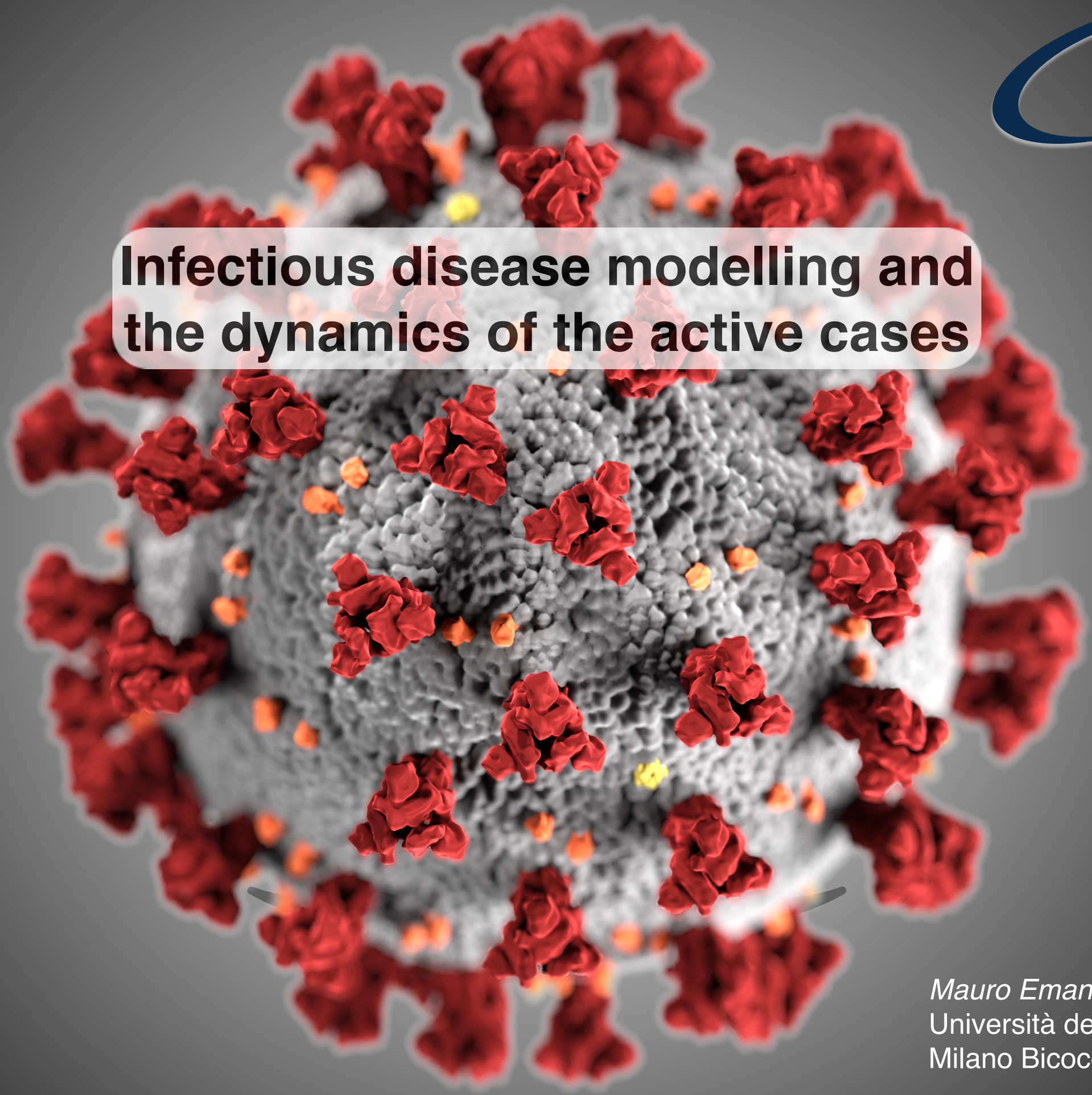
Infectious disease modelling and the dynamics of the active cases



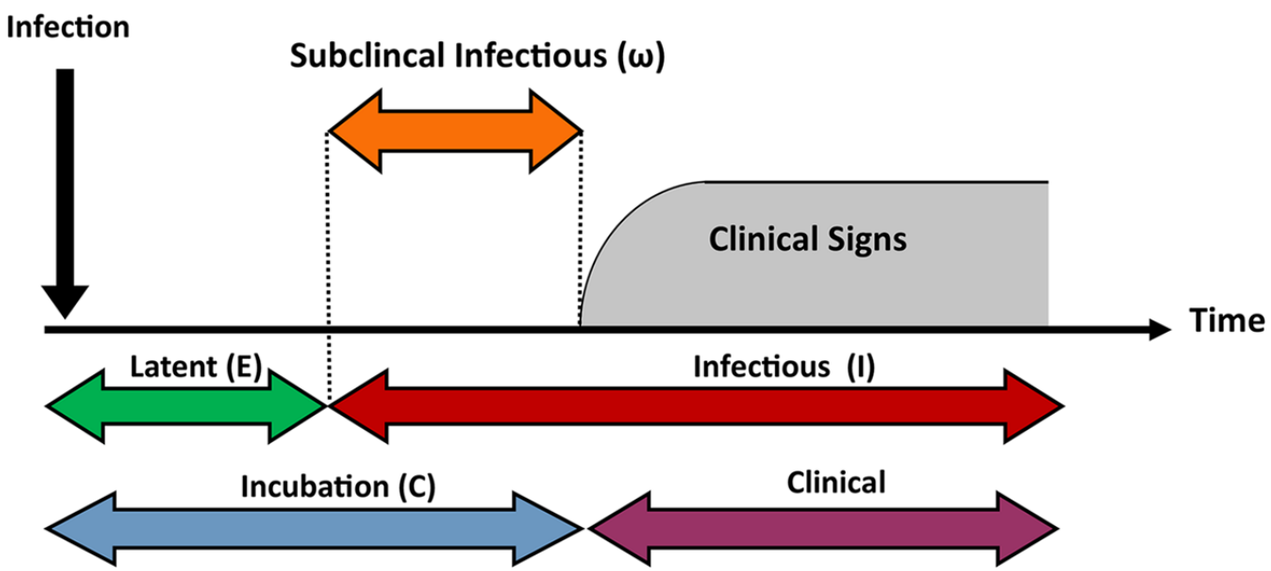
Mauro Emanuele Dinardo
Università degli Studi di Milano Bicocca and INFN

Introduction

Goal: we want to describe the dynamics of the spread of a disease among a population, in particular we are interested in the development of infected active cases

Basic assumptions:

- the disease can spread to other individuals from the infected ones
- the infected individuals can recover (either survive or die)
- recovered individuals can not become re-infected
- both symptomatic and asymptomatic cases are considered
- the system is closed (no “natural” demographic change)



Definitions used to characterise distinct periods of infectious diseases. The latent period (**E; green arrow**) begins at the time of infection and ends at the onset of the infectious period. (**I; red arrow**). The incubation period (**C; blue arrow**) starts at the time of infection and ends at the appearance of clinical signs of disease (**purple arrow**). The difference between incubation and latency is denoted by ω (omega), and can have either a positive or negative value depending on whether transition to the infectious period occurs before or after appearance of clinical signs. The ratio of transmission occurring during incubation, and transmission occurring through the total infectious period is denoted by θ (theta; not shown)

The model

The model describe the dynamics of the number of infected active cases, i.e. the total number of currently infected individuals as a function of time

Finite difference model of infected population growth:

$$N_{n+1} = N_n + \underbrace{N_n\ K\ p\ q}_{\textit{Increment}} - \underbrace{N_n\ r}_{\textit{Decrement}}$$

Definition of the different terms:

- N_n: number of active cases at time n

- K: average number of encountered people by one infected individual. Another way to interpret K is: $K = (1-Q) \times k$, where Q is the fraction of active cases which are in quarantine and k is the average number of encountered people by one infected individual

- p: probability to spread the disease from one individual to another

The model

- **q**: probability that that the new individual doesn't have already the antibodies. This probability is equal to **1** minus the probability of being infected, which can be derived from $\mathbf{N_n}$ and it's equal to:

$$\frac{N_n + r\sum_{i=0}^{n-1} N_i}{C}$$, where **C** is the carrying capacity

- **r**: recovery rate (either to survive or die) ➜ related to the <u>average recovery time</u> through the relation $\tau = 1/r$

Final formulation of the model:

$$N_{n+1} = N_n + N_n Kp\left(1 - \frac{N_n + r\sum_{i=0}^{n-1} N_i}{C}\right) - N_n r$$

The <u>basic reproduction number</u>, $\mathbf{R_0}$, and the <u>effective reproduction number</u>, $\mathbf{R_n}$ at time n, as referred in epidemiology, are $R_0 = Kp/r$ and $R_n = Kp\left(1 - \frac{N_n + r\sum_{i=0}^{n-1} N_i}{C}\right)\Big/ r$, respectively

The model

From the finite difference equation we can construct the differential equation:

$$\frac{dN(t)}{dt} = N(t)Kp\left(1 - \frac{N(t) + r\int_0^t N(t')dt'}{C}\right) - N(t)r$$

The integral-differential equation can not be solved in a closed form ➜ it has to be treated numerically (**Kp**, i.e. the growth rate will be denoted **g**)

Refining the carrying capacity

The carrying capacity, **C**, does not correspond to the total population of the system, **T**, but it's rather a subset that should evolve together with the number of infected people. In fact if the disease spreads out from a certain region, the probability that the individual encountered by an infected one doesn't have already the antibodies, **q**, can not be computed as the ratio of the total number of non infected individuals over **T**, because the encountered individual is not extracted uniformly from **T**, but rather from a subset which is "close" to the subset of infected individuals

The model

Final system of equations:

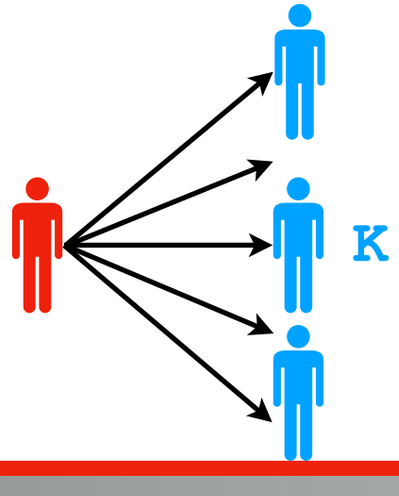
$$\begin{cases} N_{n+1} = N_n + N_n K p \left(1 - \dfrac{N_n + r\sum_{i=0}^{n-1} N_i}{C_n}\right) - N_n r \\ C_{n+1} = C_n + \left(N_{n+1} - N_n(1-r)\right) K \left(1 - \dfrac{C_n}{T}\right) \end{cases}$$

Differential form:

$$\begin{cases} \dfrac{dN(t)}{dt} = N(t) K p \left(1 - \dfrac{N(t) + r\int_0^t N(t')dt'}{C(t)}\right) - N(t) r \\ \dfrac{dC(t)}{dt} = \left(\dfrac{dN(t)}{dt} + N(t) r\right) K \left(1 - \dfrac{C(t)}{T}\right) \end{cases}$$

Which derives from:

$C_{n+1} = C_n + N_n K p \left(1 - \dfrac{N_n + r\sum_{i=0}^{n-1} N_i}{C_n}\right) K \left(1 - \dfrac{C_n}{T}\right)$, in turns the carrying capacity is the total number of encountered individuals by the infected ones, in fact the increment of the carrying capacity depends upon the new infected cases, N_n K p q, each of which encounter K number of individuals, multiplied by the probability of not being already accounted for



N.B. the term Kp is estimated from a fit to the data, while p is estimated by scanning the χ^2/d.o.f. and choosing the average best over multiple periods

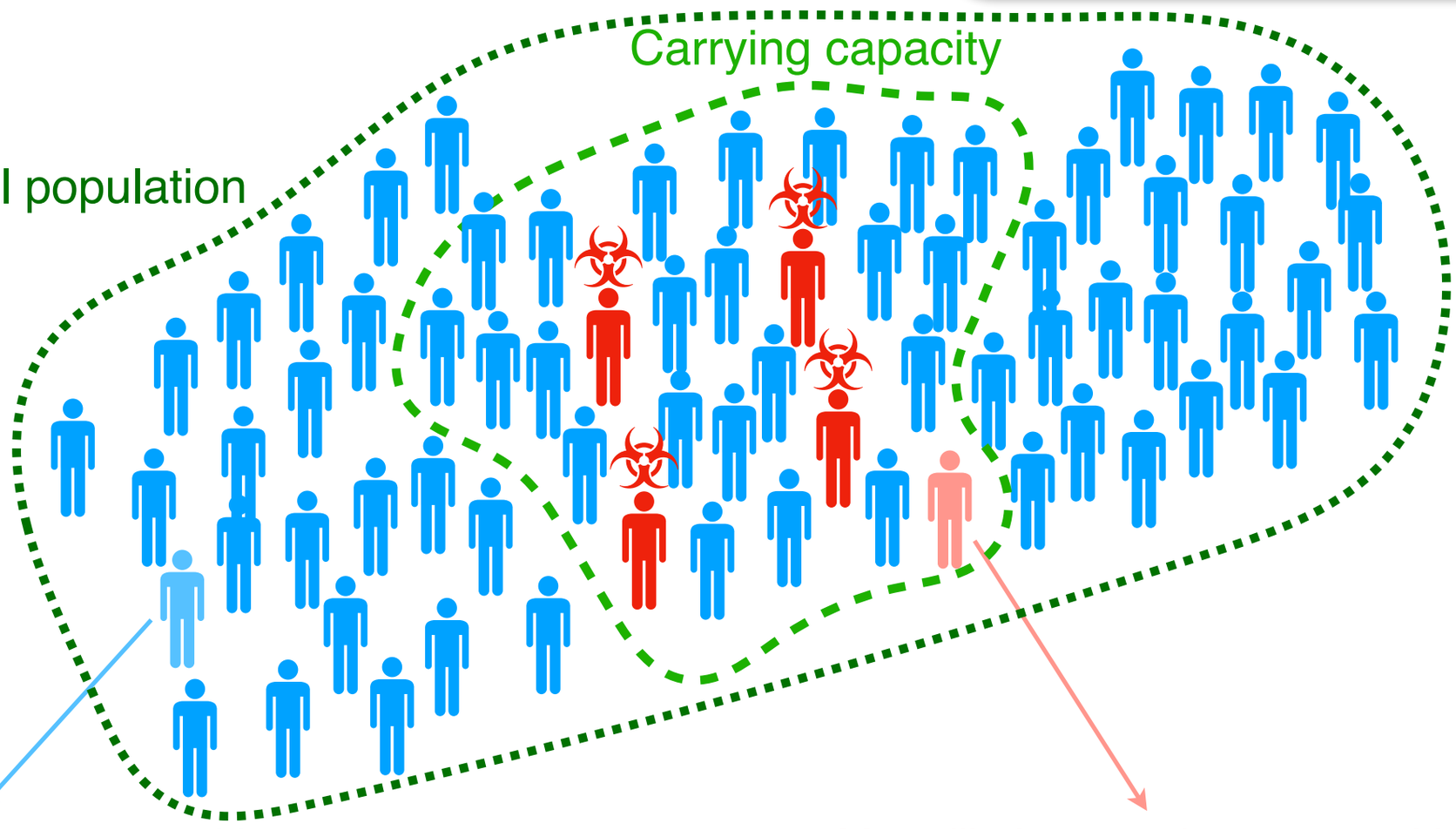
The model

Final system of equations:

$$\begin{cases} N_{n+1} = N_n + N_n K p \left(1 - \dfrac{N_n + r\sum_{i=0}^{n-1} N_i}{C_n}\right) - N_n r \\ C_{n+1} = C_n + \left(N_{n+1} - N_n(1-r)\right) K \left(1 - \dfrac{C_n}{T}\right) \end{cases}$$

Herd immunity established at `R = 1`:

- `Kpq/r = 1` ➜ `q = r/(Kp)`
- `%infected = 100•(1 - q)` ➜ `100•(1 - r/(Kp))`



Unlikely to get in contact with an infected individual

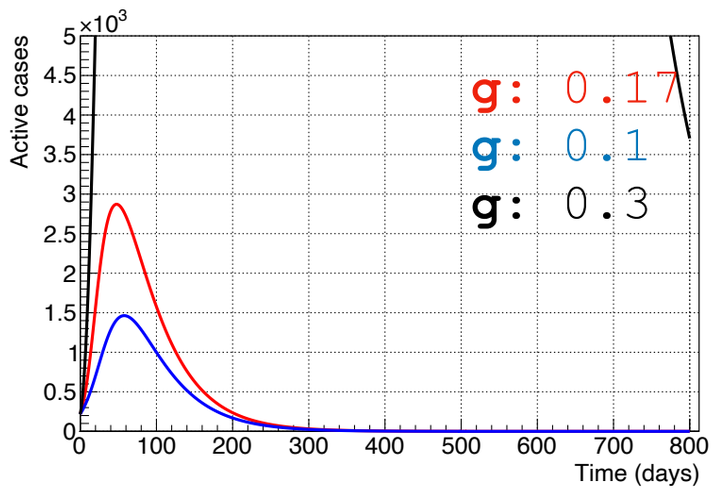
More likely to get in contact with an infected individual

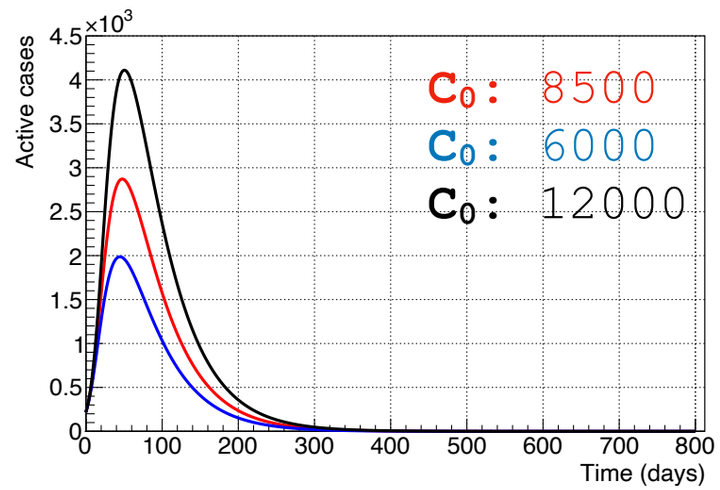
The model

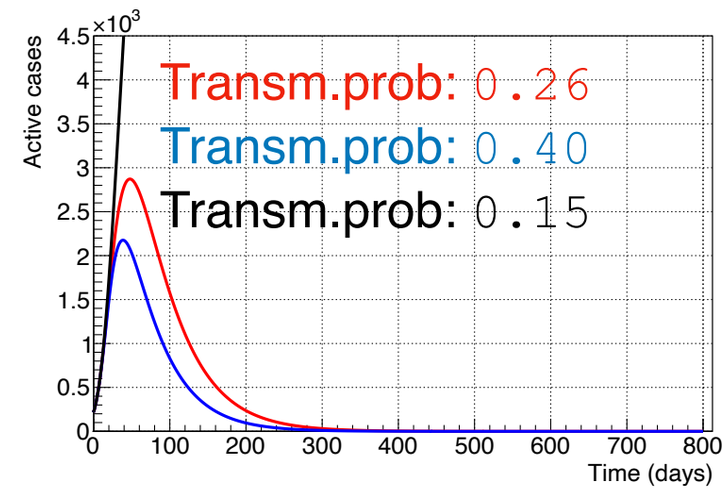
Impact of the different parameters on the model

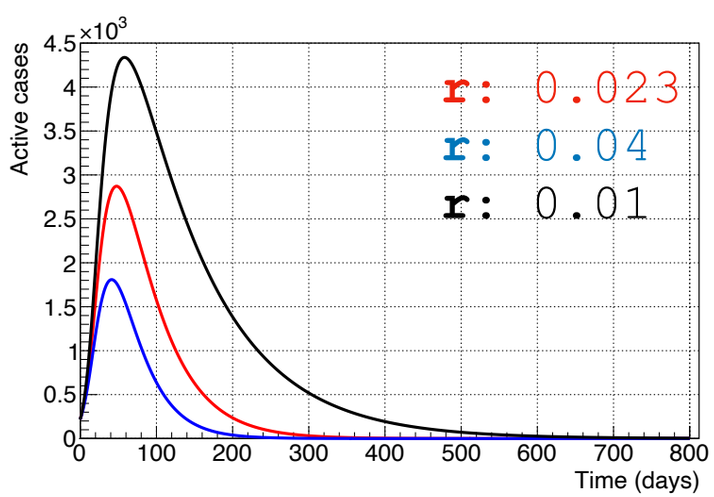
Default model parameter values:

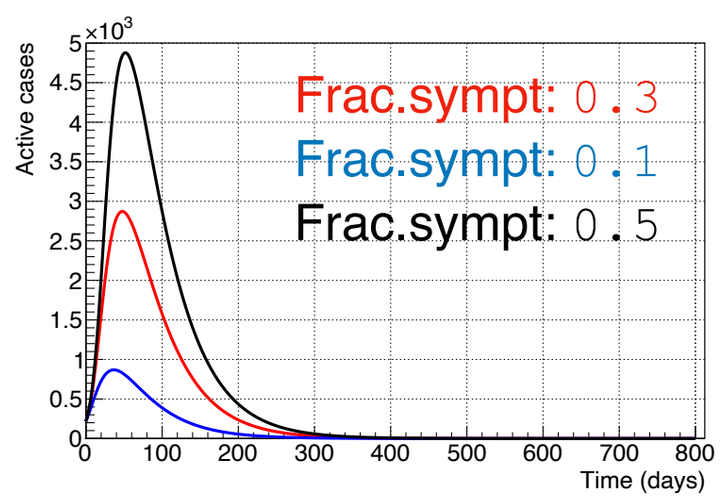
- **N_0** (initial population): 220
- **g:** 0.17
- **C_0:** 8500
- **r:** 0.023
- **Total population:** 60×10^6
- **Frac.sympt:** 0.3
- **Transmission prob.:** 0.26

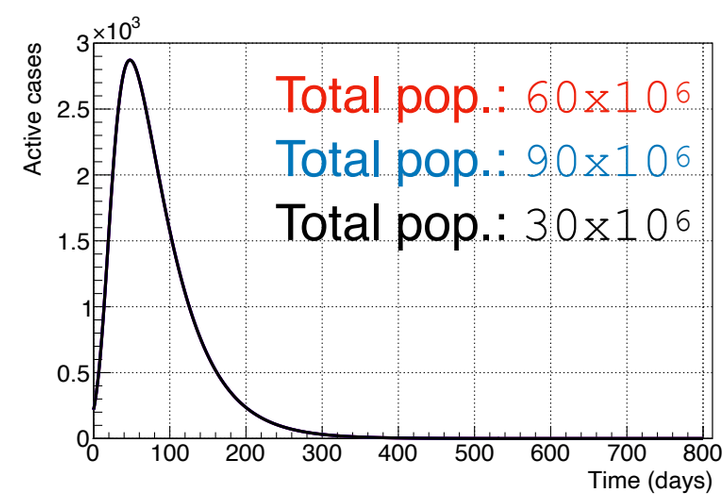












The model

The model is very similar to the S.I.R. model (**S**usceptible, **I**nfected, **R**ecovered from `Kermack` and `McKendrick`, `1927`), in fact:

$$\begin{cases} \frac{dS}{dt} = -\beta \frac{IS}{N} \\ \frac{dI}{dt} = \beta \frac{IS}{N} - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases}$$

We can identify `β` with `Kp`, `γ` with `r`, `N` with `C`, `I(t)` with `N(t)`, and `S(t)` with

$$C - \left(N(t) + r \int_0^t N(t')dt' \right)$$

The key difference is that the model takes into account the non uniform spread of the disease with the equation:

$$C_{n+1} = C_n + \left(N_{n+1} - N_n(1-r) \right) K \left(1 - \frac{C_n}{T} \right)$$

Probability of being infected

Derivation of the coefficient **q** for the model of infected active cases

- The term **q** is equal to **1** minus the probability of being infected
- The probability of being infected can be derived from $\mathbf{N_n}$ and it's equal to the sum of the terms $N_i\tilde{K}$ divided by **C** (**K-tilde** is just a short writing for the product **K p q**)

$$
\begin{aligned}
&N_0 \\
&N_1 = N_0 + \boxed{N_0\tilde{K}} - rN_0 \\
&N_2 = N_1 + \boxed{N_1\tilde{K}} - rN_1 \\
&N_3 = N_2 + \boxed{N_2\tilde{K}} - rN_2 \\
&\ldots
\end{aligned}
\quad \Rightarrow \quad
\begin{aligned}
&N_0 \\
&\boxed{N_0\tilde{K}} = N_1 - N_0 + rN_0 \\
&\boxed{N_1\tilde{K}} = N_2 - N_1 + rN_1 \\
&\boxed{N_2\tilde{K}} = N_3 - N_2 + rN_2 \\
&\ldots
\end{aligned}
$$

Summing up all the terms and simplifying, we obtain:

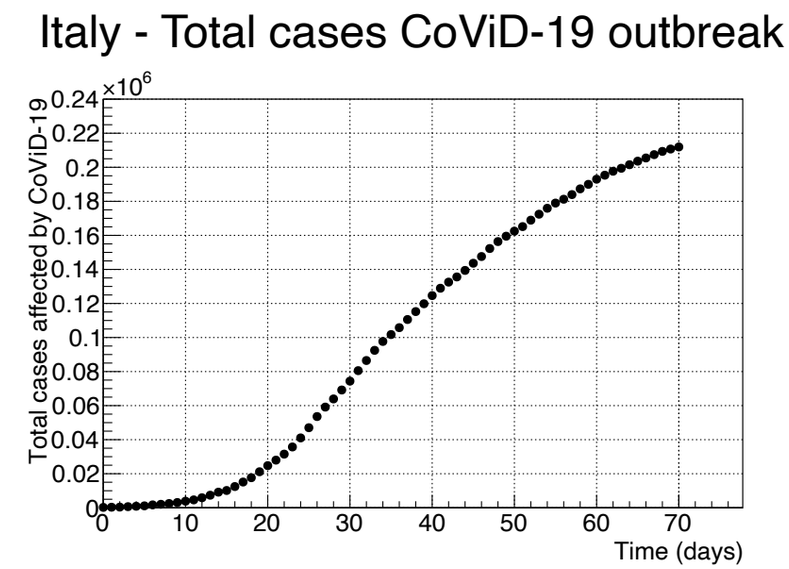
$$q = 1 - \frac{N_n + r\sum_{i=0}^{n-1} N_i}{C}$$

Terminology

Just for completeness we would like to relate the equation to typical plots shown in epidemiology

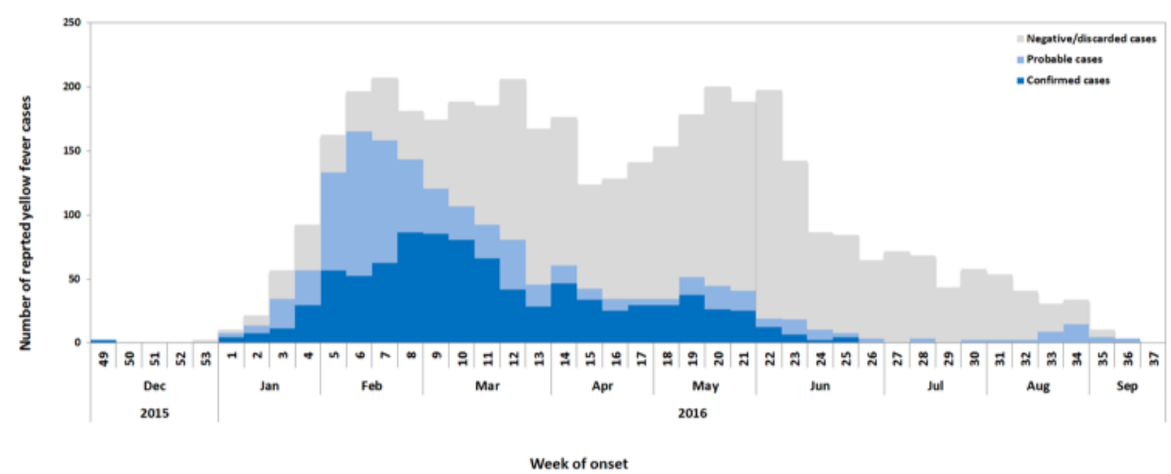
$$N_{n+1} = N_n + N_n K p q - N_n r$$

- The sum $\sum_n N_n K p q$ represent the total cases, i.e. the cumulated number of infected case as a function of time



- The derivative of the total cases, i.e. the term $N_n K p q$, represents the so called epidemic curve, i.e. the curve that shows the number of cases at the time of disease onset

- The term N_n itself instead represents the so called active cases, i.e. the total number of currently infected individuals as a function of time



Data source: Data as of 15 September 2016. Data for the past four weeks are subject to revision pending ongoing investigation and reclassification.

Model parameters

Parameter name	Symbol	Derivation
Recovery rate	**r**	Data plot Δrecovered/active cases
Growth rate	**g**, i.e. **KP**	Fit to the data
Transmission probability	**p**	Average best χ^2/d.o.f. on multiple periods
Fraction of symptomatic	N.A.	Obtained from literature
Total population	**T**	Demography
Initial conditions	N_0 and C_0	Fit to the data

Summary

- Three free floating parameters during the fit (**g**, N_0, C_0)
- One parameter is determined by χ^2/d.o.f. scan over multiple periods (**p**)
- One parameter determined from data plot (**r**)
- Two parameters determined from literature (**T**, fraction of symptomatic)

Application to data: **Italy - 3/5/2020**

We will present results obtained from different refinements of the analysis:

1. Fit to single, self-consistent, time periods
2. Fit to multiple time-periods simultaneously
3. Fit to multiple time-periods, considering also data smearing

History CoViD-19 Italy 2020

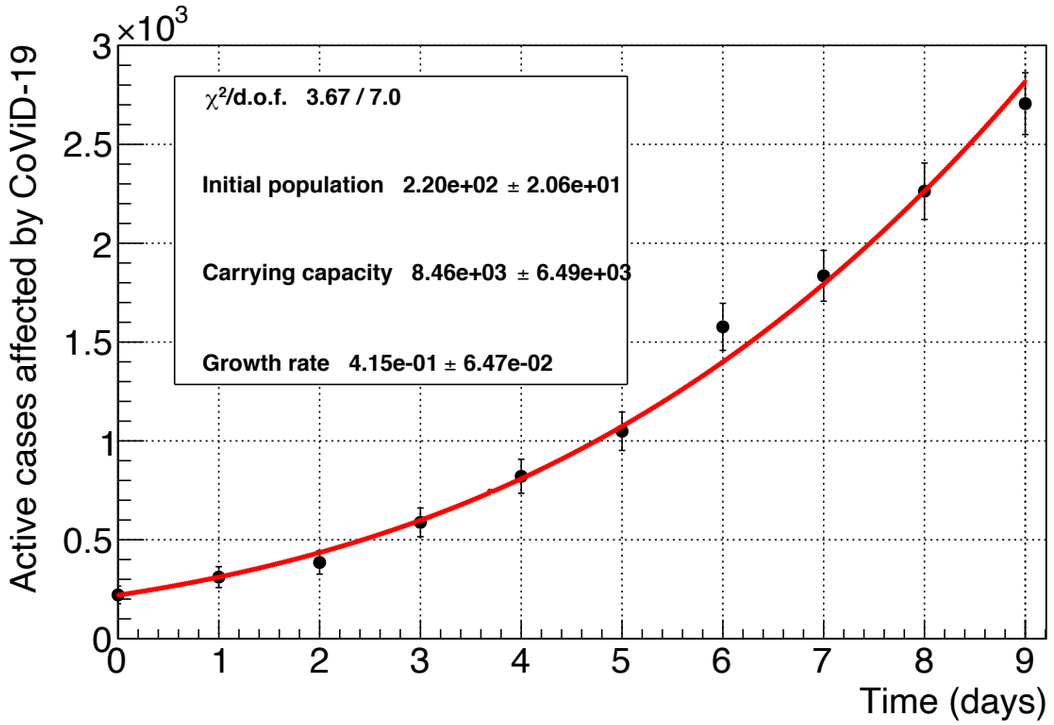
- Friday 21st February ➜ Codogno become red zone
- Monday 24th February ➜ Data recording starts
- Wednesday 4th March ➜ Lock down schools and universities
- Sunday 8th March ➜ Lombardy isolated
- Tuesday 10th March ➜ Restrictions on movements
- Monday 23rd March ➜ Total lock down
- Monday 4th May ➜ Release lock down and entering in the so called phase-2
- Monday 15th June ➜ Start of the so called phase-3 (free movements with masks)

- Start period **1**
- Start period **2**
- Start period **3**
- Start period **4**

Some references:

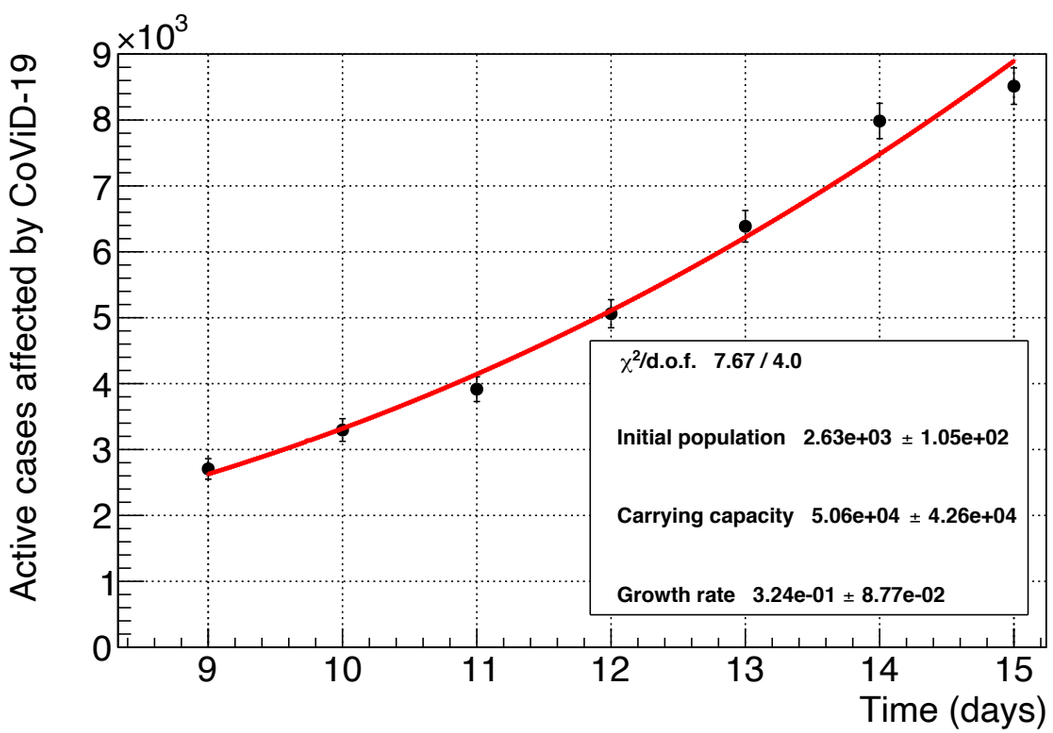
- Fraction of asymptomatic:
  - ★ https://www.repubblica.it/salute/medicina-e-ricerca/2020/03/16/news/coronavirus_studio_il_50-75_dei_casi_a_vo_sono_asintomatici_e_molto_contagiosi-251474302/
- Latent/incubation period:
  - ★ https://www.internazionale.it/notizie/graham-lawton/2020/03/26/sintomatici-diffondono-virus
  - ★ https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30746-7/fulltext

Application to data: fit to single time period



Number of <u>active</u> cases, fit data from 0th to 8th day (<u>period 1</u>):

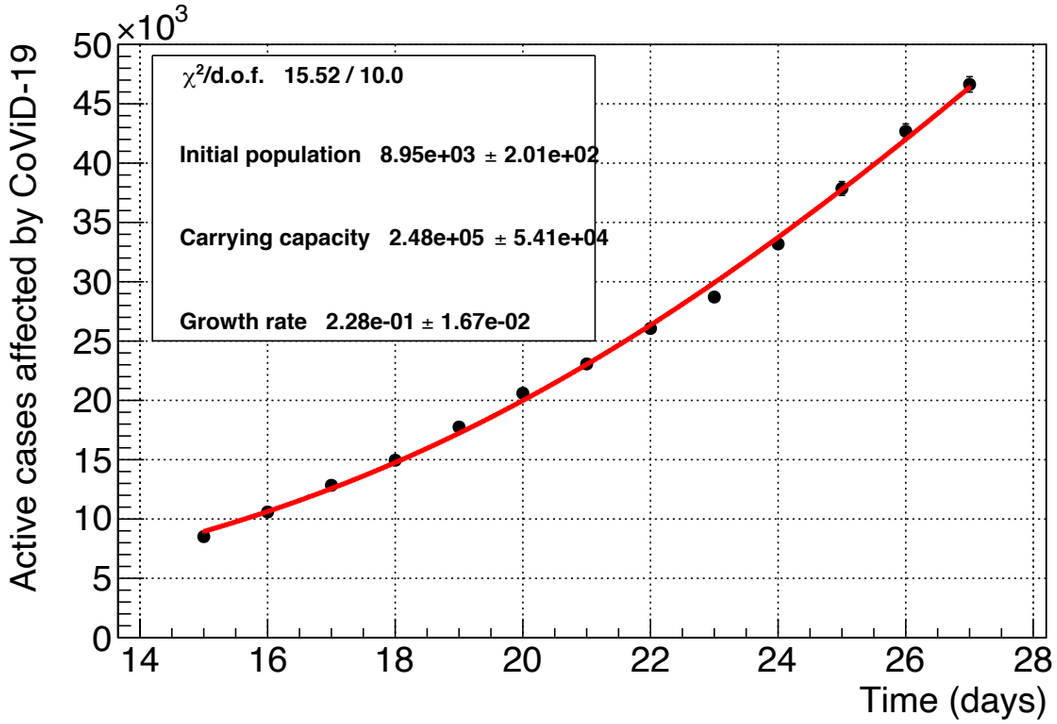
- N_0: 220
- g: 0.415
- C_0: 8460±6490
- N_9: 2817
- C_9: 23538±18057



Number of <u>active</u> cases, fit data from 9th to 14th day (<u>period 2</u>):

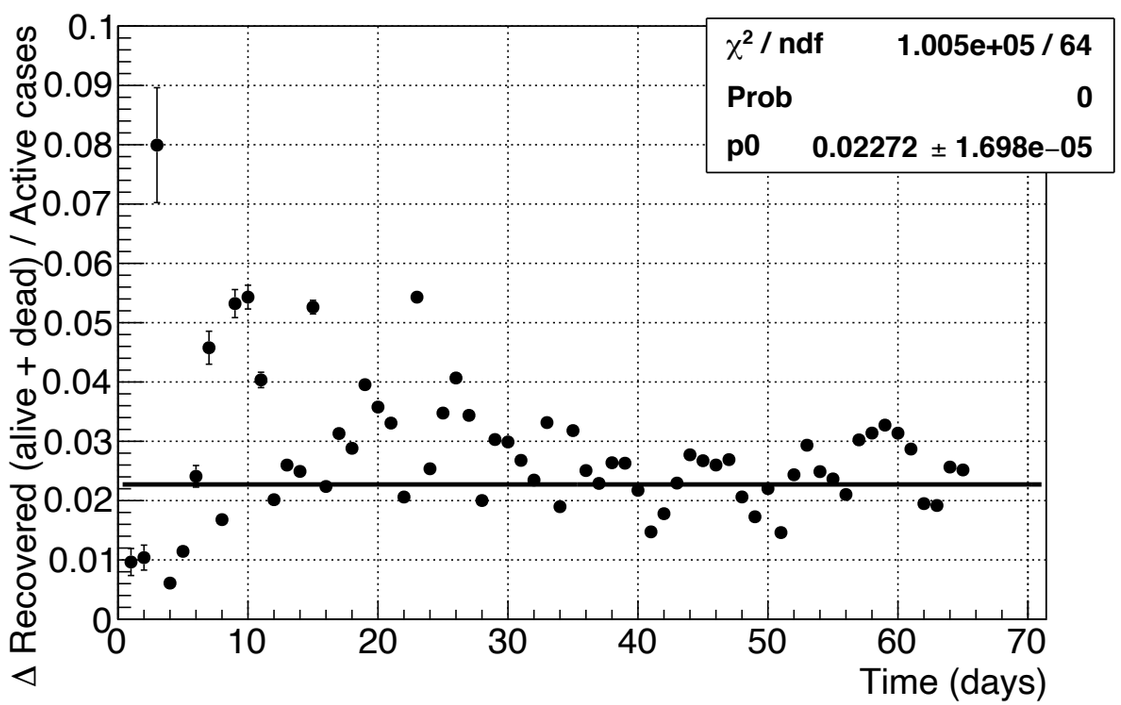
- N_0: 2630
- g: 0.324
- C_0: 50600±42600
- N_{15}: 8895
- C_{15}: 79587±67004

Application to data: fit to single time period



Number of active cases, fit data from 15th to 26th day (period 3):

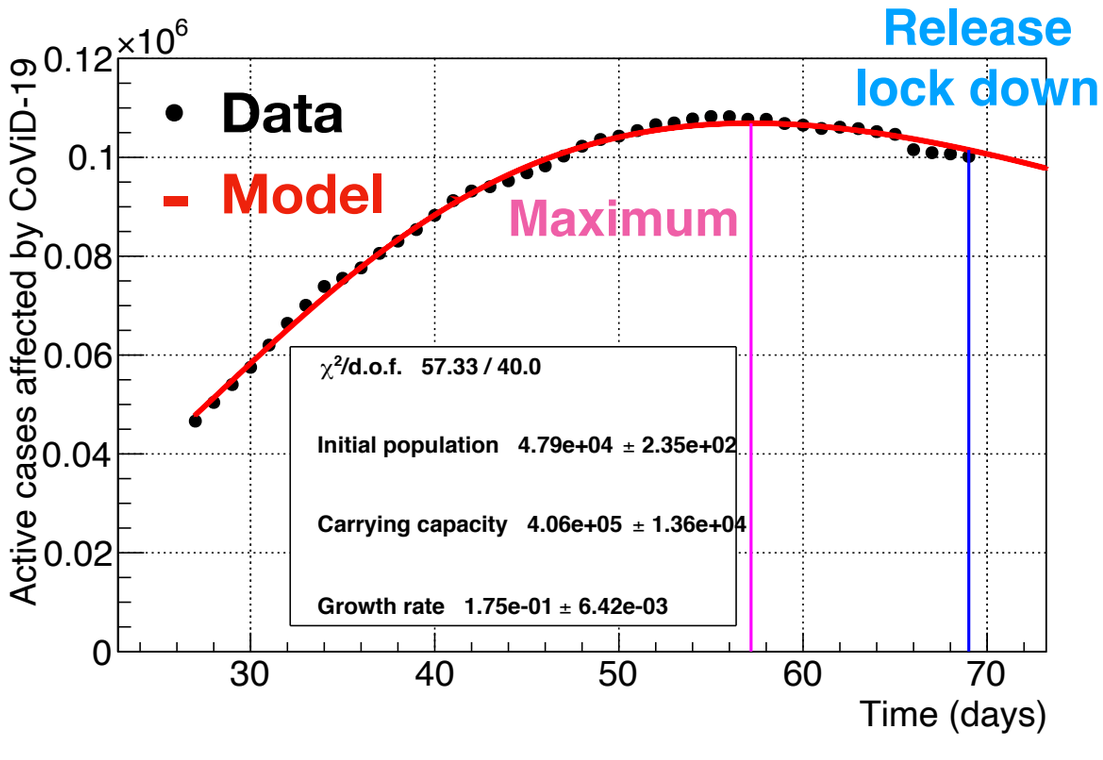
- N_0: 8950
- g: 0.228
- C_0: 248000±54100
- N_{27}: 46390
- C_{27}: 376923±82224

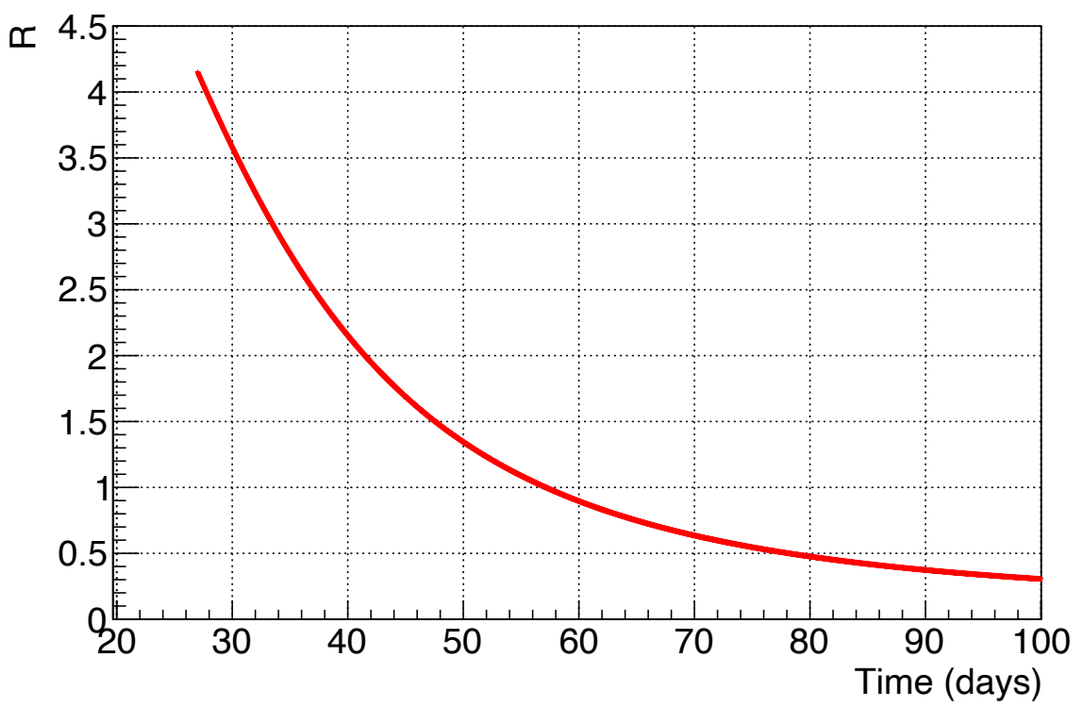


Recovery rate (r) is derived from data:

- r ~ 0.023 ➜ τ ~ 44 days

Application to data: fit to single time period





Number of active cases, fit data from 27th day till release (period 4)

For all fit:

- Errors rescaled by x3

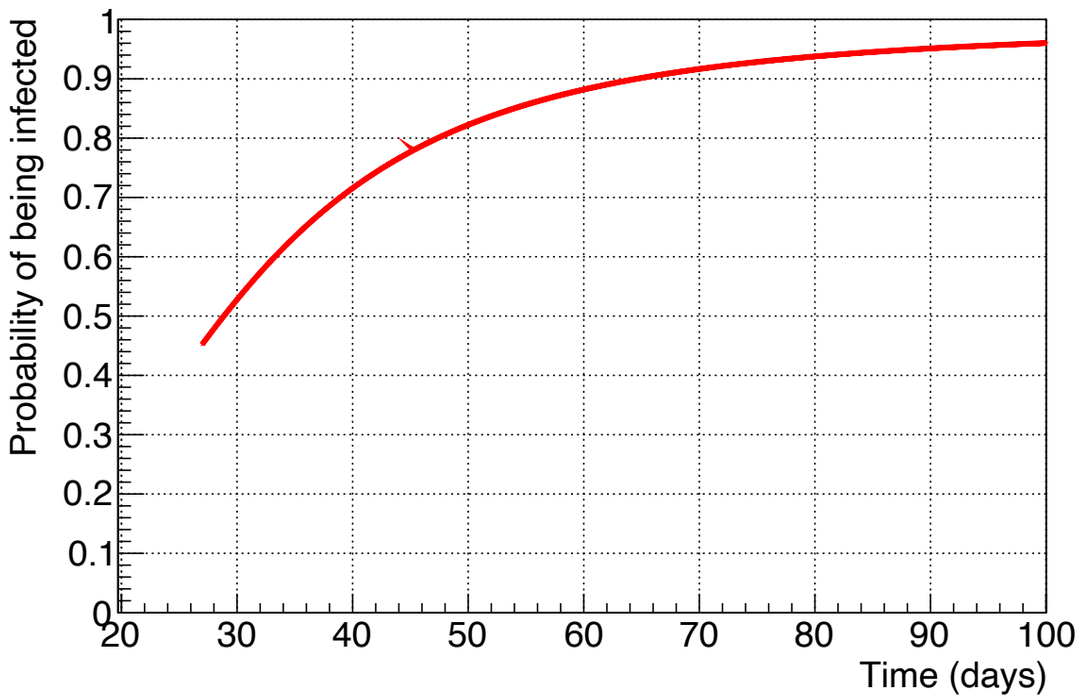
Fitted parameters:

- Initial population (N_0)
- Growth rate (**g**)
- Carrying capacity (C_0)

Fixed parameters:

- Recovery rate (**r** = 0.023)
- Fraction of symptomatic (**0.3**)
- Transmission probability (**p** = 0.26 - see control plots)
- Total population (**T** = 60×10^6)

Application to data: fit to single time period



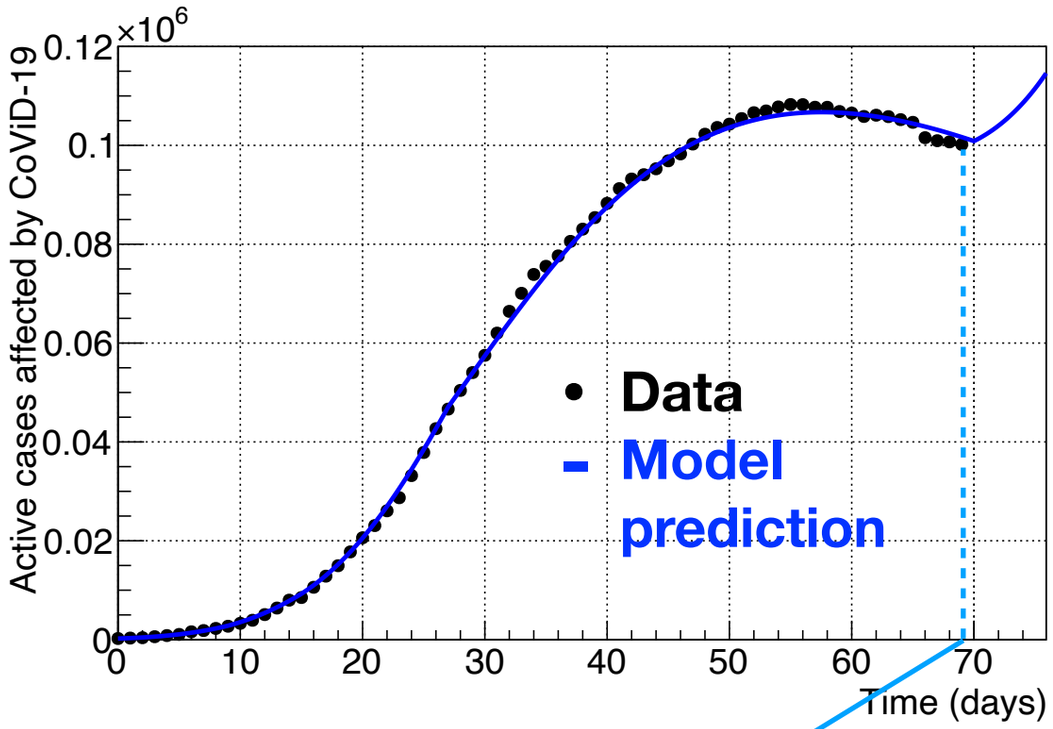
Probability of being infected, i.e. **1-q**

Parameter summary last period:

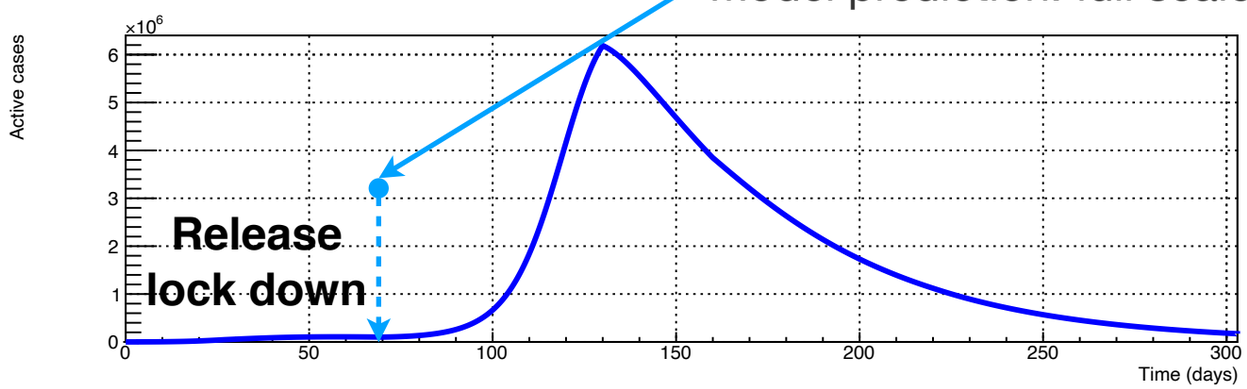
- **r = 0.023** ➜ plot Δrecovered/active
- **g = 0.175** ➜ fit
- **p = 0.26** ➜ χ^2/d.o.f. scan
- Symptomatic fraction **= 0.3** ➜ literature
- **T = 60x10^6** ➜ demography

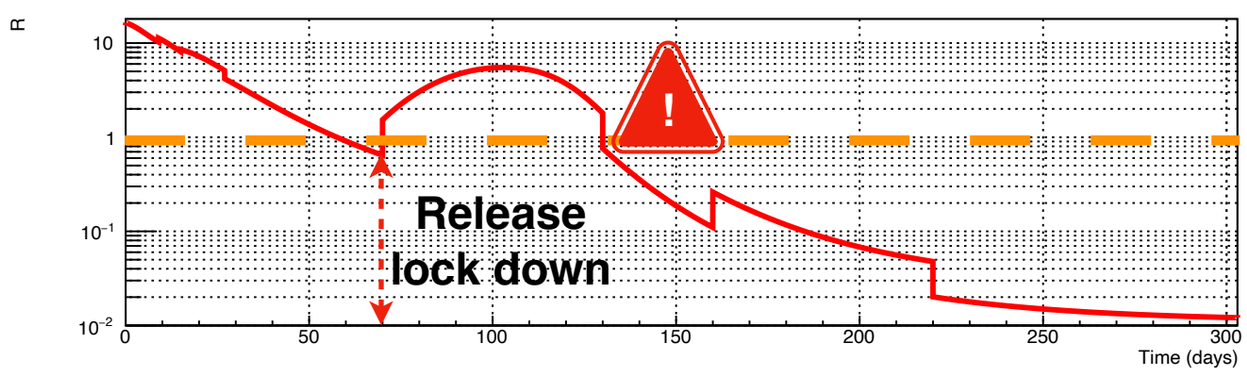
Nuisance parameters: **N_0** and **C_0**

Predictions: first scenario



Model prediction: full scale





- Normal lifestyle (`g = 0.415`) for 9 days
- Mild soc.distance (`g = 0.324`) for 6 days
- Medium soc.distance (`g = 0.228`) for 12 days
- Hard soc.distance (`g = 0.175`) for **42** days

Past Future

- Normal lifestyle (`g = 0.415`) for 60 days
- Hard soc.distance (`g = 0.175`) for 30 days
- Normal lifestyle (`g = 0.415`) for 60 days
- Hard soc.distance (`g = 0.175`) for 120 days

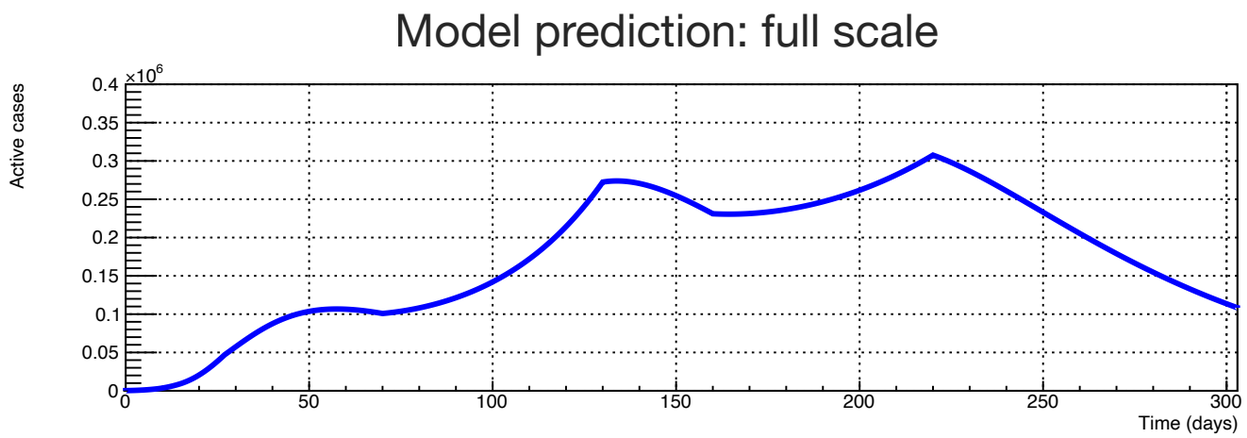
Other parameters:

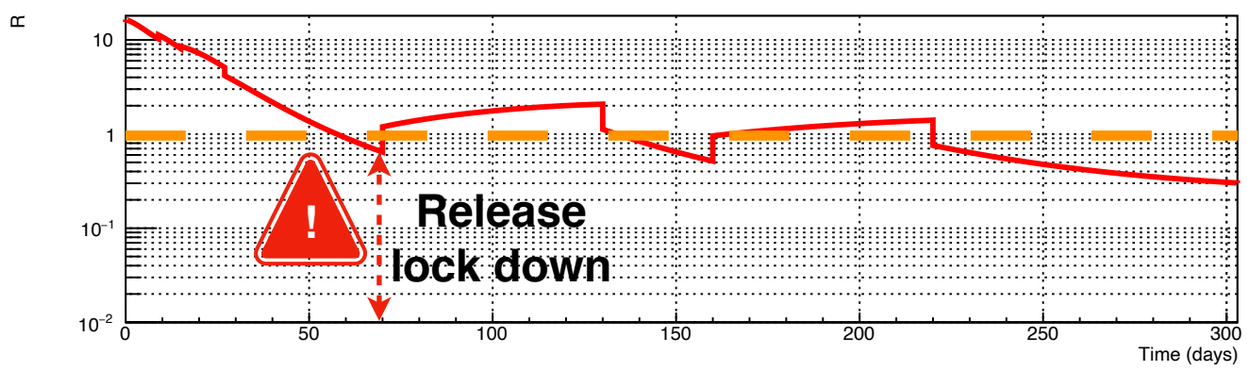
- N_0: `220`
- `r`: `0.023`
- Symptomatic fraction: `0.3`
- Transmission probability: `0.26`

Herd immunity:

- Normal lifestyle (`g = 0.415`): **94%**
- Mild soc.distance (`g = 0.324`): **93%**
- Medium soc.distance (`g = 0.228`): **90%**
- Hard soc.distance (`g = 0.175`): **87%**

Predictions: second and third scenarios

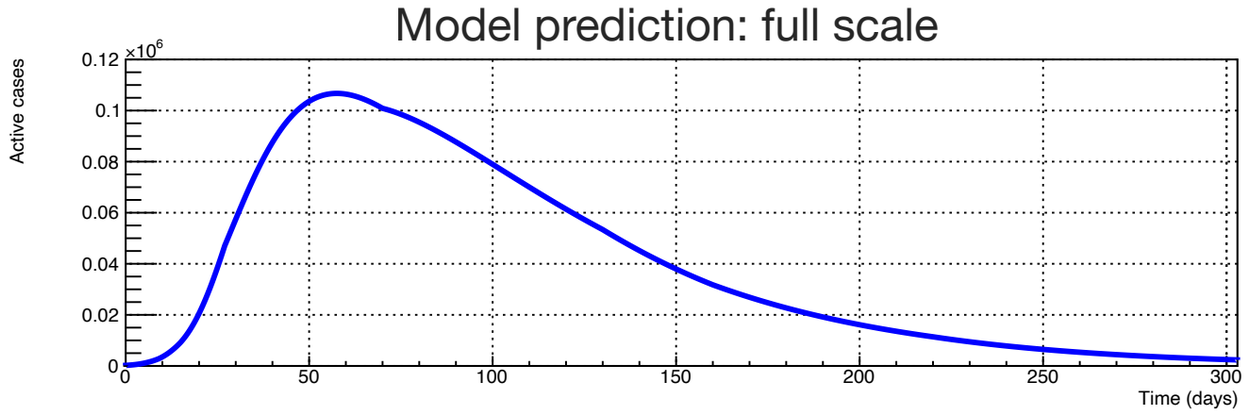


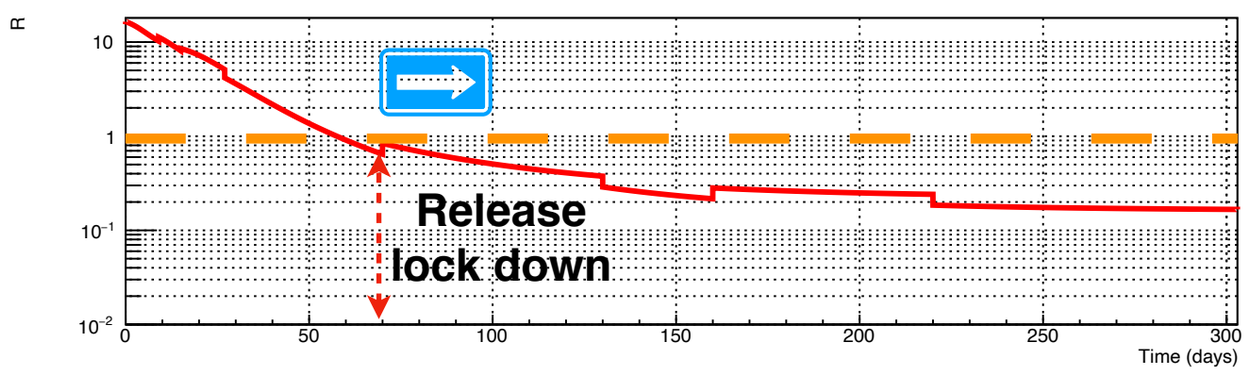


- Normal lifestyle (`g = 0.415`) for 9 days
- Mild soc.distance (`g = 0.324`) for 6 days
- Medium soc.distance (`g = 0.228`) for 12 days
- Hard soc.distance (`g = 0.175`) for **42** days

Past Future

- Mild soc.distance (`g = 0.324`) for 60 days
- Hard soc.distance (`g = 0.175`) for 30 days
- Mild soc.distance (`g = 0.324`) for 60 days
- Hard soc.distance (`g = 0.175`) for 120 days





==========

- Normal lifestyle (`g = 0.415`) for 9 days
- Mild soc.distance (`g = 0.324`) for 6 days
- Medium soc.distance (`g = 0.228`) for 12 days
- Hard soc.distance (`g = 0.175`) for **42** days

Past Future

- Medium soc.distance (`g = 0.228`) for 60 days
- Hard soc.distance (`g = 0.175`) for 30 days
- Medium soc.distance (`g = 0.228`) for 60 days
- Hard soc.distance (`g = 0.175`) for 120 days

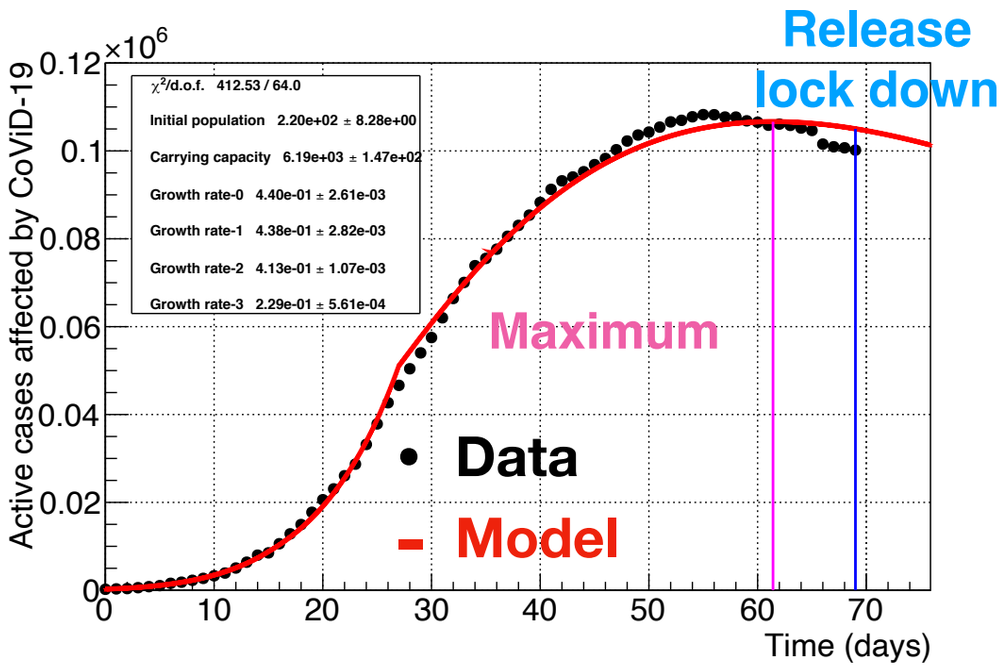
Refinements of the analysis

The model was first tested on different periods, independently, as shown in the previous slides. Then the analysis was further improved by performing:

1. a global fit, i.e. a joint fit, to multiple periods simultaneously, imposing continuity on the number of active cases ($\mathbf{N_n}$) and on the carrying capacity ($\mathbf{C_n}$) across periods, i.e. $\mathbf{N_n}$ and $\mathbf{C_n}$ at the end of period `m` must be equal to $\mathbf{N_n}$ and $\mathbf{C_n}$ at the beginning of period `m+1`

2. a global fit with `logNormal` smearing to take into account effects such as latent/ incubation periods or in general delays in data reporting, law enforcement, or social behaviour

Application to data: fit to multiple time-periods



Number of active cases, **joint fit** periods **1 + 2 + 3 + 4**

For all fit:

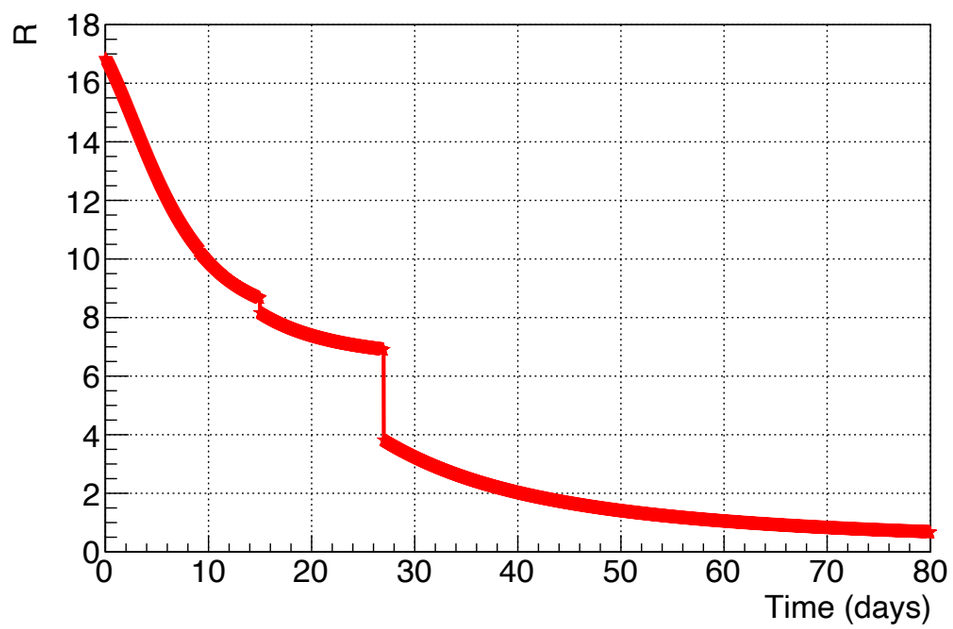
- Errors rescaled by x3

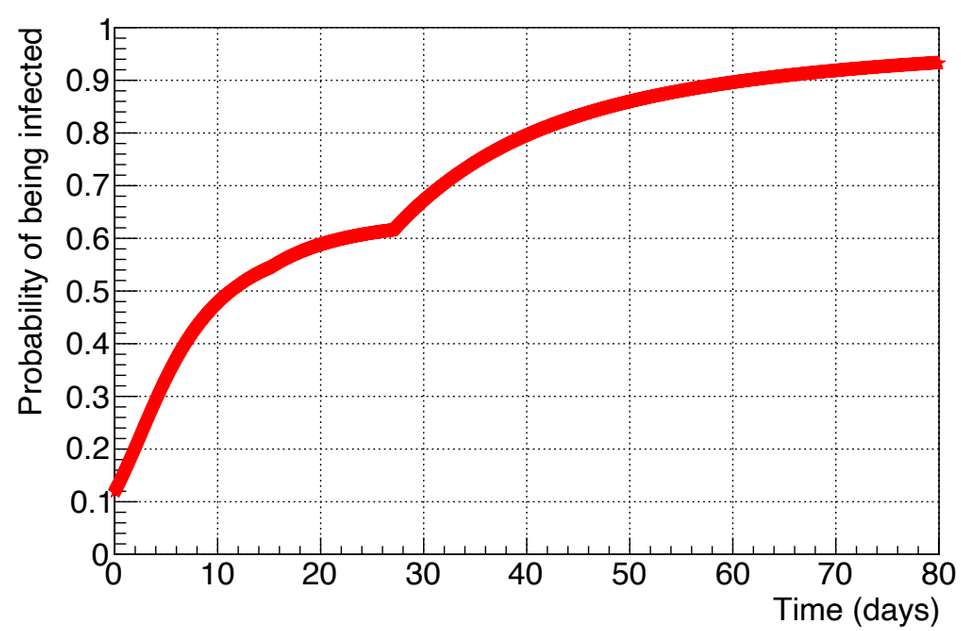
Fitted parameters:

- Initial population (N_0)
- Growth rates periods 1, 2, 3, 4 (**g**)
- Carrying capacity (C_0)

Fixed parameters:

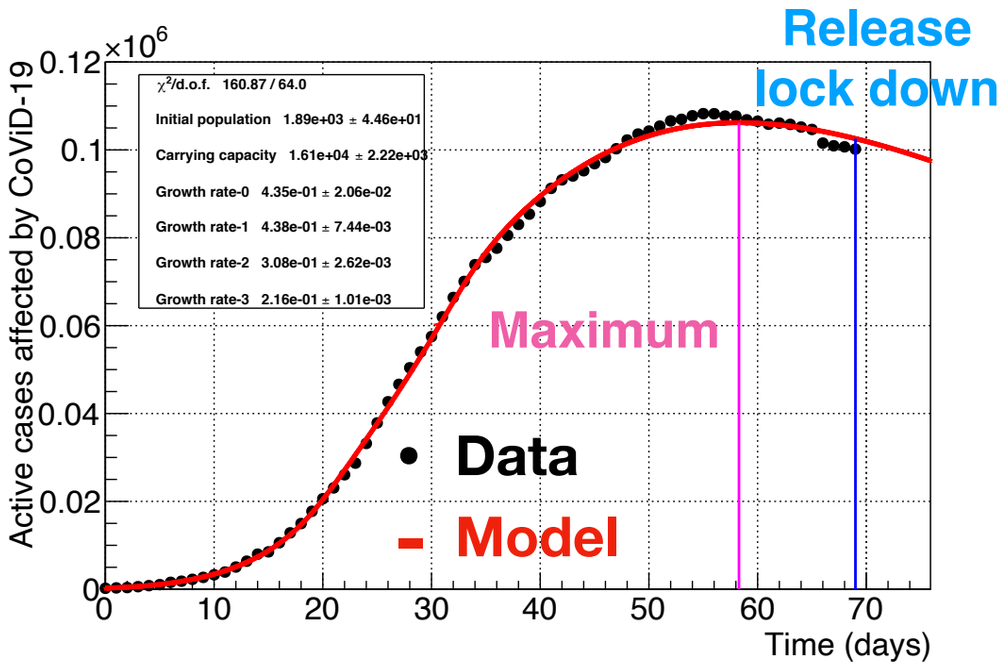
- Recovery rate (**r** = 0.023)
- Fraction of symptomatic (0.3)
- Transmission probability (**p** = 0.26 - see control plots)
- Total population (**T** = 60×10^6)





Probability of being infected, i.e. **1-q**

Application to data: fit to multiple time-periods



Number of active cases, **joint fit** periods **1 + 2 + 3 + 4** with **logNormal**(μ=2,σ=0.3) smearing, i.e. convolution, to mimic effects such as latent/incubation periods (logNormal is used to preserve causality)

For all fit:

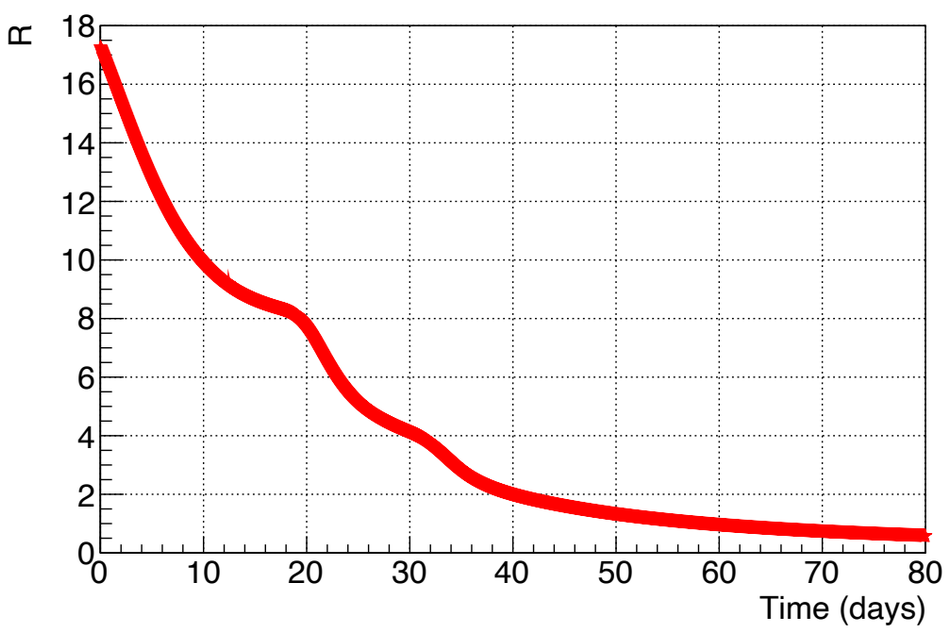
- Errors rescaled by x3

Fitted parameters:

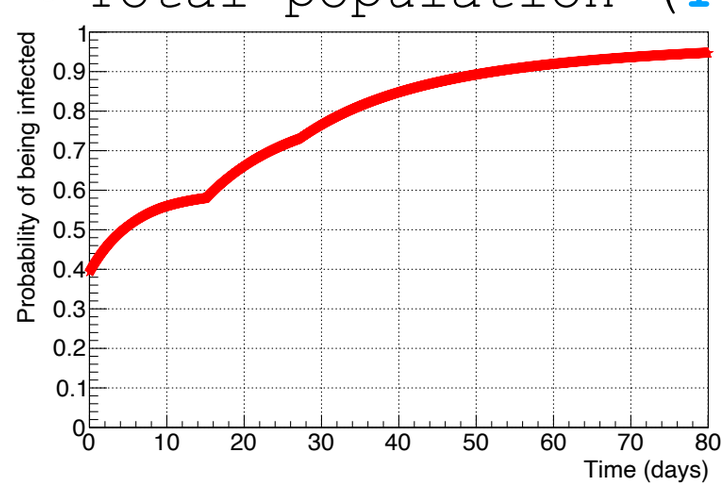
- Initial population (**N_0**)
- Growth rates periods 1, 2, 3, 4 (**g**)
- Carrying capacity (**C_0**)

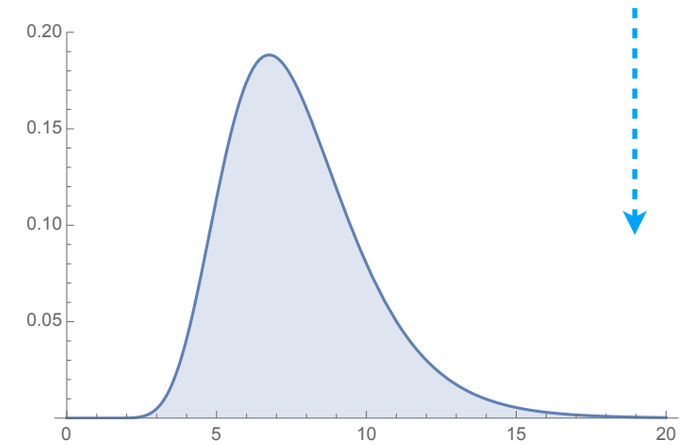
Fixed parameters:

- Recovery rate (**r** = 0.023)
- Fraction of symptomatic (**0.3**)
- Transmission probability (**p** = 0.26 - see control plots)
- Total population (**T** = 60x10⁶)

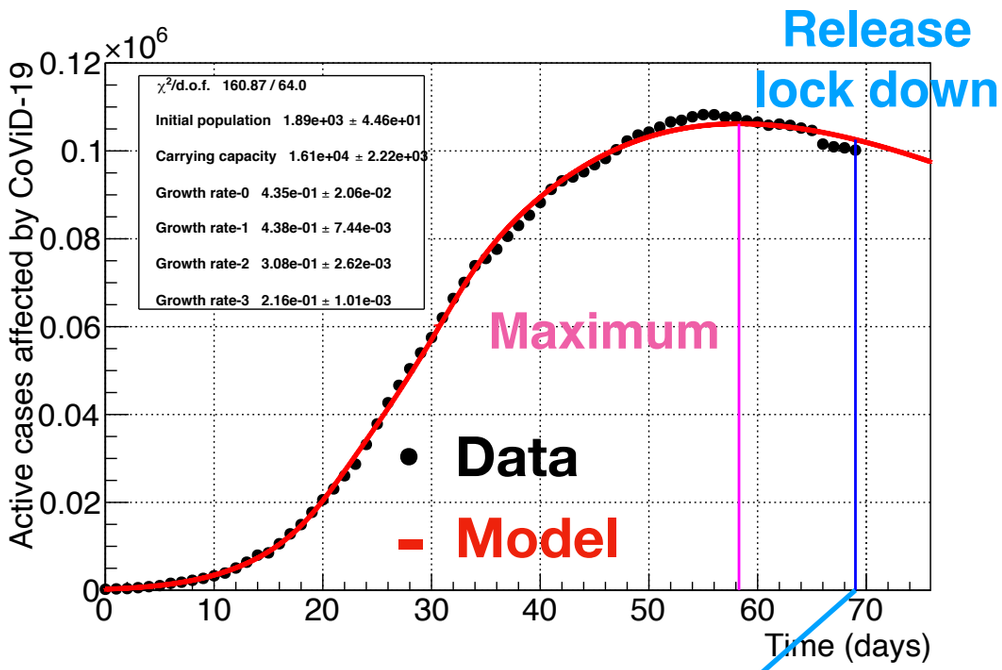


Probability of being infected, i.e. **1-q**





Predictions

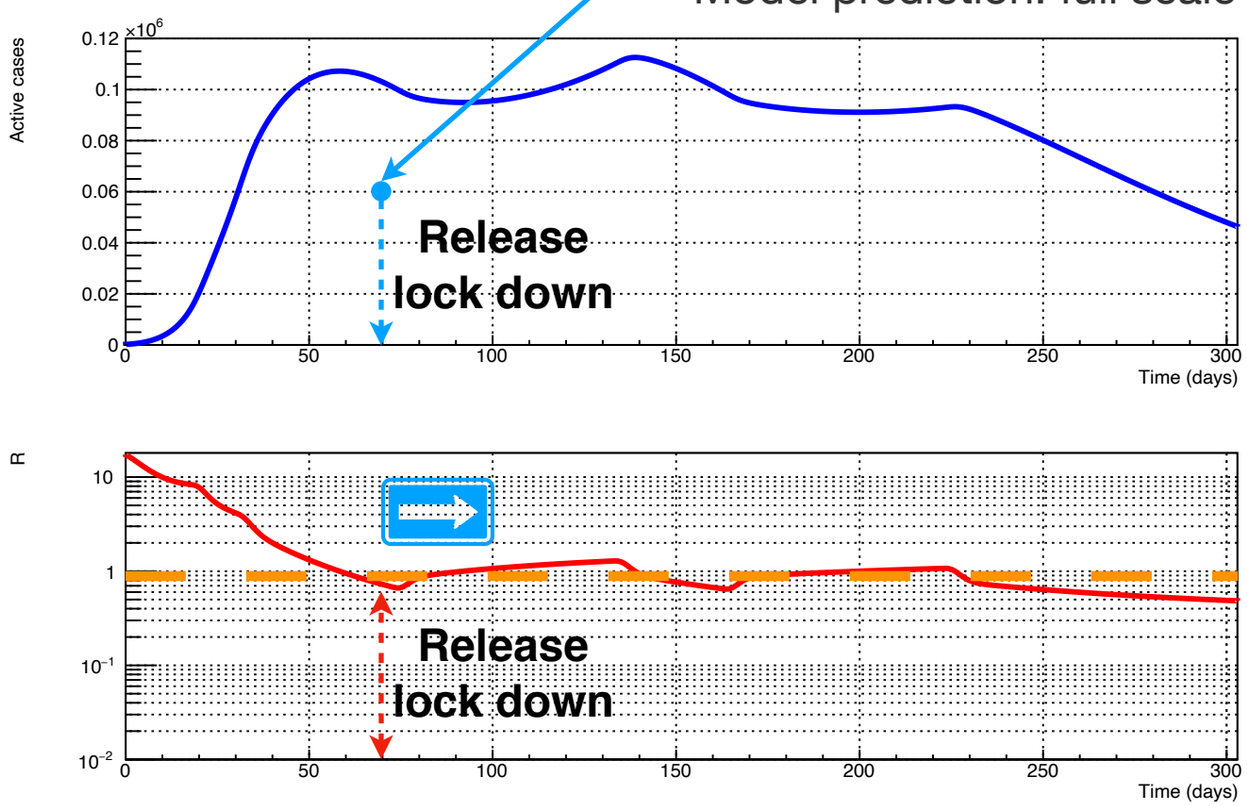


- Normal lifestyle (g = 0.435) for 9 days
- Mild soc.distance (g = 0.438) for 6 days
- Medium soc.distance (g = 0.308) for 12 days
- Hard soc.distance (g = 0.216) for **42** days

Past Future

- Medium soc.distance (g = 0.308) for 60 days
- Hard soc.distance (g = 0.216) for 30 days
- Medium soc.distance (g = 0.308) for 60 days
- Hard soc.distance (g = 0.216) for 120 days

Model prediction: full scale



Other parameters:

- N_0: 1890, C_0: 16100
- r: 0.023
- Symptomatic fraction: 0.3
- Transmission probability: 0.26

Herd immunity:

- Normal lifestyle (g = 0.435): **95%**
- Mild soc.distance (g = 0.438): **95%**
- Medium soc.distance (g = 0.308): **93%**
- Hard soc.distance (g = 0.216): **89%**

Considerations

The model has in total 7 parameters:

- recovery rate **r** ➜ obtained from data plot Δrecovered/active
- growth rate **g** ➜ obtained from fit to the data
- transmission probability **p** ➜ obtained from χ^2/d.o.f. scan over multiple periods
- fraction of symptomatic ➜ obtained from literature
- total population **T** ➜ demography
- initial conditions N_0 and C_0 ➜ obtained from fit to the data

We extracted as much information as possible from the available data and we tried to avoid any a priori hypothesis, though the model unavoidably rely on a couple of tacit assumptions:

- the different periods are well defined, this of course might not be entirely realistic especially for those practices, e.g. wearing a mask, which were not dictated by law, and because the model doesn't take into account the latent/incubation periods, though the smearing with `logNormal` does not show significative effects

- possible differences/variations, region-by-region, double/under counting, etc…, in the criteria of measuring the data, are averaged out

Considerations

Even so the model seems to reasonably describe the data and therefore we tried to project possible evolutions of the disease in the future by exploring three different scenarios:

1. normal lifestyle
2. lock down schools and universities: “mild soc.distance”
3. mild soc.distance + restrictions on movements: “medium soc.distance”

➜ only the third scenario is capable of keeping the effective reproduction number below **1**

- All other scenarios will unavoidably bring **R > 1** with the consequence of clogging the hospitals, even though **R** will no longer reach values as high as at the beginning of the contagion (**~20**)
- It’s well known that the herd immunity, which sets off at a percentage of infected population of ~90%, can stop the infection from spreading, but there is also another phenomenon that can do so even earlier, which is the <u>saturation of the carrying capacity</u>, **C**. In fact the movements of people can not be described with a Brownian motion because a person typically follows the same paths. At saturated carrying capacity the active cases population decreases exponentially: $N(t)=N_0 e^{-rt}$

By any means this work pretend to describe the true dynamic of the infection spread, though it’s interesting to notice that with just few parameters, and reasonable assumptions, the model is able to follow the evolution of the active cases

How to act …

How to act in order to dump the spread of the disease:

1. Decrease the factor p (*probability to spread the disease from one individual to another*): adopt masks, gloves, disinfections, etc… Mutations of the pathogen can also change p

2. Decrease factor k (*average number of encountered people by one individual*): enforce social distancing

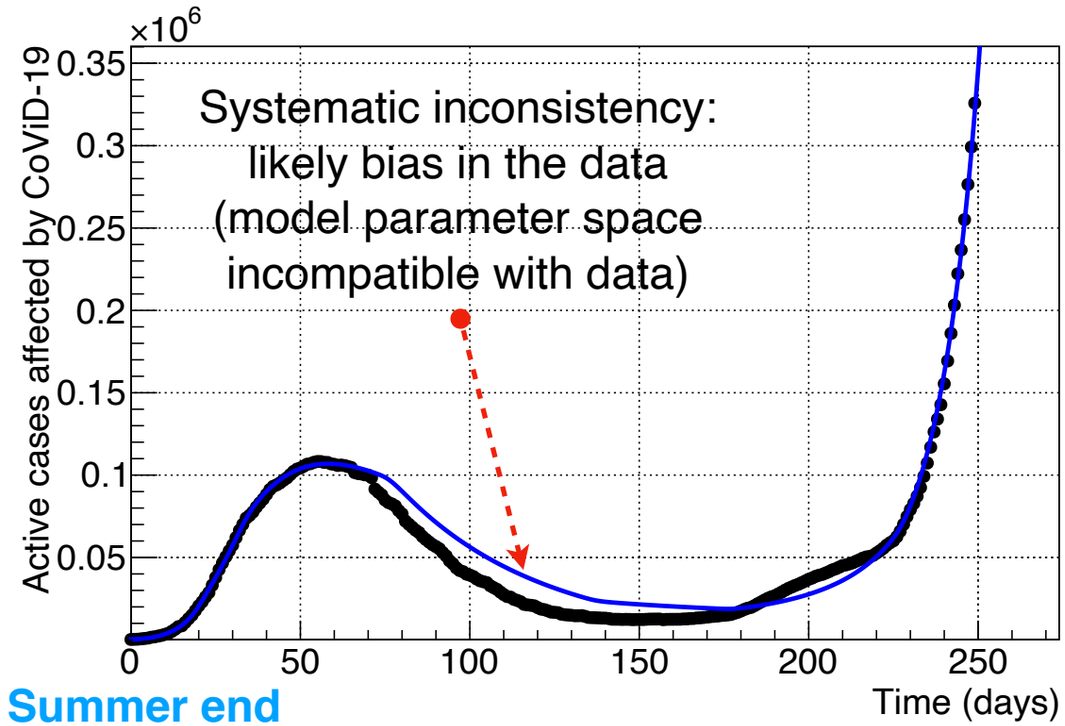
3. Increase factor Q (*fraction of active cases which are in quarantine*): quickly discover the new infected cases and put them in quarantine

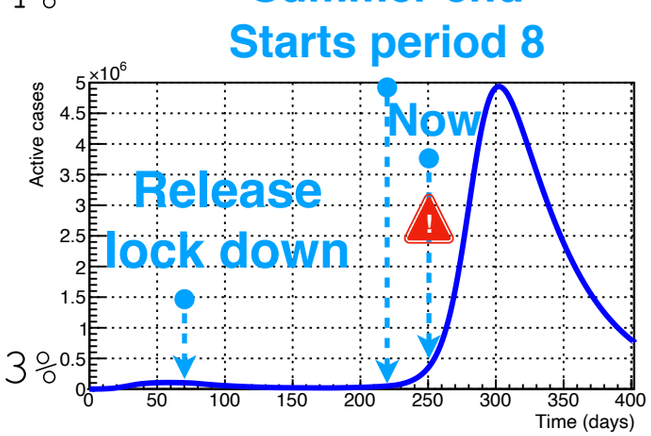
4. Increase factor r (*recovery rate*): develop a vaccine (r is likely to be age-dependent). Mutations of the pathogen can also change r

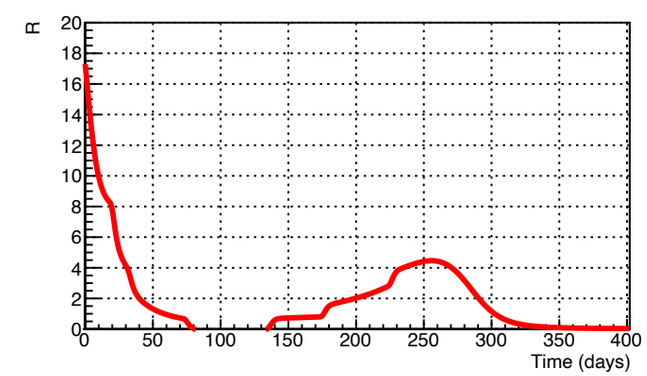
New data … 31/10/2020

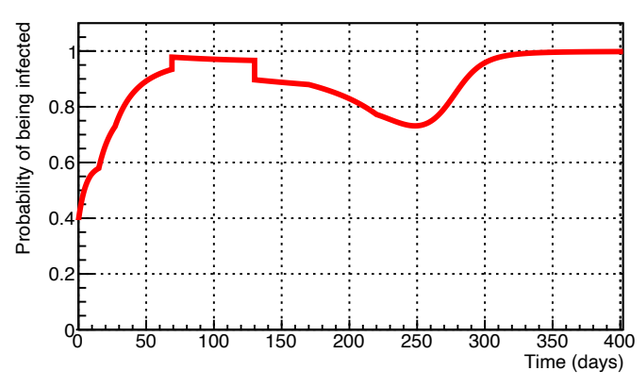
After the end of the lockdown several data have been collected and they have been found quite compatible with the following parameters:

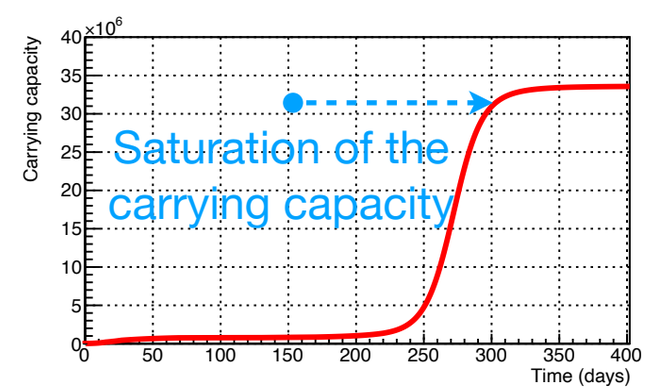
- **period 5 (end lockdown + 61 days)**
  - recovery rate: `r = 0.025`
  - transmission prob.: `p = 26%/8 = 3.3%`
  - growth rate: `g = 0.435/8 = 0.054` (fixed)
- **period 6 (end period 5 + 40 days):**
  - recovery rate: `r = 0.023`
  - transmission prob.: `p = 26%/2.5 = 10.4%`
  - growth rate: `g = 0.154` (from fit)
- **period 7 (end period 6 + 50 days):**
  - recovery rate: `r = 0.023`
  - transmission prob.: `p = 26%/1.5 = 17.3%`
  - growth rate: `g = 0.299` (from fit)
- **period 8 (end period 7 + 200 days): proj.**
  - recovery rate: `r = 0.023`
  - transmission prob.: `p = 26%`
  - growth rate: `g = 0.4` (from fit)









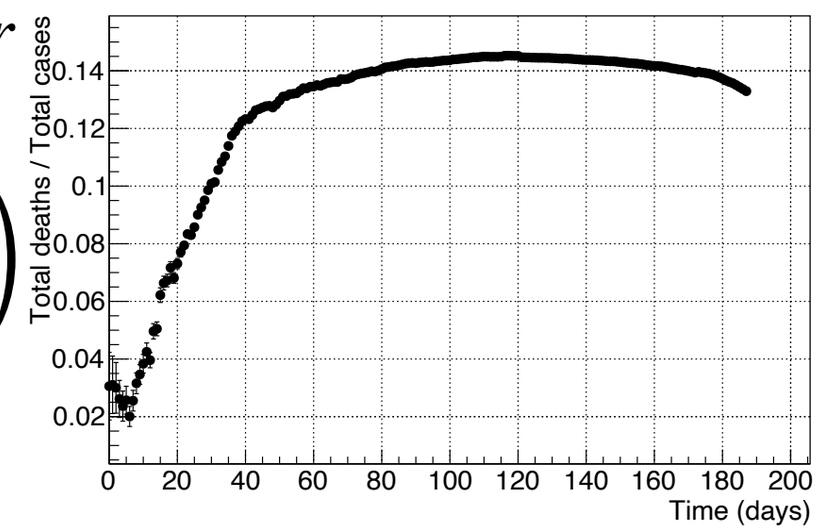


The model

The model is further improved by explicitly taking into account the mortality fraction caused by the disease (**m**): the affected parameters are the total population, **T**, and the average number of encountered people, **K**

$$\begin{cases} N_{n+1} = N_n + N_n\left(K\frac{T_n}{T_0}\right)p\left(1 - \frac{N_n + r\sum_{i=0}^{n-1} N_i}{C_n}\right) - N_n r \\ C_{n+1} = C_n + \left(N_{n+1} - N_n(1-r)\right)\left(K\frac{T_n}{T_0}\right)\left(1 - \frac{C_n}{T_n}\right) \\ T_{n+1} = T_n - N_n r m \end{cases}$$

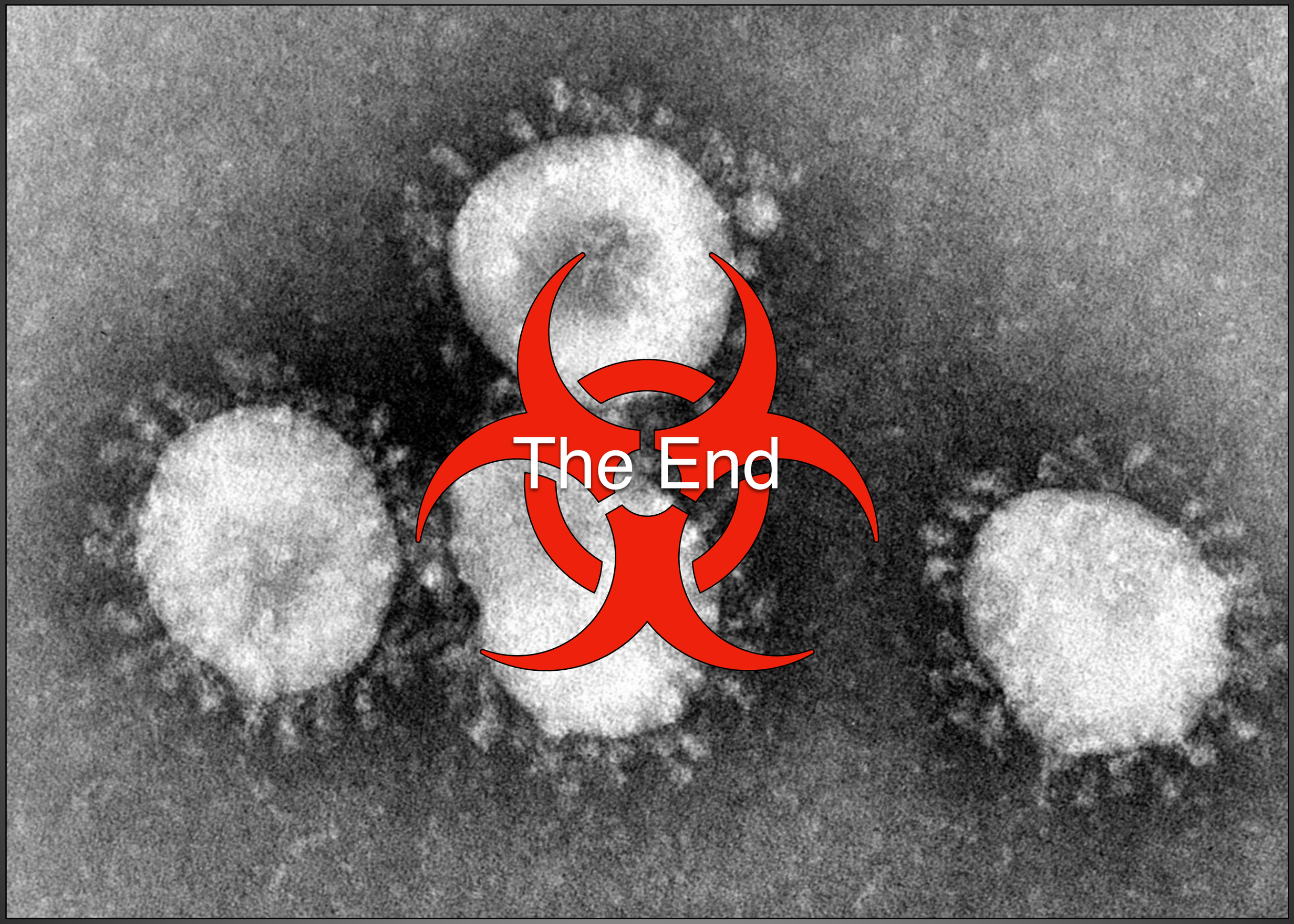
Mortality fraction, **m**, from data



- The total population, **T**, decreases by the amount $N_n\, r\, m$, i.e. $r\, m$ is the fraction of deceased patients
- The average number of encountered people, **K**, decreases proportionally to the actual population fraction, T_n/T_0, where T_0 is the initial population

Differential form:

$$\begin{cases} \frac{dN(t)}{dt} = N(t)\left(K\frac{T(t)}{T_0}\right)p\left(1 - \frac{N(t) + r\int_0^t N(t')dt'}{C(t)}\right) - N(t)r \\ \frac{dC(t)}{dt} = \left(\frac{dN(t)}{dt} + N(t)r\right)\left(K\frac{T(t)}{T_0}\right)\left(1 - \frac{C(t)}{T(t)}\right) \\ \frac{dT(t)}{dt} = -N(t)rm \end{cases}$$

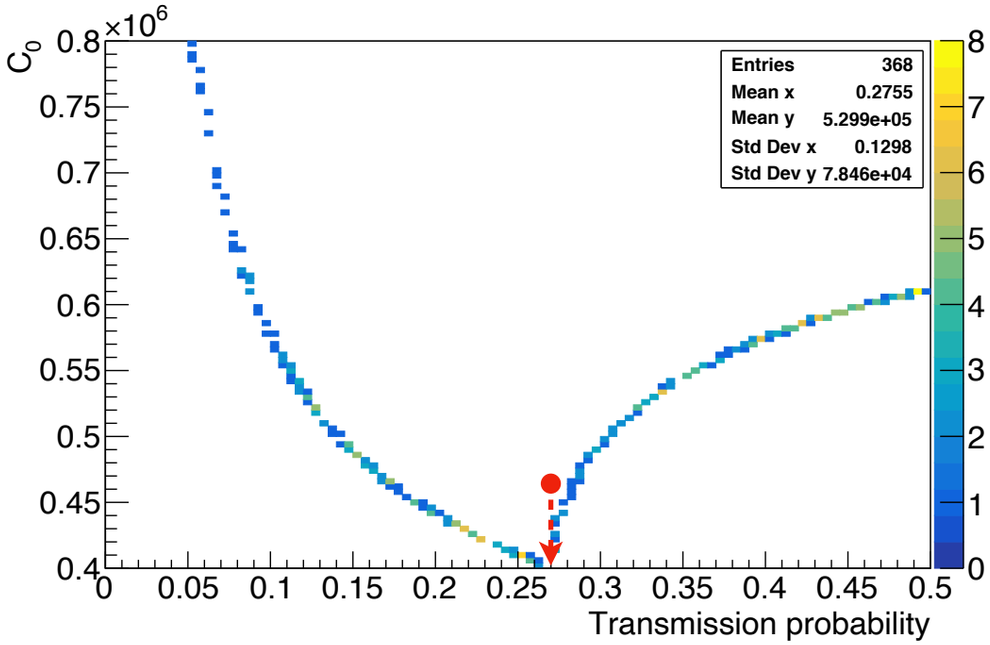

The End

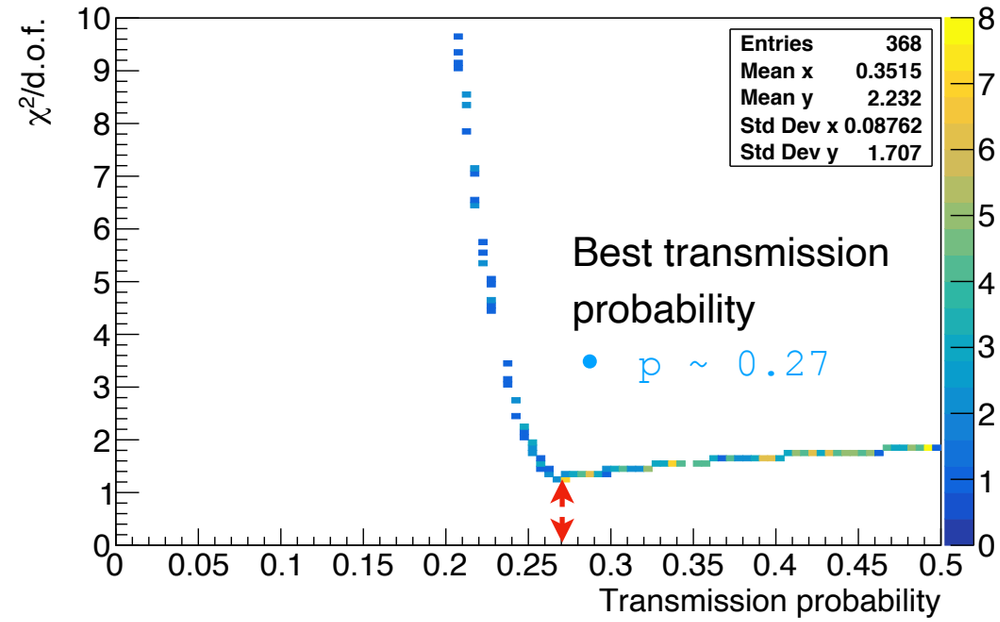
Control plots

Possible model ameliorations/extensions:

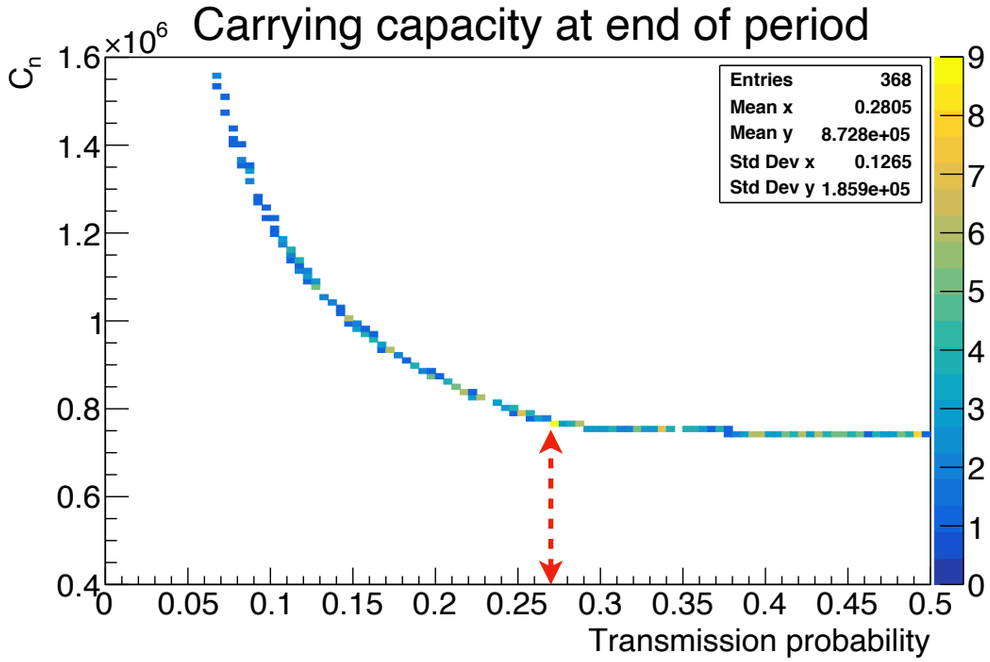
- Take into account natural demographic change

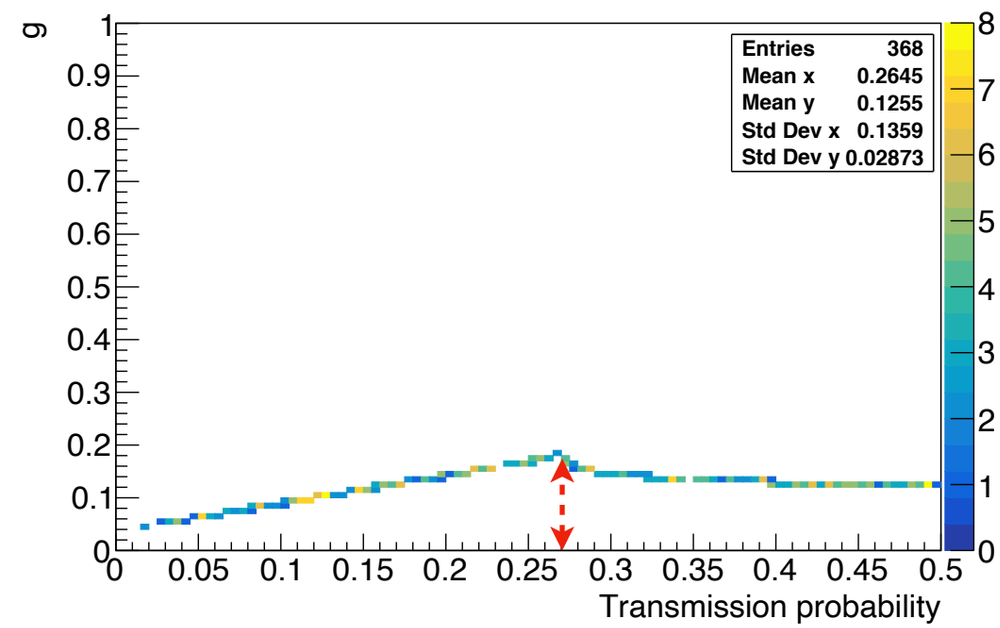
Application to data: fit to single time period





Carrying capacity at end of period





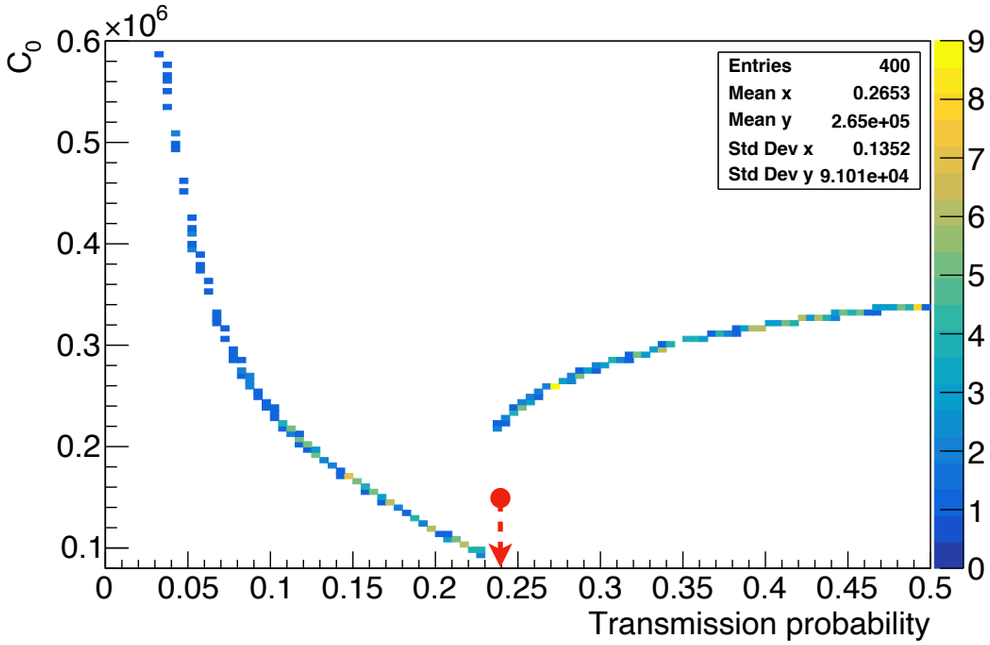
Fit data from 27th day till release (period 4)

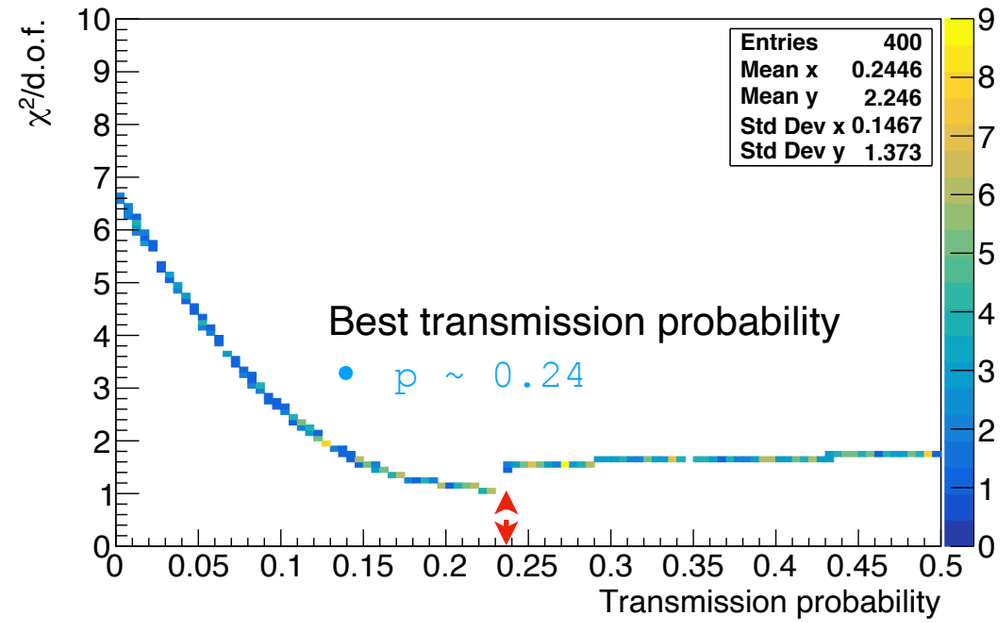
- 400 extractions of transmission probability, **p**

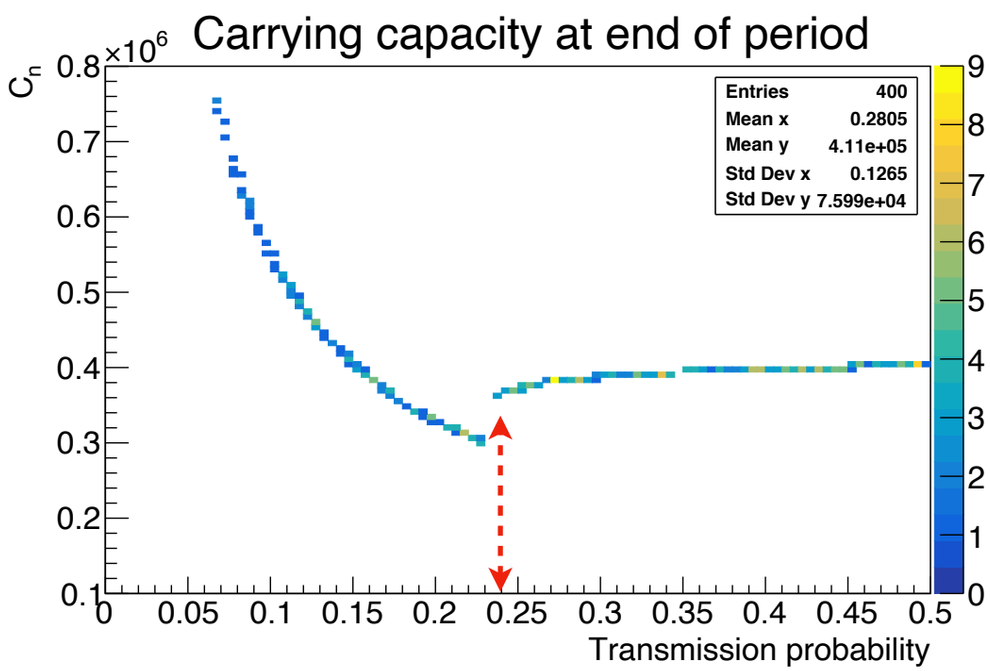
Fixed parameters:

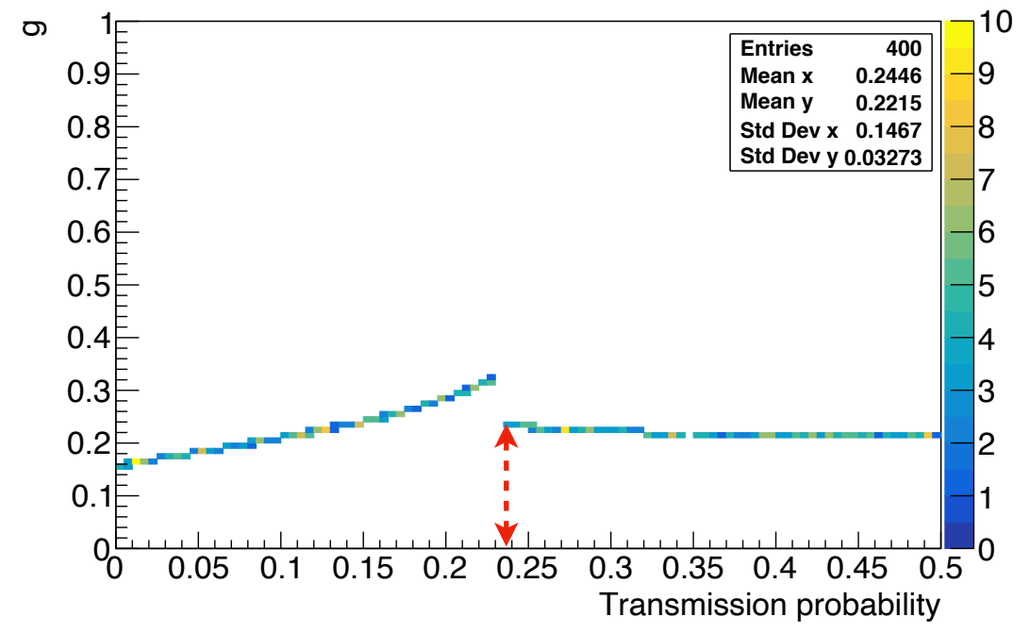
- **r = 0.023** ➜ plot Δrecovered/active
- Symptomatic fraction = **0.3** ➜ literature
- **T = 60x10⁶** ➜ demography

Application to data: fit to single time period









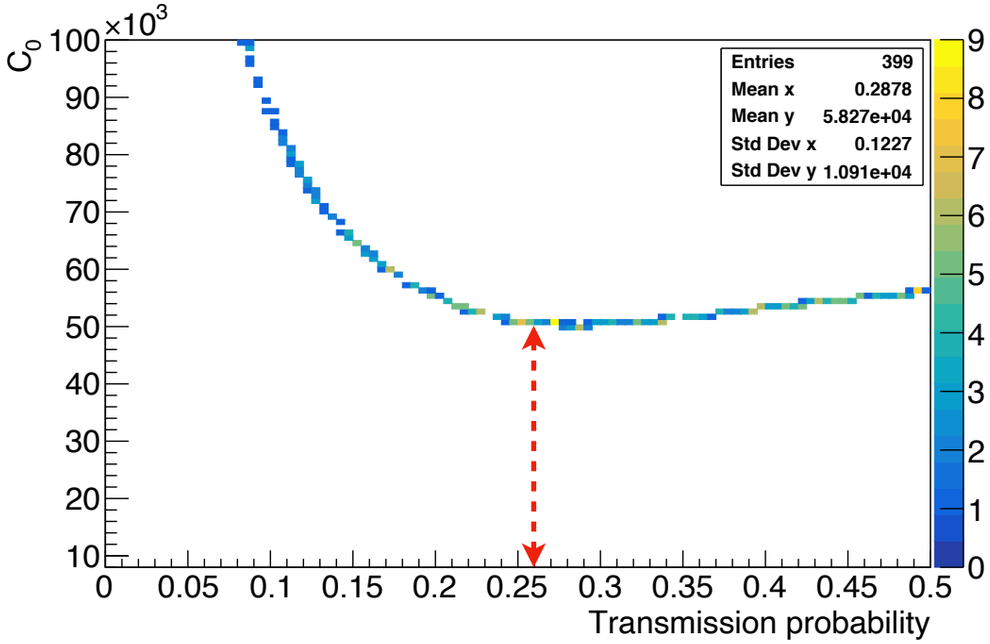
Fit data from 15th to 26th day (period 3)

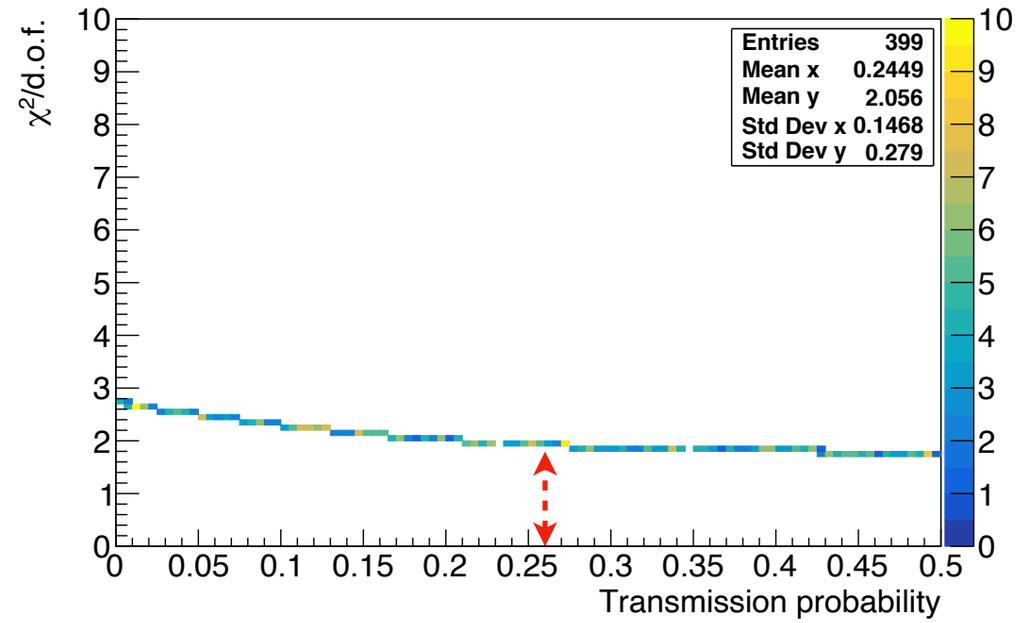
- 400 extractions of transmission probability, p

Fixed parameters:

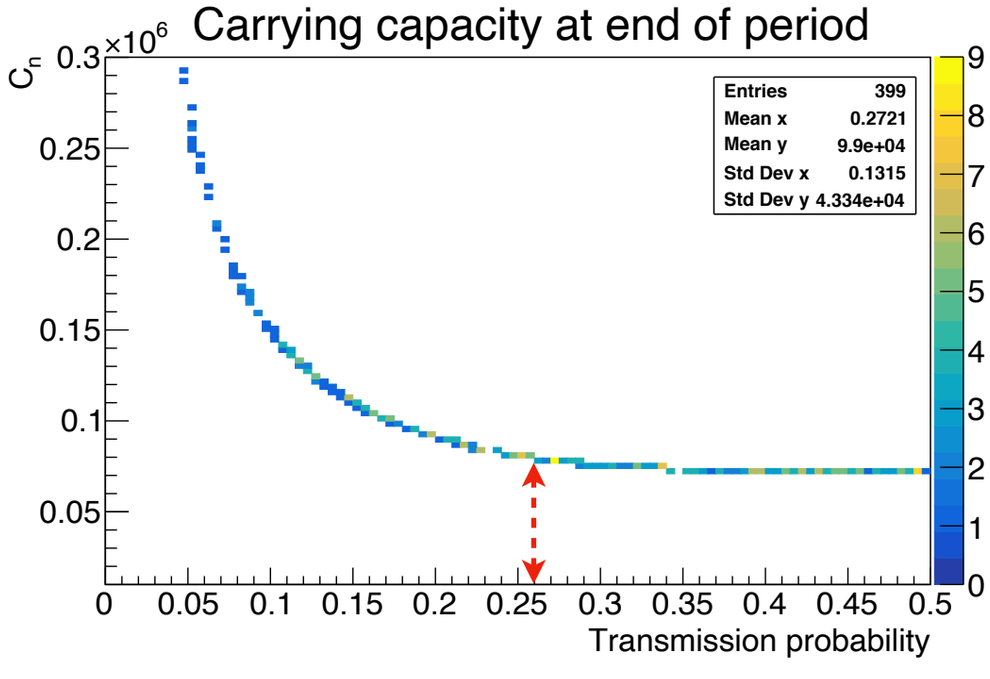
- r = 0.023 ➜ plot Δrecovered/active
- Symptomatic fraction = 0.3 ➜ literature
- T = 60x10⁶ ➜ demography

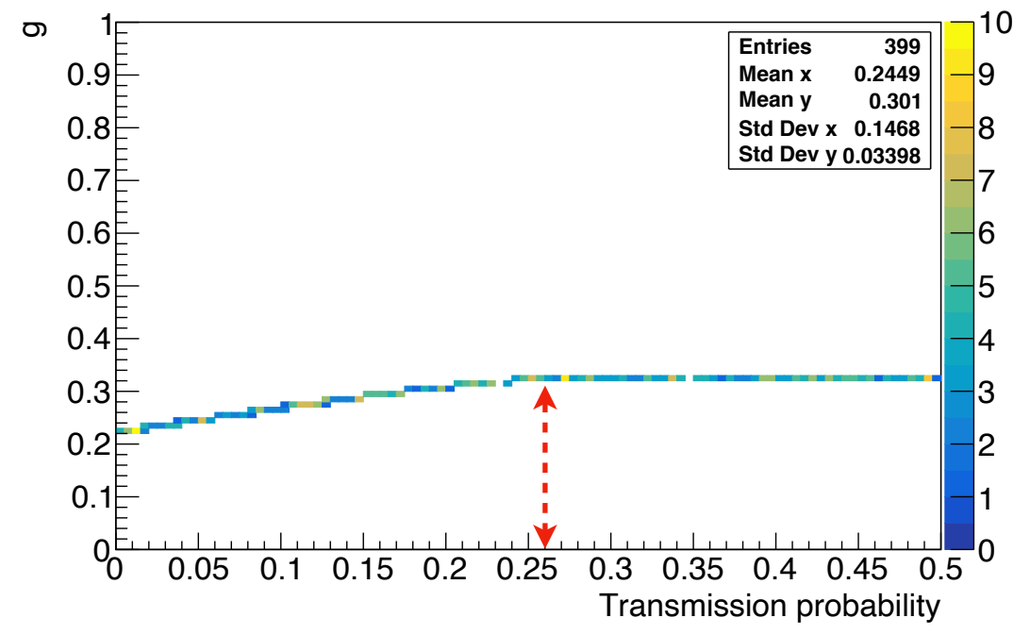
Application to data: fit to single time period





Carrying capacity at end of period





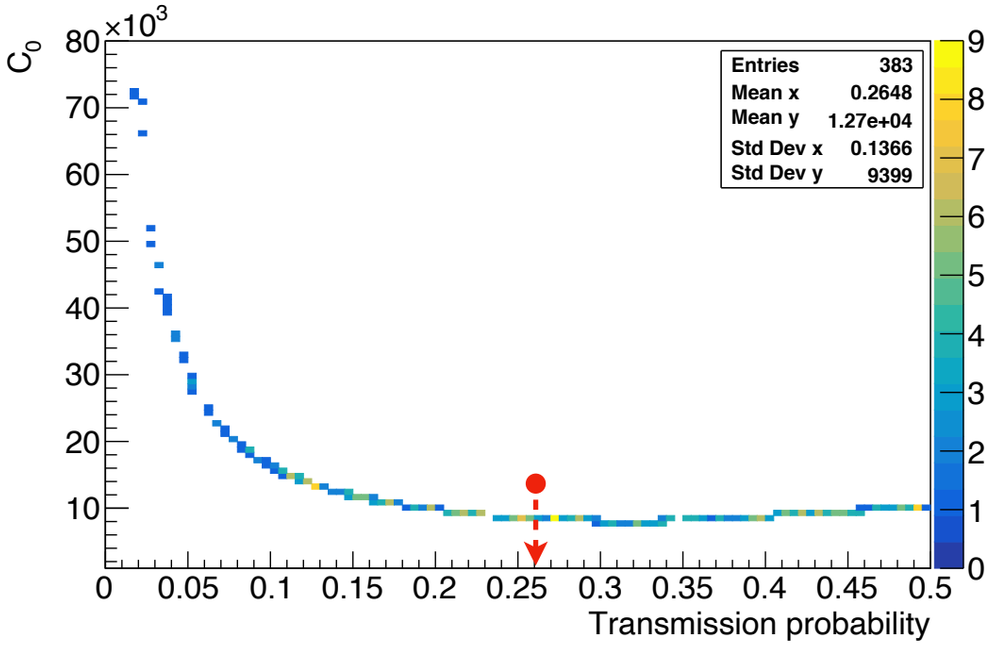
Fit data from 9th to 14th day (period 2)

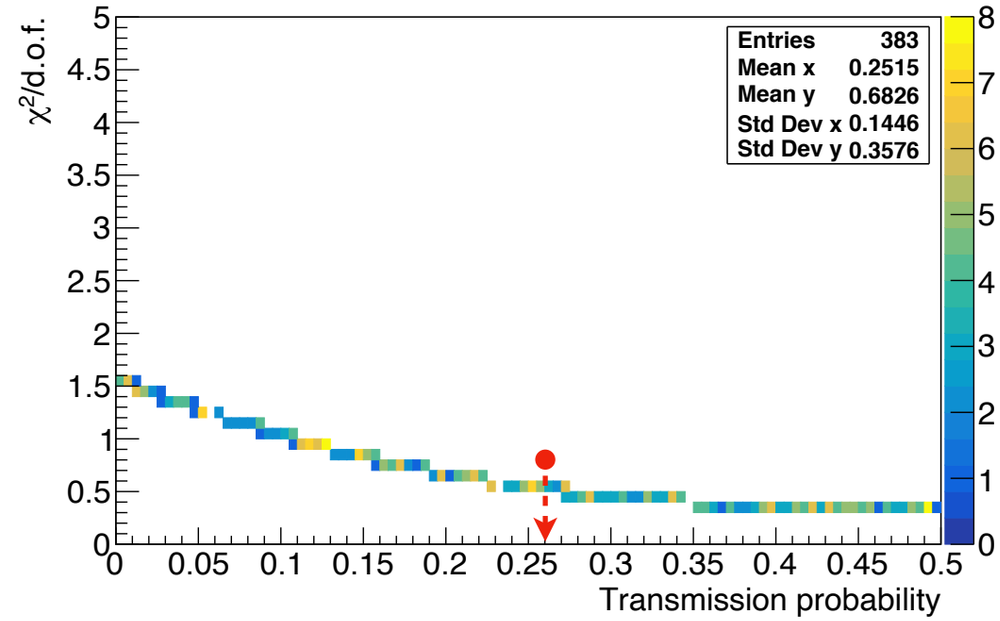
- 400 extractions of transmission probability, p

Fixed parameters:

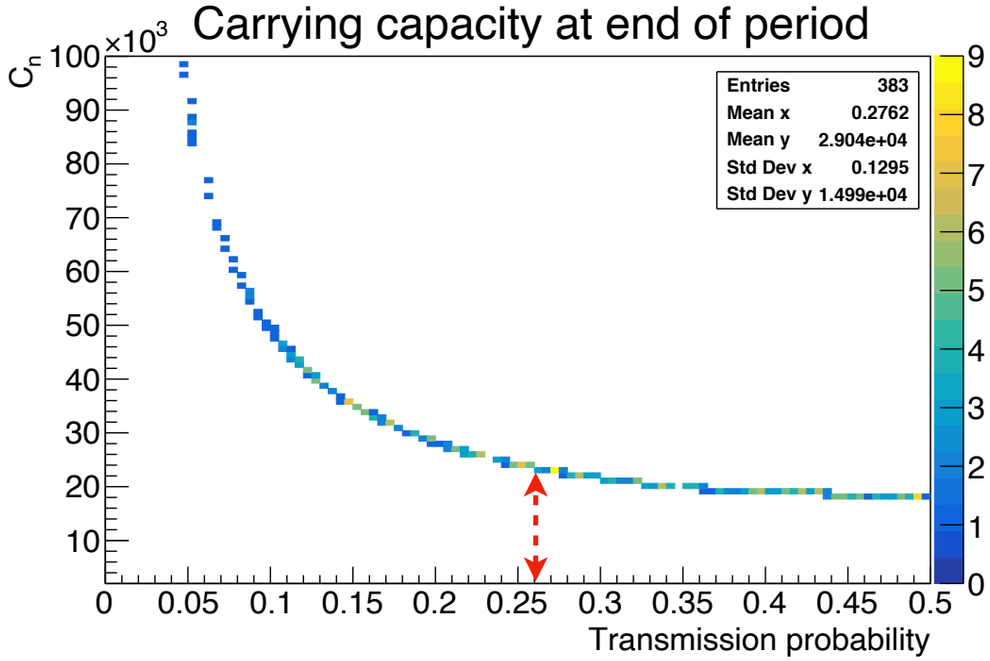
- r = 0.023 ➜ plot Δrecovered/active
- Symptomatic fraction = 0.3 ➜ literature
- T = 60x10⁶ ➜ demography

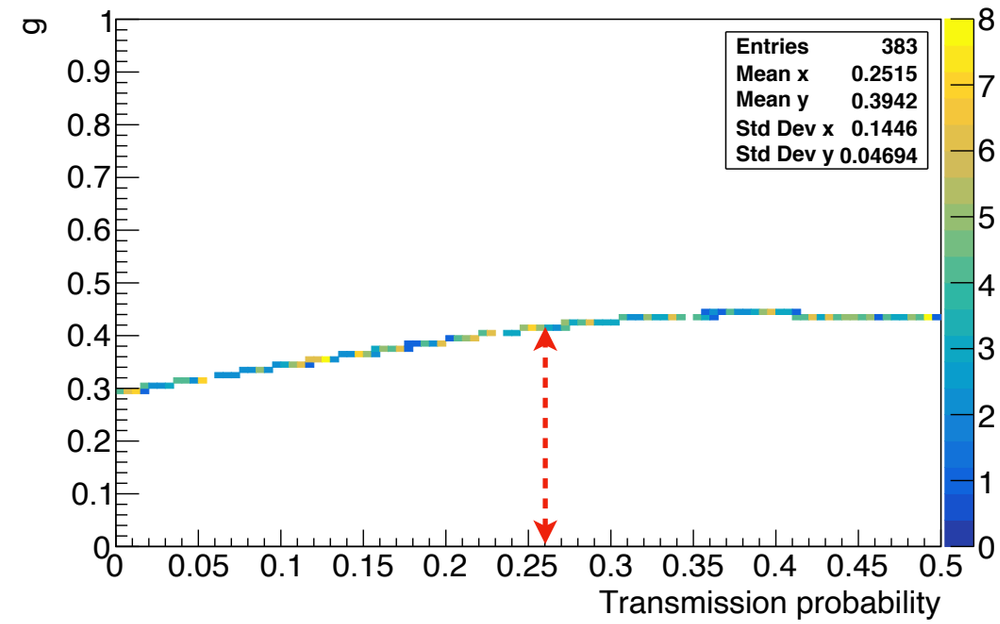
Application to data: fit to single time period





Carrying capacity at end of period





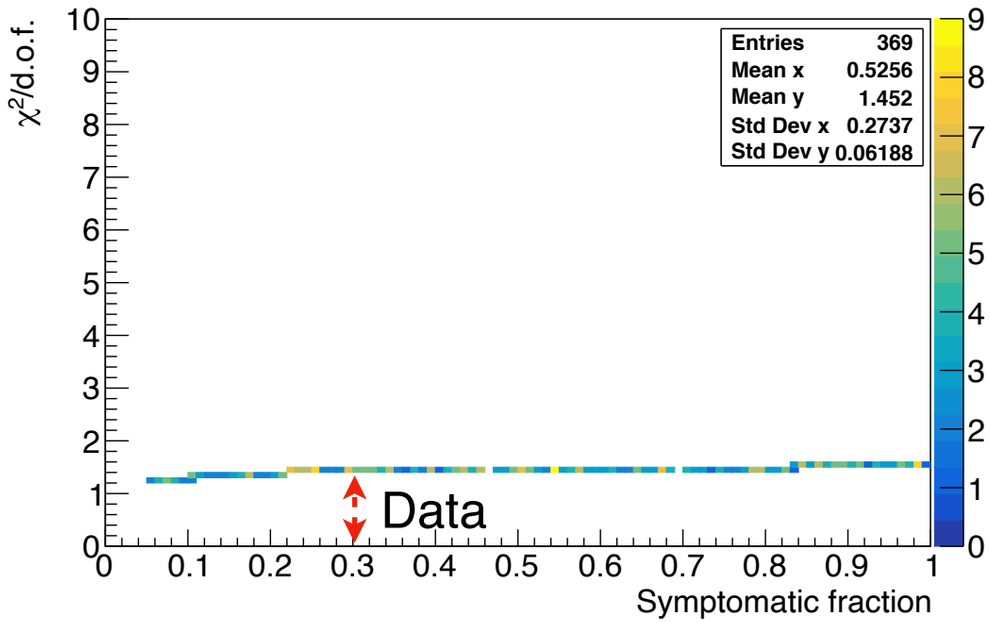
Fit data from 0th to 8th day (period 1)

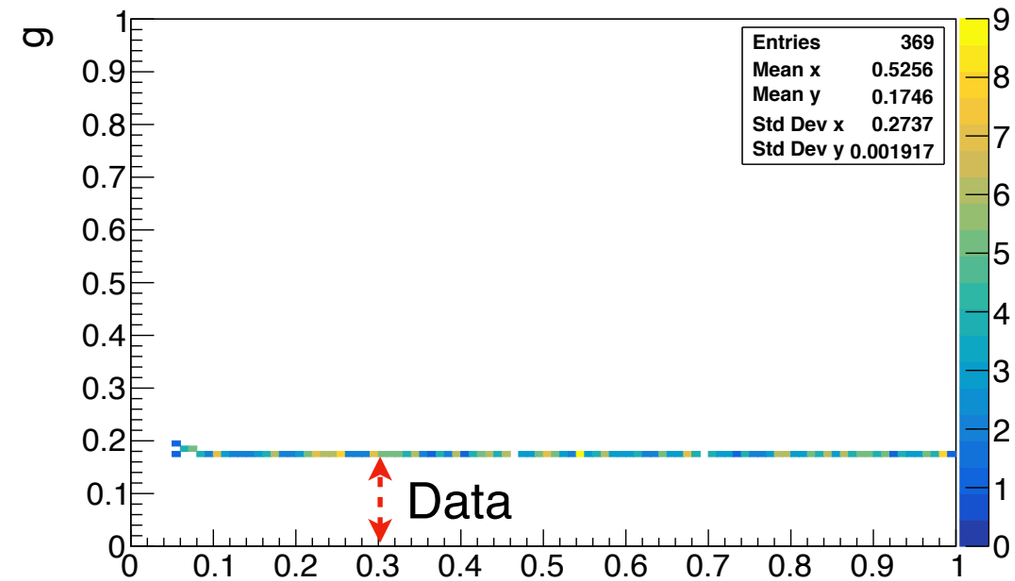
- 400 extractions of transmission probability, p

Period 3 and 4 have slightly different transmission probability ➜ averaging:

- p ~ 0.26

Application to data: fit to single time period





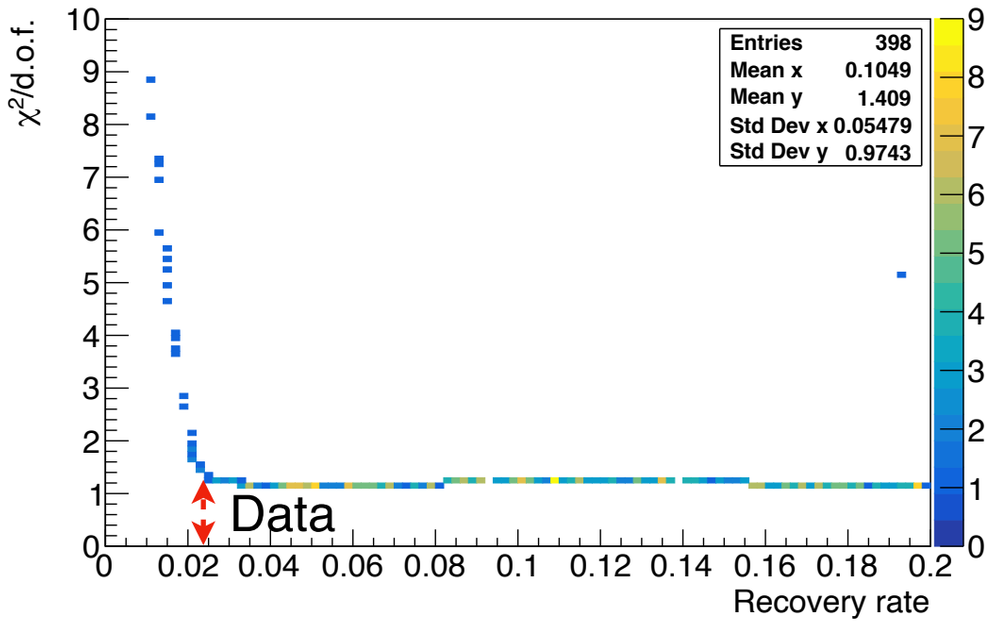
Fit data from 27th day till release (period 4)

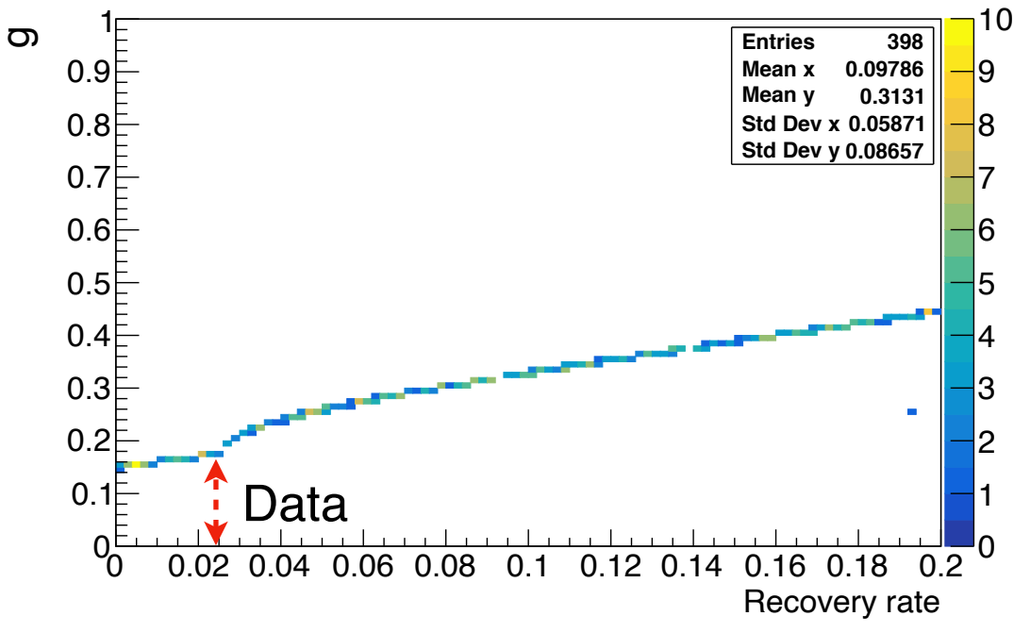
- `400` extractions of symptomatic fraction

Fixed parameters:

- `r = 0.023` ➜ plot Δrecovered/active
- `p = 0.26`
- `T = 60x10⁶` ➜ demography

Application to data: fit to single time period





Fit data from 27th day till release (period 4)

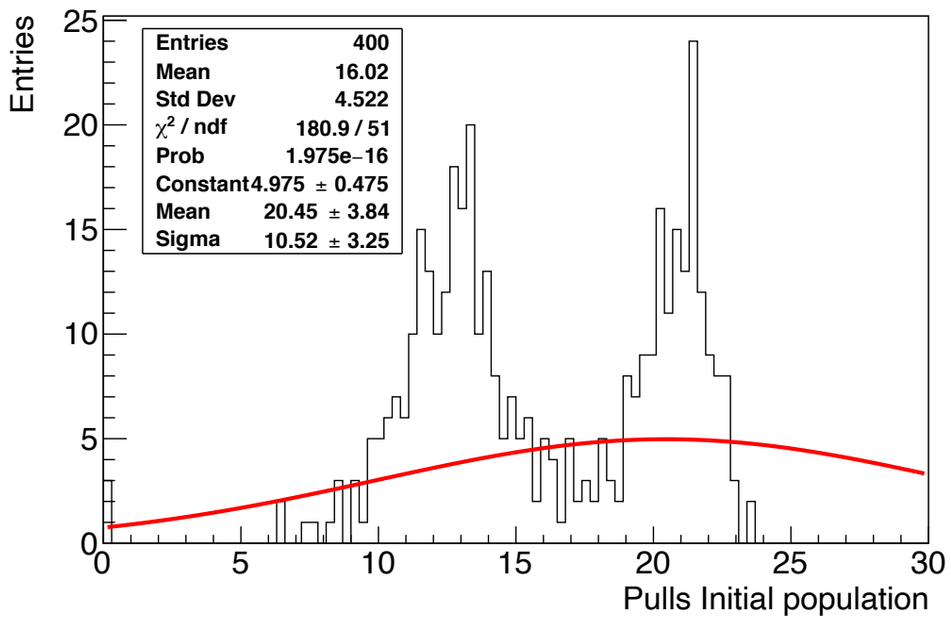
- `400` extractions of `r`

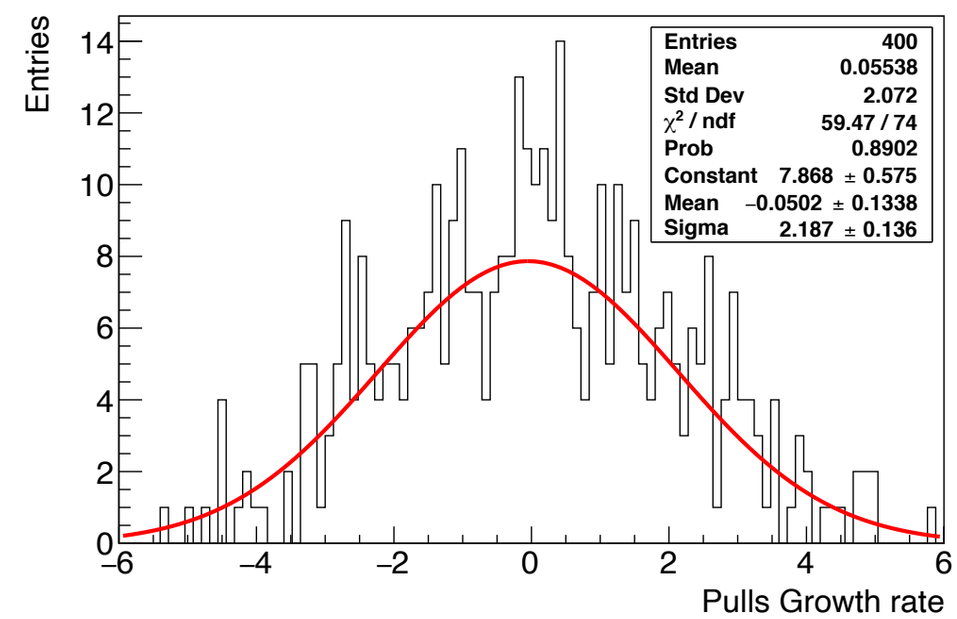
Fixed parameters:

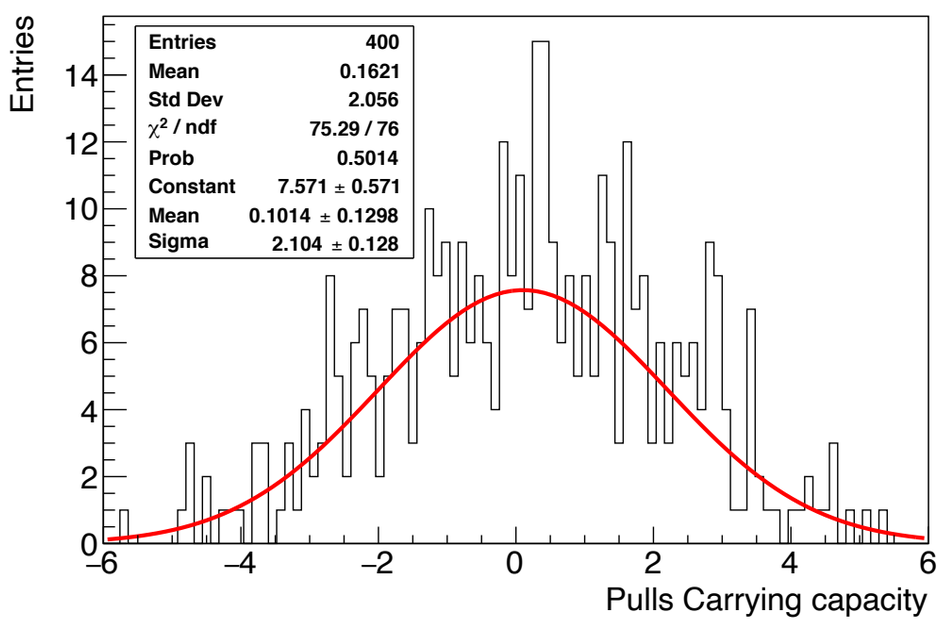
- `p = 0.26`
- Symptomatic fraction `= 0.3` ➜ literature
- `T = 60x10⁶` ➜ demography

Pseudo experiment studies

Generated 400 pseudo experiments of <u>period 4</u>, and re-fit them in order to check stability of the fit procedure

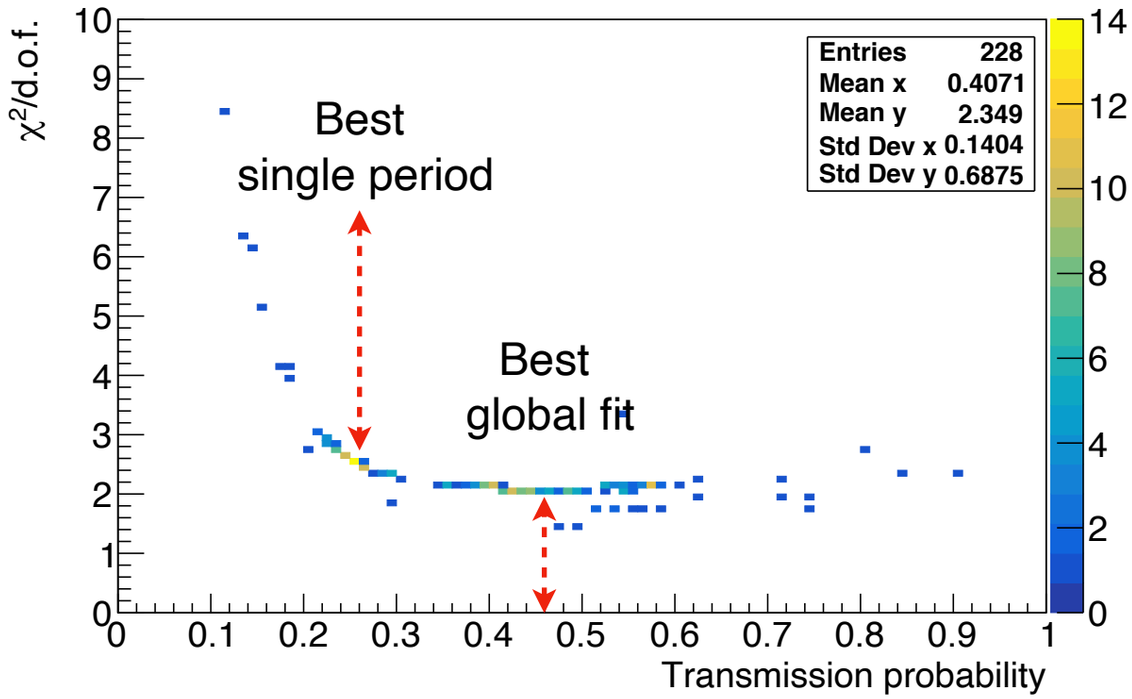






Uncertainties seems to be underestimated by a factor of 2

Application to data: fit to multiple time-periods



Joint fit periods **1 + 2 + 3 + 4**, with smearing

- `1000` extractions of transmission probability, `p`

Fixed parameters:

- `r = 0.023` ➜ plot Δrecovered/active
- Symptomatic fraction `= 0.3` ➜ literature
- `T = 60x10⁶` ➜ demography

Best transmission probability

- `p ~ 0.46`