**Background**

Schematics of the surveillance system, [Dominika A Kalkowska 2015](https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-015-0791-5).

Previous papers about the surveillance

* [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)
* [Ranta 2001](https://onlinelibrary-wiley-com.ez.lshtm.ac.uk/doi/pdf/10.1111/0272-4332.t01-1-216174)

The time to detection from genome paper. [Megan Auzenbergs 2023](https://gatesopenresearch.s3.eu-west-1.amazonaws.com/manuscripts/14920/7b375af8-4124-4905-b376-94a924dd9ce1_13272_-_megan_auzenbergs_v2.pdf)

**Model concept**

ES: Environmental surveillance, AFP: Acute flaccid paralysis.

**Basic model**

The objective of this study is to quantify the lead time of environmental surveillance detection. According to Ranta 2001’s study, the model is decomposed the three parts: the epidemic model, wastewater part, and the testing strategy.

TODO: What is means about the expert opinion in the literature?

The SIWR or EPISENS-M model is frequently used regarding the wastewater model. Here we adopted the extension of the SIWR model. However, [Andrew F. Brouwer 2019](https://www.pnas.org/doi/10.1073/pnas.1808798115) showed that the pathogen diffusion rate from swage is theoretically identifiable, but practically impossible to determine the parameter due to the large uncertainty. When we arbitrarily chose the large pathogen diffusion rate, the virus volume in wastewater is proportional to the infectious individuals, who are shedding the virus. Even though evidence about the delay in the detection of virus from the virus shedding in individuals is present, the objectives of this study are quantifying the lead time of ES compared to the AFP surveillance, and the influence of delayed detection from one individual would be negligible since the infectious population increases rapidly in the early phase, which period is our focus.

TODO: find a paper describing the delayed detection of poliovirus from a single infectee (not prolonged shedding periods).

* [W. J. Lodder 2012](https://www.pnas.org/doi/10.1073/pnas.1808798115): Human challenge paper about polio. The second peak was regarded as the secondary transmission around day 20.
* [Manasi Majumdar 2018](https://watermark.silverchair.com/jix667.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAsEwggK9BgkqhkiG9w0BBwagggKuMIICqgIBADCCAqMGCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMiiBr4f0itCp0UQ2KAgEQgIICdHB_0cXN4PLqyUQf-2iNluYZ9Dbc-jA-UR2-P_BeKV1zzbR9SVYjdxdx2RY-UjcDSz_rrpYYBPcFIpQUyAEdQnCciddZjhxbD1-G9U3e193aWuOvVTcR-MK48hCYi5TLfOZeI0-0QW6jRjbBo0w9no6pDyz9RQCzUPIL4lEZYgnGLUKgNtEqvRy7MXBQpN1FyPo-Cp9KQs6GV7KatOKnjYn4Ty_TyuJ3Tun5HiVnFXdD_o8cpBspSKU8iAkUF8ub3pwlGke4EkNEzbRDYYKC9OgXsi2fmInD1uhs6uUPPwCu6diDBhwVggcgpVEaHea04A6K1KYMuE04kfhleoYYcIXzTYHtxnFmM9cLKF7hCXy4zSzZWEA8QeErJbCc-cWfpRCIT_R7PCPkEWJdqqGpGPoUwWLq0CFYe0_nzPrVvbWDWnTWVJ1OiNcLtyeSMuP0QpMWLcxN-0NUXug-TtdgW_kWjA8hoWaEhCPuMZ7hELoZ088dJvAAyFkDTGzW47-vyWffsYCyYxnny8S5U_KTOc3piUceCC38WK0Jy5VlqsUSVAiKa7CayxP31SdhrEDcqFXcr1G-dqD-mHKYEH0G2jJTd6aYN15D6WQ5HK2mfG5CSsUKklOJiX1Kcp0ansk9y5BU920XsXvYh3qQhulOqPzmnEXnp08tDmWx0UpjS66bYWEVjj2FYpW5vt0Gxfrzml2PzYwnIrVYQeSTGuzdSlQ6psQjYj4F1IXf9Ukp8b2K-8j0NstDFinCOi7kpo_TYl6C9sRvxq5x8aU_fPKW3EEaEi4ZNsGIAZl1Qv799V_5ptU__rXOH5sHSvsVBIJ4TlFYtUk) Detection of OPV in sewage in the UK was due to the immigrants coming from the country where OPV was still used.

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.

A picture containing text, line, plot, diagram

Description automatically generated

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.

The obtained shape parameter for gamma distribution is almost 2, so I decide to use the simple SEIR model without multiple compartments for Infectious individuals.

**Testing quality and strategy**

Among the quality indicators for environmental surveillance, the number of sampling (150 stool specimens annually) and sampling timing are present. The number of samplings increases the testing sensitivity. The testing frequency is also a key to detecting a virus in an early stage. We also have to consider the accuracy of the identification of polioviruses among all isolates. Sewage is correctly working or not by investigating the samples containing other enteroviruses in sewage samples.

The testing quality is varied from location to location. Sometimes

TODO: How can we incorporate the number of samples and the timing of sampling?

**Other factors delay the detection of environmental surveillance.**

TODO: [Yifan Zhu 2021](https://pubmed.ncbi.nlm.nih.gov/33548842/) Early warning of COVID-19 via wastewater-based epidemiology.

**The diffusion process of the polio virus**

Candidate models are the gravity model, radiation model ([Filippo Simini 2012](https://www.nature.com/articles/nature10856)), and impedance model ([Kankoé Sallah 2017](https://ij-healthgeographics.biomedcentral.com/articles/10.1186/s12942-017-0115-7)).

Gravity model? Radiation model? Random network model?

We mainly use the radiation model since this model is shown to fit the mobilisation well, and parameter-free to estimate the regional difference.

The probability of commuting from i to j per day is given by

The probability of mobility α (including the travel volume per day and the duration of travelling, see the relationship between the travel distance and travel duration, [John R. Giles 2020](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7486699/pdf/pnas.201922663.pdf)).

The *P*i and *P*j denotes the population size at location i and j, respectively, where *S*ij is the total population in the circle of radius centred at i, but excluding the source and destination population.

Then the meta-population model is formulised as

Where *E*i represents the individuals with latent period at location *i*, and *I*k,j represents the infectious individuals at stage k (denoting a1, a2, and AFP) at location *j*.

For the gravity model, we are restricted to use that model due to data scarcity and need for parameter specification.

In polio situation, radiation model fits better than the gravity model ([Arend Voorman 2023](https://pubmed.ncbi.nlm.nih.gov/34483024/)).

Limitations: [Amy Wesolowski 2015](https://journals.plos.org/ploscompbiol/article/file?id=10.1371/journal.pcbi.1004267&type=printable) found that the both gravity model and the radiation model fail to predict the human mobility due to accessibility, availability of transport, and cost of travel between locations.

**Vaccination coverage.**

[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) District level OPV0, OPV1, Hexavalent1~4 vaccination coverage was written with the sample size of around 5 ~ 800. See page 48 – 54,

**Coverage of ES area**.

1. The first strategy is setting ES based on the large population size region.

2. Focus on the three city areas, for each changing the coverage area evenly.

A picture containing text, map, diagram, plot

Description automatically generated

Figure. South Africa male population under 5 with 10km resolution obtained from the [WorldPop](https://www.worldpop.org/) project.

A picture containing text, map, diagram, plot

Description automatically generated

Figure. South Africa male population under 5 with 20km resolution obtained from the [WorldPop](https://www.worldpop.org/) project.

A picture containing diagram, plot, line, screenshot

Description automatically generated

Figure. South Africa male population under 5 with 100km resolution obtained from the [WorldPop](https://www.worldpop.org/) project.

A picture containing text, diagram, screenshot, line

Description automatically generated

**Summarised statistics**

We used the cumulative probability of detecting the first polio cases through AFP surveillance of ES over time. Among *N* 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.

All the code for simulations and visualisations are implemented in Julia v1.8.3.

------------------------------------------------------------------------------------------------------------------------------------

Rough writing zone

Objectives

* Calculate the lead time
* **Calculate the lead time varying the catchment area of ES**
* Sensitivity analysis. Calculate the results, varying the assumptions.

Results (or Appendices results)

* Table. Summary of outbreak situations and ES implemented areas.
* Figure. Single model cumulative probability for AFP, ES (overall and conditional), and Box plots.
  + Table. Prop. of both, AFP only, ES only. Also, lead time (mean, min, 25th, 50th , 75th, max).
* Figure. Sample epidemic curve for one simulation sets.
  + ? Figure. Sample epidemic curve for various assumptions?
* Figure. Meta-population model cumulative probability for AFP, ES and box plots for the lead time.
* Figure. Conditional probability, varying the catchment area for ES (drawing AFP as baseline.), boxplots for the lead time.
* Lead time, sensitivity analysis

**Model specification**

A picture containing diagram, line, font, text

Description automatically generated

For the environmental surveillance. Amount of viral shedding (hazard function) is expressed as *λ*0(Ia,+IAFP). The probability of detecting virus per one sample, ωt, is represented as

For each time, we take *n*t (0 or 1) and let test sensitivity *P*ES,test for one collected samples be 0.95, and the number of polio-positive sample, wt is given by.

**AFP surveillance**

Among the new infections, γ2Ia, to be diagnosed with the poliovirus, we have to consider the probability of seeking healthcare, stool sampling, and sensitivity of stool sampling. Each probability is denoted as the *P*H, *P*AFP,sample,*P*AFP,test . Then, the probability of AFP case detected as polio case is modelled as

**Parameter specification**

|  |  |  |
| --- | --- | --- |
| Parameters | Values | Ref. |
| Basic reproduction number (R0) | 10 | [Kimberly M. Thompson 2020](https://onlinelibrary.wiley.com/doi/pdf/10.1111/risa.13484) |
| Population size (N),  should be under 15 years old | 10,000 | WorldPop |
| Latent period (γ1) | 4 days | [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) | 15.02 days | Fitted to [Radboud J. Duintjer Tebbens 2013](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7890644/pdf/nihms-1670167.pdf) |
| Paralysis-to-infection ratio (pAFP) | 1/200 for WPV1, | [Neal Nathanson 2010](https://academic.oup.com/aje/article/172/11/1213/194806?login=false) |
| 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 |
| 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) |

**Spatio-temporal SEIR model incorporating the radiation model.**

A picture containing diagram, line, font, plan

Description automatically generated

Differential equation is written as

Where λi is given by

**ES surveillance**

Viral shedding from the population can be approximated by the number of infectious individuals, *I*a,i,+*I*AFP,i ([Andrew F. Brouwer 2019](https://www.pnas.org/doi/10.1073/pnas.1808798115)). The probability of detecting virus per one sample at location i at time t is modelled as

*g* is the scaling parameter. Let *n*t,i ­be the number of sampling at location i at time t (taking only 0 or 1), and *P*ES,test be the test sensitivity for one collected samples (0.95). The number of polio-positive sample, *w*t,i is given by

It is the sampling location could be unmatched with the compartment model version of the sampling site. It indicates that if we increase the spatial resolution, it automatically increases the number of samples corresponding to each site. Further consideration is required for this points.

**AFP surveillance**

Among the new infections, γ2Ia, to be diagnosed with the poliovirus, we must consider the probability of seeking healthcare, stool sampling, and sensitivity of stool sampling. Each probability is denoted as the *P*H, *P*AFP,sample,*P*AFP,test . Then, the probability of AFP case detected as polio case is modelled as

Since the probability is the same across the location, and this sampling process follows the binomial distribution (sequence of Bernoulli trials), this can be rewritten as

**ES surveillance, Parameter relationships**

A picture containing text, line, plot, diagram

Description automatically generated

A picture containing text, diagram, screenshot, line

Description automatically generated

**Temporal results for single compartment model.**

I assumed monthly environmental sampling. (the initial date of sampling is randomly chosen from 1 to 30).

is set as maximum ω(t) become 0.9 when 10 infection occurs. (Results would not be changed if this value is set to be 0.8)

The below figure is produced with R0=2

A picture containing diagram, line, plot, text

Description automatically generated

**Table and Figure for the lead time of Environmental surveillance.** Take a difference of the date of AFP surveillance. Lead time more than 0 (early detection in environmental surveillance) is 85% among 1000 simulations.

A picture containing screenshot, diagram, line, text

Description automatically generated

A picture containing text, screenshot, font, number

Description automatically generated