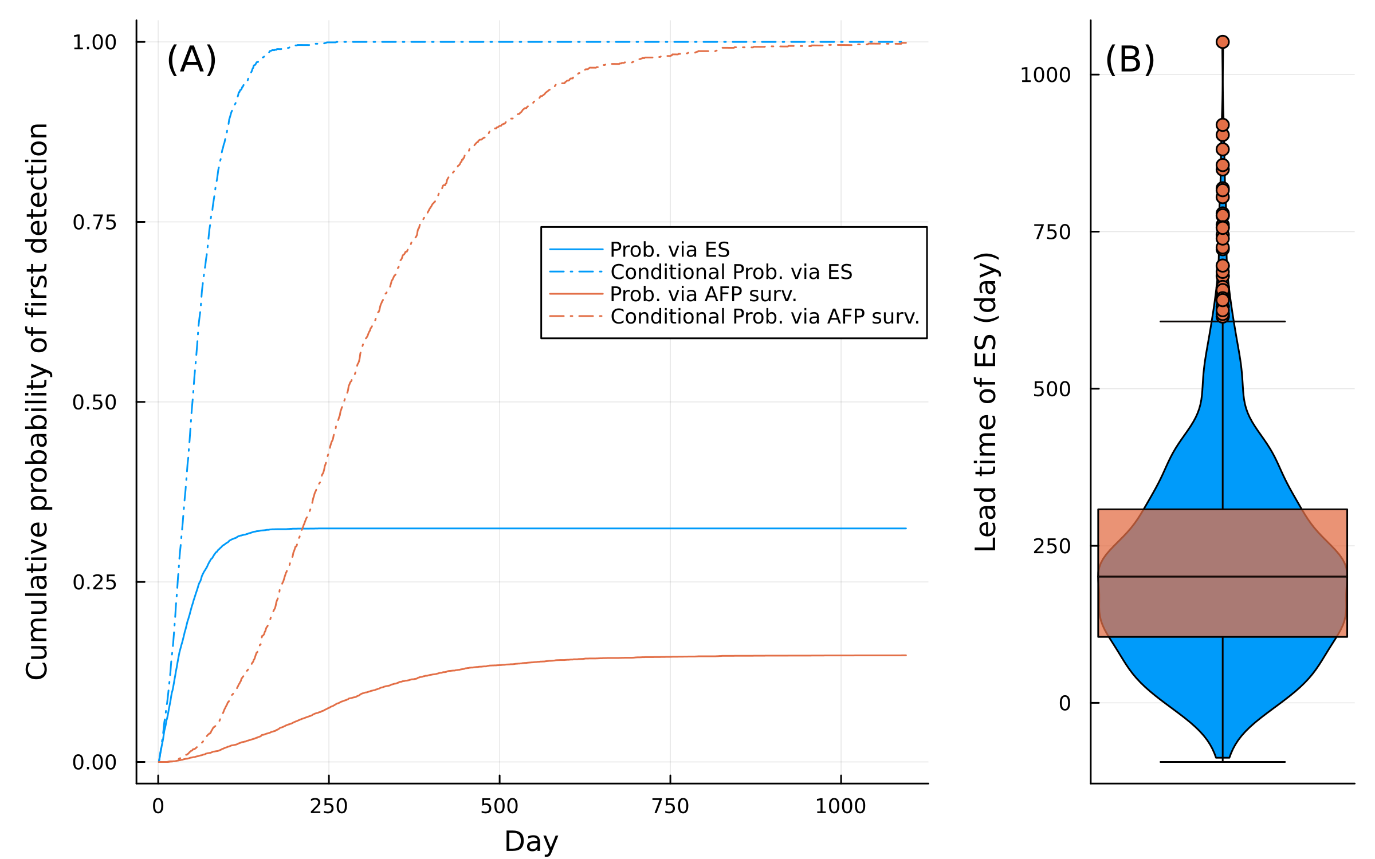


Figure r3. One single homogenous population. Cumulative probability for both surveillance, and box plots of lead time for a single homogenous population. (Write the number of extinctions for each pattern.) (TODO: Create figure). N\_tot: Int64 100000, N\_unvac: Int64 8070 under 10,000 simulations. R0=14, Re(0) = 1.13.

A graph of a number of cases

Description automatically generated

* Figure r13. Histogram of final AFP cases. N\_tot: Int64 100000, N\_unvac: Int64 8070 under 10,000 simulations. R0=14, Re0 = 1.13.

## Spatial characteristics and several example trajectories

Table r1. Vaccination coverage used for each region. TODO: add a column of EVP.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| District | Sample size | OPV0 | OPV1 | HEXA1 | HEXA2 | HEXA3 | HEXA4 | EVP |
| West Coast | 74 | 97.3 | 95.9 | 97.3 | 97.3 | 89.2 | 82.4 | 94.3 |
| Cape Winelands | 179 | 96.1 | 94.4 | 95.0 | 94.4 | 93.9 | 88.8 | 92.8 |
| Overberg | 143 | 97.2 | 94.4 | 94.4 | 94.4 | 96.5 | 84.6 | 94.3 |
| Eden | 121 | 96.7 | 88.4 | 95.9 | 95.0 | 90.9 | 89.3 | 93.3 |
| Central Karoo | 33 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 97.0 | 98.0 |
| Namakwa | 27 | 100.0 | 96.3 | 96.3 | 100.0 | 100.0 | 92.6 | 97.9 |
| Pixley ka Seme | 32 | 93.8 | 87.5 | 90.6 | 81.3 | 78.1 | 68.8 | 85.0 |
| Z F Mgcawu | 61 | 98.4 | 98.4 | 96.7 | 100.0 | 96.7 | 83.6 | 97.3 |
| Frances Baard | 200 | 99.0 | 98.0 | 99.0 | 99.5 | 98.5 | 90.5 | 97.3 |
| Cacadu | 348 | 97.1 | 95.4 | 96.0 | 96.3 | 94.8 | 87.6 | 94.1 |
| Amathole | 220 | 98.6 | 96.4 | 97.7 | 96.8 | 95.0 | 86.4 | 95.1 |
| Chris Hani | 79 | 98.7 | 92.4 | 92.4 | 97.5 | 93.7 | 87.3 | 95.0 |
| Joe Gqabi | 110 | 95.5 | 92.7 | 92.7 | 91.8 | 93.6 | 81.8 | 92.0 |
| O.R.Tambo | 372 | 95.4 | 90.3 | 86.6 | 87.9 | 84.4 | 76.3 | 85.6 |
| Xhariep | 81 | 97.5 | 95.1 | 96.3 | 92.6 | 93.8 | 76.5 | 93.8 |
| Lejweleputswa | 285 | 98.6 | 97.9 | 99.3 | 97.9 | 97.9 | 87.4 | 96.6 |
| Thabo Mofutsanyane | 50 | 100.0 | 96.0 | 98.0 | 98.0 | 98.0 | 86.0 | 95.8 |
| Fezile Dabi | 128 | 91.4 | 81.3 | 82.0 | 82.8 | 82.8 | 82.0 | 81.2 |
| Ugu | 187 | 97.3 | 96.8 | 97.9 | 98.4 | 96.3 | 88.8 | 96.1 |
| Umgungundlovu | 60 | 96.7 | 96.7 | 93.3 | 96.7 | 96.7 | 83.3 | 94.5 |
| Uthukela | 700 | 95.3 | 93.7 | 93.3 | 92.7 | 92.7 | 80.6 | 91.0 |
| Umzinyathi | 205 | 97.6 | 97.6 | 97.1 | 96.6 | 95.6 | 83.9 | 94.6 |
| Amajuba | 470 | 96.0 | 95.3 | 95.3 | 93.6 | 93.8 | 83.2 | 92.8 |
| Zululand | 399 | 96.5 | 95.5 | 95.0 | 95.2 | 93.7 | 77.9 | 92.7 |
| Umkhanyakude | 243 | 97.1 | 95.5 | 93.8 | 94.2 | 93.8 | 87.2 | 92.2 |
| Uthungulu | 986 | 92.4 | 92.3 | 92.7 | 91.7 | 91.3 | 82.8 | 90.3 |
| iLembe | 359 | 79.9 | 78.6 | 76.9 | 76.3 | 76.6 | 72.4 | 75.4 |
| Gert Sibande | 284 | 99.3 | 98.2 | 98.2 | 98.2 | 97.9 | 84.9 | 95.9 |
| Nkangala | 328 | 90.2 | 87.5 | 88.7 | 88.4 | 86.9 | 78.0 | 86.5 |
| Ehlanzeni | 396 | 93.2 | 92.9 | 89.9 | 89.9 | 89.6 | 81.1 | 87.9 |
| Mopani | 680 | 93.5 | 91.5 | 91.9 | 92.1 | 90.4 | 81.6 | 89.9 |
| Vhembe | 589 | 93.0 | 92.5 | 92.2 | 91.9 | 90.8 | 82.0 | 90.0 |
| Capricorn | 493 | 87.4 | 85.8 | 86.0 | 86.6 | 85.8 | 71.6 | 84.4 |
| Waterberg | 111 | 96.4 | 95.5 | 96.4 | 94.6 | 92.8 | 73.9 | 93.1 |
| Bojanala | 299 | 94.3 | 91.6 | 93.0 | 91.3 | 90.6 | 83.3 | 90.3 |
| Ngaka Modiri Molema | 484 | 97.1 | 96.7 | 94.8 | 95.2 | 93.2 | 82.0 | 92.8 |
| Dr Ruth Segomotsi Mompati | 106 | 97.2 | 97.2 | 97.2 | 95.3 | 95.3 | 84.9 | 94.4 |
| Dr Kenneth Kaunda | 210 | 95.7 | 94.8 | 94.8 | 93.3 | 91.9 | 83.8 | 92.1 |
| Sedibeng | 514 | 94.6 | 92.6 | 94.2 | 93.8 | 93.0 | 84.0 | 91.9 |
| Sisonke | 597 | 96.0 | 95.1 | 94.6 | 94.1 | 94.5 | 87.1 | 92.8 |
| Alfred Nzo | 271 | 98.2 | 95.9 | 96.7 | 96.7 | 95.6 | 83.4 | 94.4 |
| John Taolo Gaetsewe | 5 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 98.1 |
| Sekhukhune | 345 | 92.8 | 89.6 | 91.6 | 91.9 | 90.7 | 82.6 | 89.8 |
| West Rand | 382 | 97.4 | 96.9 | 97.4 | 96.1 | 95.8 | 90.1 | 94.9 |
| Buffalo City | 211 | 98.6 | 95.3 | 95.3 | 95.3 | 94.8 | 86.3 | 93.2 |
| City of Cape Town | 101 | 97.0 | 98.0 | 97.0 | 97.0 | 94.1 | 90.1 | 94.7 |
| Ekurhuleni | 890 | 97.1 | 96.1 | 96.6 | 95.5 | 94.9 | 86.9 | 94.1 |
| eThekwini | 887 | 94.0 | 92.6 | 92.4 | 91.7 | 91.8 | 83.8 | 90.3 |
| City of Johannesburg | 1710 | 97.0 | 96.4 | 95.6 | 95.3 | 94.8 | 87.8 | 93.4 |
| Mangaung | 187 | 98.9 | 97.9 | 97.3 | 98.4 | 97.9 | 87.7 | 96.2 |
| Nelson Mandela Bay | 326 | 94.5 | 93.3 | 93.9 | 92.9 | 92.0 | 86.5 | 91.5 |
| City of Tshwane | 621 | 98.9 | 96.3 | 96.5 | 95.3 | 95.8 | 86.5 | 94.5 |

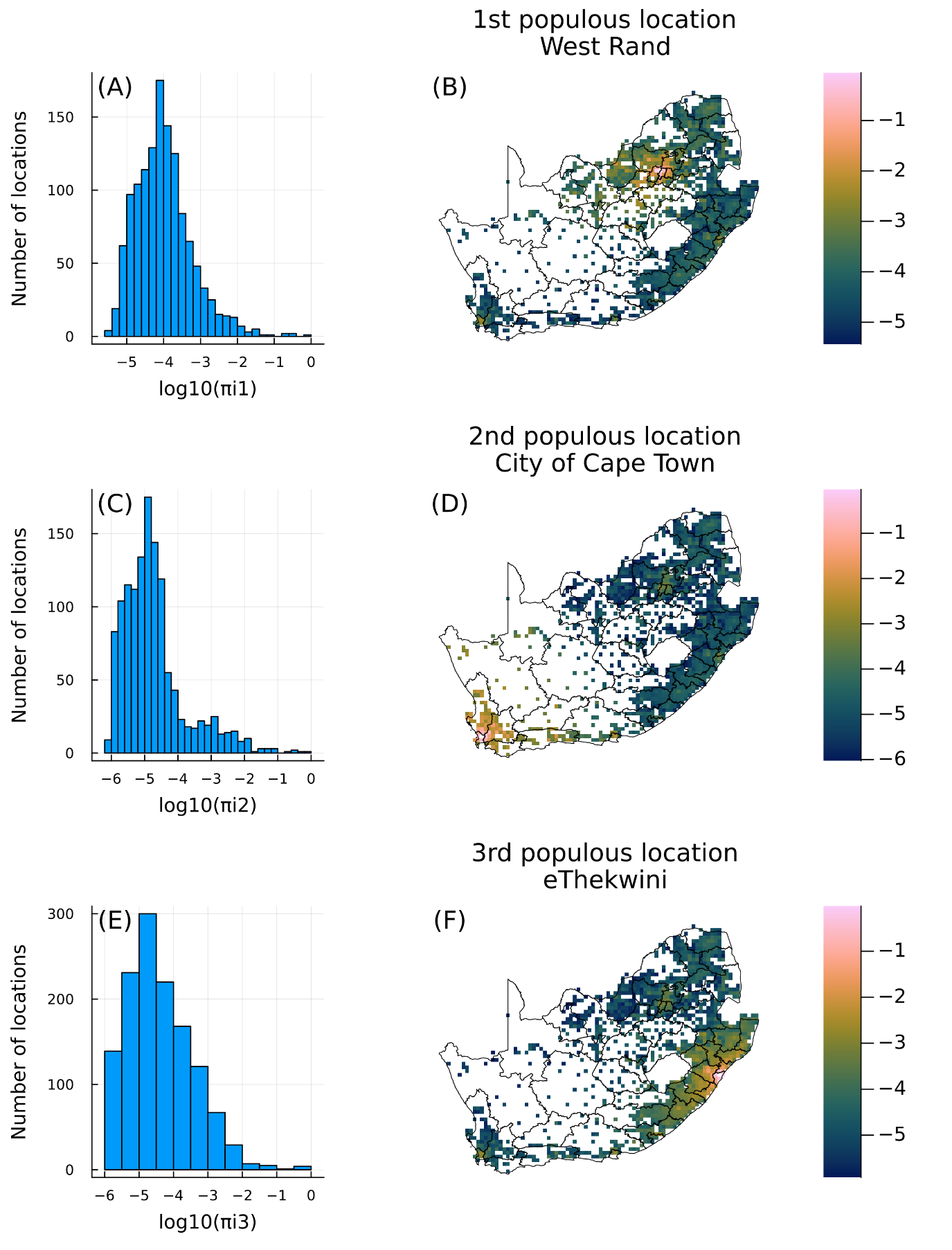


Figure r1. Histogram and example map for the parameter πij. (for 4 places).

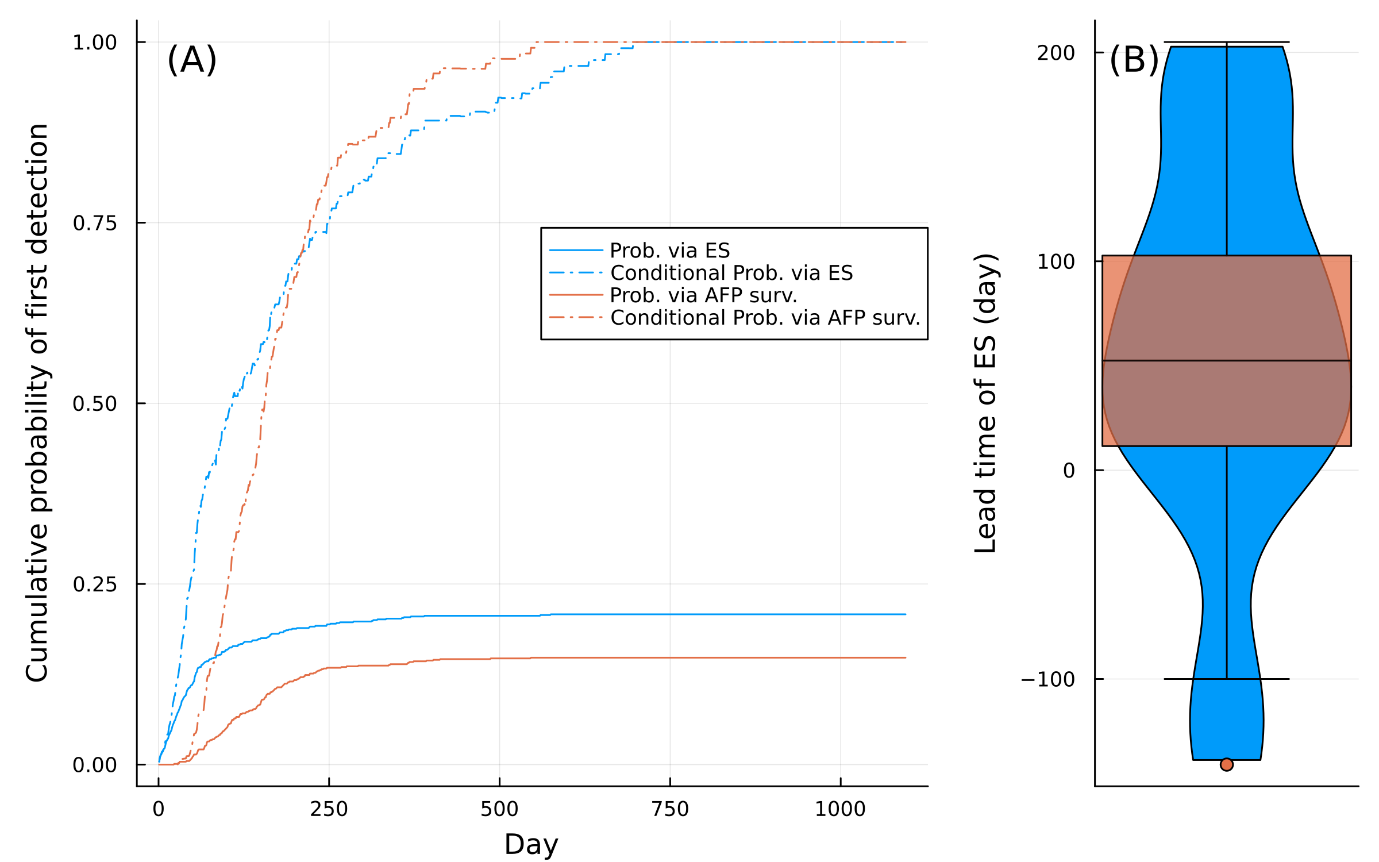
Table r2. Top 20 population grid location used in the present study. (TODO: Cumulative population coverage).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Longitude | Latitude | District | Whole population | EVP | Unimmunized population |
| 27.77 | -26.15 | West Rand | 88395 | 94.9 | 4508 |
| 18.57 | -33.82 | City of Cape Town | 66488 | 94.7 | 3516 |
| 30.83 | -29.79 | eThekwini | 55556 | 90.3 | 5408 |
| 28.15 | -25.96 | City of Johannesburg | 51209 | 93.4 | 3369 |
| 27.96 | -26.15 | City of Johannesburg | 48254 | 93.4 | 3175 |
| 30.83 | -29.60 | iLembe | 46696 | 75.4 | 11483 |
| 18.57 | -34.01 | City of Cape Town | 40666 | 94.7 | 2151 |
| 18.37 | -33.82 | City of Cape Town | 40057 | 94.7 | 2118 |
| 27.96 | -25.96 | City of Johannesburg | 37234 | 93.4 | 2450 |
| 27.96 | -25.39 | City of Tshwane | 35821 | 94.5 | 1983 |
| 28.15 | -26.15 | Ekurhuleni | 35621 | 94.1 | 2110 |
| 25.47 | -33.82 | Nelson Mandela Bay | 34878 | 91.5 | 2962 |
| 27.77 | -26.34 | West Rand | 27991 | 94.9 | 1427 |
| 27.96 | -25.77 | City of Tshwane | 25726 | 94.5 | 1424 |
| 28.15 | -25.58 | City of Tshwane | 25023 | 94.5 | 1385 |
| 27.77 | -26.54 | Sedibeng | 24300 | 91.9 | 1965 |
| 27.77 | -25.96 | West Rand | 20628 | 94.9 | 1052 |
| 28.34 | -25.58 | City of Tshwane | 19920 | 94.5 | 1103 |
| 28.34 | -26.15 | Ekurhuleni | 19597 | 94.1 | 1161 |
| 18.37 | -34.01 | City of Cape Town | 17993 | 94.7 | 952 |

A screenshot of a graph

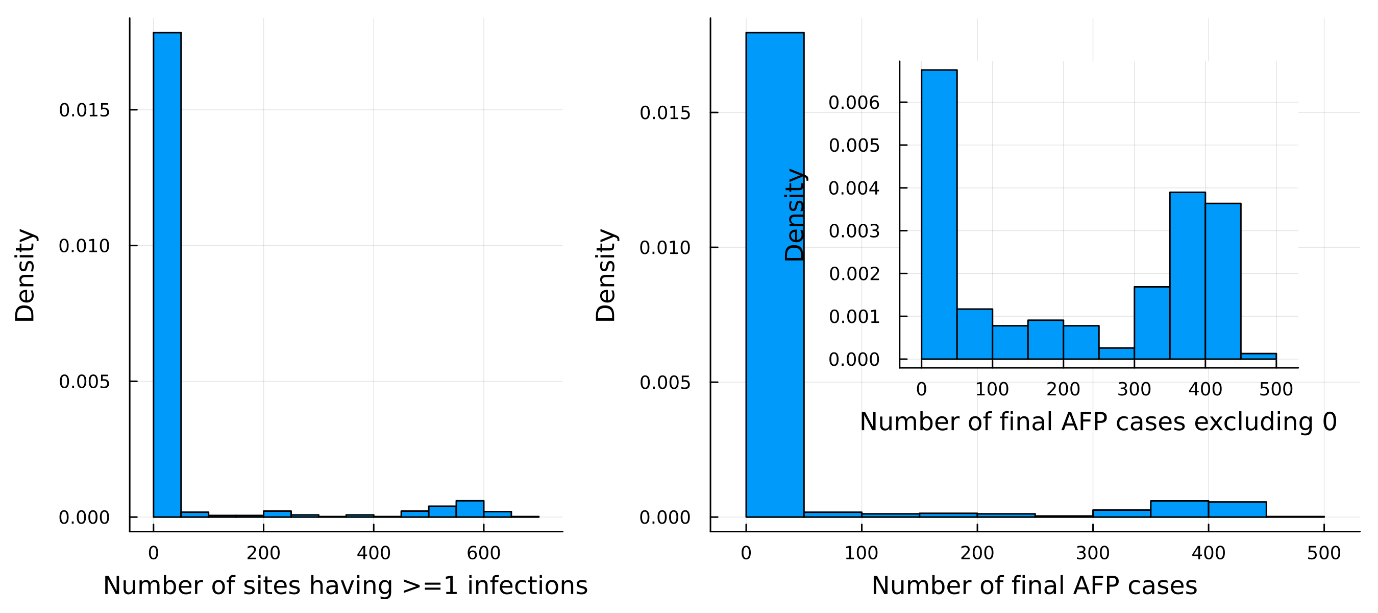
Description automatically generated

Figure r2. Example epidemic trajectories for the baseline settings.



* Figure r5. Spatial baseline results. Cumulative probability for both surveillance, and box plots of lead time for a single homogenous population. (Write the number of extinctions for each pattern.)

1000 simulations.



* Figure r11. Summary simulation characteristics of simulations. (A) Histogram of locations at least one infection. (B) Histogram of final size of AFP. (C) Histogram of final size of AFP excluding 0. (D) Weekly number of AFP over time.

## Sensitivity analysis

A group of graphs with numbers

Description automatically generated

A group of graphs with numbers

Description automatically generated

A group of graphs with different colored lines

Description automatically generated

Figure r6. Prob of polio detection (A), (C), (E) and Lead time of ES detection (B), (D), (F). We varied the following parameters: population coverage of catchment area for (A) and (B); sampling frequency for (C) and (D); environmental surveillance detection sensitivity (E) and (F). Np90 is given by -log(1-0.9)/g where g is the baseline hazard. **R0 = 14.0 Re0 = 1.13.**

A group of graphs with different colored lines

Description automatically generated

Figure r6\_2. Prob of polio detection (A), (C), (E) and Lead time of ES detection (B), (D), (F). We varied the following parameters: population coverage of catchment area for (A) and (B); sampling frequency for (C) and (D); environmental surveillance detection sensitivity (E) and (F). Np90 is given by -log(1-0.9)/g where g is the baseline hazard. **R0 = 13.0 Re0 = 1.05.**

A graph of a graph

Description automatically generated

Figure r8. Effect of the basic reproduction number (the effective reproduction number among unvaccinated individuals) on the prob of polio detection (A), (C), (E) and lead time of ES (B), (D), (F) for each parameter: population coverage of catchment area for (A) and (B); sampling frequency for (C) and (D); environmental surveillance detection sensitivity (E) and (F).

A collage of graphs

Description automatically generated

* Figure r9. Effect of the contact rates of individuals between different regions on the prob of polio detection (A), (C), (E) and lead time of ES (B), (D), (F) for each parameter: population coverage of catchment area for (A) and (B); sampling frequency for (C) and (D); environmental surveillance detection sensitivity (E) and (F).

TODO: ES early detect cases?

A graph of different sizes and sizes of different sizes

Description automatically generated

* Figure r13?. Precent of population coverage :10.995414. R0=13.

A graph of different sizes of blue dots

Description automatically generated

* Figure r14?. Precent of population coverage :50.20926. R0=13.

Table . Outbreaks

Table 2. Outbreaks of polio in outside African continent, Pakistan and Afghanistan.

Figure . Test Figure.

Outbreaks of polio in outside African continent, Pakistan and Afghanistan.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Discussion

* In order to effectively find a polio case, 40% coverage of ES catchment area in the order of populous area is required.
* ES performs better, but under the setting,
* Finding 1. There is a substantial variability of lead time simulations.
* Finding 2.

**About the detection probabilities**

[Avram Levy 2022](https://www-sciencedirect-com.ez.lshtm.ac.uk/science/article/pii/S0048969722053657?via%3Dihub) , [Hiroki Ando 2023](https://pubmed.ncbi.nlm.nih.gov/36867995/) for COVID-19 pandemic. TODO: check how these estimates are obtained (assuming the population catchment size?).

**Limitations**

- [Harriet L. Mills 2014](https://journals.plos.org/ploscompbiol/article/file?id=10.1371/journal.pcbi.1003561&type=printable) Spatial resolution for the meta-population model.   
- [Amy Wesolowski 2015](https://journals.plos.org/ploscompbiol/article/file?id=10.1371/journal.pcbi.1004267&type=printable) evaluates the radiation model, and the gravity model has poor power to predict the actual mobilisation.   
- [Olga E Hart 2020](https://pubmed.ncbi.nlm.nih.gov/32371231/) Temperature is not considered for the detection probability.

aaa

# Recommendations

Test.

The discussion should end with a paragraph linking the current findings with recommendations for further work. However, it may be appropriate to present the recommendations as a separate section. Your recommendations must follow from your findings and your analysis of them, and not simply be a list of unrelated ‘good ideas’

# Reference list

TODO: Zotero replacement.

1. Ranta, J., Hovi, T. & Arjas, E. Poliovirus Surveillance by Examining Sewage Water Specimens: Studies on Detection Probability Using Simulation Models. *Risk Anal.* **21**, 1087–1096 (2001).

2. O’Reilly, K. M. *et al.* Surveillance optimisation to detect poliovirus in the pre-eradication era: a modelling study of England and Wales. *Epidemiol. Infect.* **148**, e157 (2020).

# Appendices