Stochastic Disease Spread Models

Wolfgang Hörmann

Zeynep Gökçe İşlier

Refik Güllü

Jannuary 23, 2019

SIR (Susceptible-Infected-Recovered) Model

- 3 compartments:
  S (susceptible) $\rightarrow$ I (infected) $\rightarrow$ R (recovered)
- $N = R(t) + I(t) + S(t)$
- The rate of transitions from susceptible (S) to infected (I) is proportional to $I \cdot S$.
- The transition from infected (I) to recovered (R) is proportional to I.
- When $I = 0$ no more transaction can occur
  $\Rightarrow$ absorbing state is $I = 0$.

Continuous Time Markovian SIR Disease Spread Models

Parameters: $\mu \ldots$ expected disease length; $\lambda \ldots$ infection rate
$N \ldots$ size of population

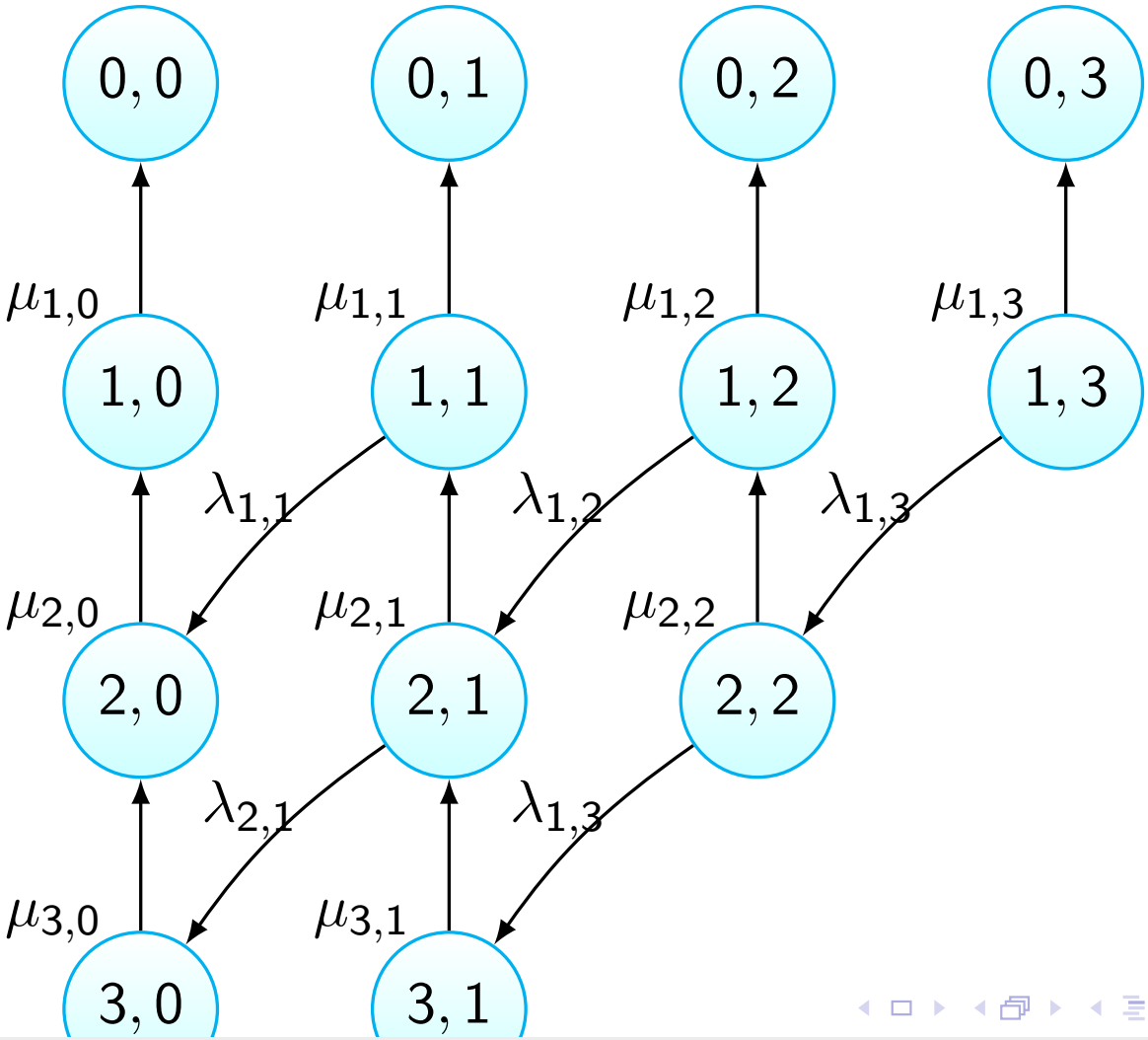
- disease duration for an infected $\sim$ Exponential(μ)
- transition rate from infected to recovered $\mu_{is} = \mu_i = i\mu$
- number of contacts for every infected individual $\sim$ Poisson(λ)
- total number of contacts $\sim$ Poisson($I\,S\,\lambda/N$)
  (Hernandez Suarez et al., 2010; Nasell, 2002)
- transition rate from susceptible to infected $\lambda_{i,s} = i\,s\,\lambda/N$
- N $= S + I + R$
- State space $\Omega = (I(t), S(t))$

Markovian SIR Disease Spread Models continued

Important quantities studied in the literature

1. Probability of an outbreak
2. Final outbreak size distribution ("Attack Rate")
3. Time to extinction ("Length of Epidemic")
4. Distribution of the maximum number of infected

State Space Diagram for SIR with $N = 4$



First Step Analysis for Expected time to extinction

- No return to a state previously visited $\Rightarrow$
  $E[T_{i,s}]$ = expected time to extinction can be calculated.
- Absorbing states: $(0, s)$ for $s = 0, 1, \ldots N$
- Boundary Conditions

for states with zero infected: $E[T_{0,s}] = 0$

for states with zero susceptible: $E[T_{i,0}] = \sum_{k=1}^{i} \frac{1}{\mu_k}$

Implementation of First Step Analysis

- Start with one infected and one susceptible:

$$E[T_{1,1}] = \frac{1}{\lambda_{1,1} + \mu_1} + \frac{\lambda_{1,1}}{\lambda_{1,1} + \mu_1} E[T_{2,0}] + \frac{\mu_1}{\lambda_{1,1} + \mu_1} E[T_{0,1}].$$

- Recursion for $E[T_{i,1}]$:

$$E[T_{i,1}] = \frac{1}{\lambda_{i,1} + \mu_i} + \frac{\lambda_{i,1}}{\lambda_{i,1} + \mu_i} E[T_{i+1,0}] + \frac{\mu_i}{\lambda_{i,1} + \mu_i} E[T_{i-1,1}].$$

$E[T_{i,s}]$ for s susceptible cases

- Recursion for $E[T_{i,s}]$:

$$E[T_{i,s}] = \frac{1}{\lambda_{i,s} + \mu_i} + \frac{\lambda_{i,s}}{\lambda_{i,s} + \mu_i} E[T_{i+1,s-1}] + \frac{\mu_i}{\lambda_{i,s} + \mu_i} E[T_{i-1,s}]$$

- For the SIR model: Replace $\lambda_{i,s}$ by $i\,s\lambda/N$ and μ_i by $i\mu$

$$E[T_{i,s}] = \frac{N}{i\,s\lambda + Ni\mu} + \frac{s\lambda}{s\lambda + N\mu} E[T_{i+1,s-1}] + \frac{N\mu}{s\lambda + N\mu} E[T_{i-1,s}]$$

Calculation of Final Size (= attack rate) Distribution

- end of epidemic

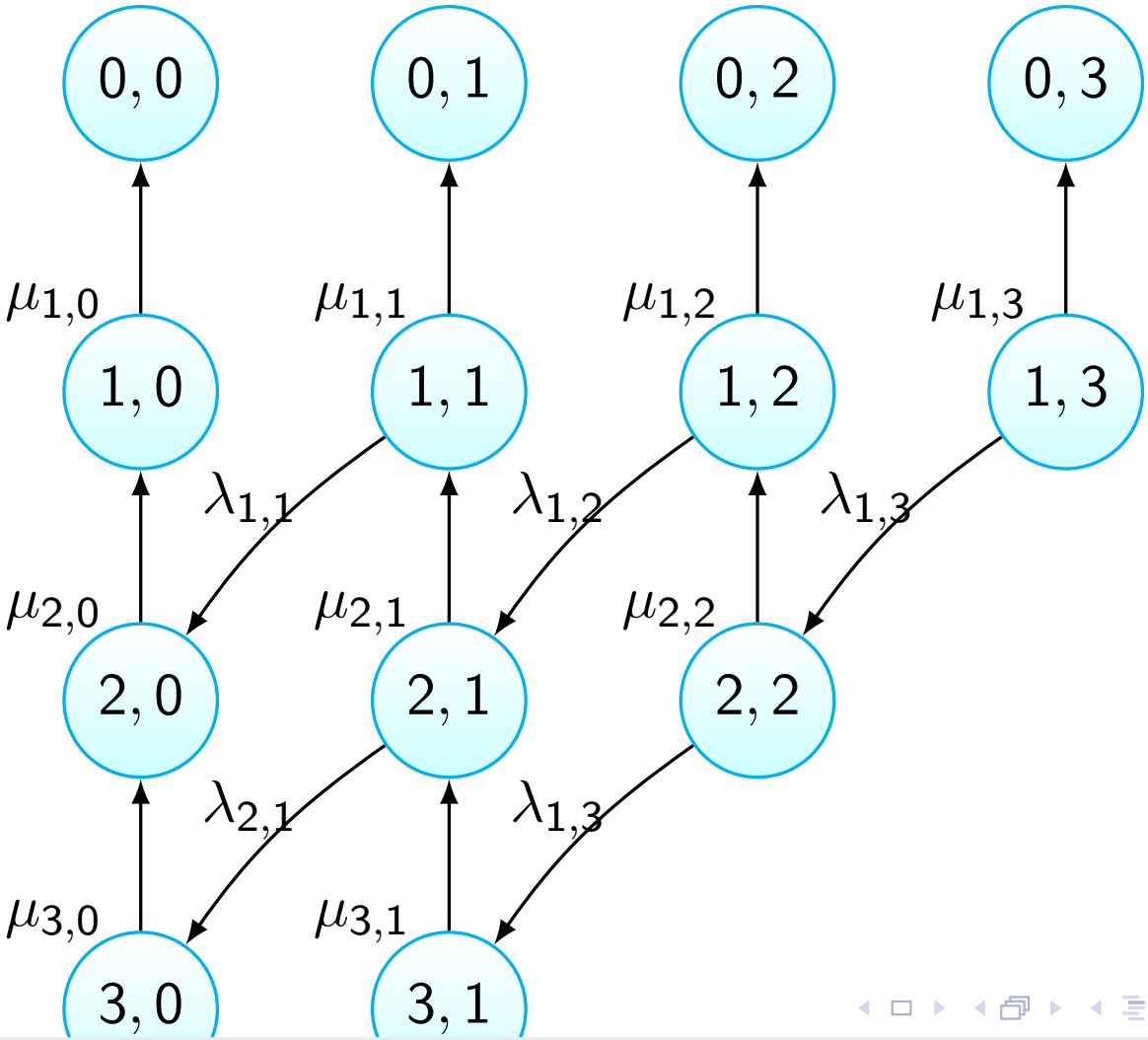
$$\tau = \inf\{t > 0 : I(t) = 0\}.$$

- m: total number of recovered individuals at τ
- To determine the final size distribution we need:

$$P_m(i, s) = \Pr\{R(\tau) = m \quad | \quad (I(0), S(0)) = (i, s)\}.$$

For example: $P_N(1, N-1)$ is the probability that, from a single infected, finally all individuals are infected.

State Space Diagram for SIR with $N = 4$



First Step Analysis for Final Size, m fixed

$$P_m(i,s) = \frac{\lambda_{i,s}}{\lambda_{i,s}+\mu_i} P_m(i+1,s-1) + \frac{\mu_i}{\lambda_{i,s}+\mu_i} P_m(i-1,s).$$

$$P_m(i,s) = \frac{\lambda s}{\lambda s+\mu N} P_m(i+1,s-1) + \frac{\mu N}{\lambda s+\mu N} P_m(i-1,s). \quad (1)$$

- **Absorbing states:** $(0,s)$ for $s = 0,1,\ldots N$

First Step Analysis for Final Size, continued

- **Boundary Condition for 0 infected**

$$P_m(0, s) = 1 \quad \text{for } s = N - m$$
$$P_m(0, s) = 0 \quad \text{for } s < N - m.$$

- **Boundary Condition for $N - m$ susceptible**

$$\begin{aligned} P_m(i, N-m) &= \frac{\mu N}{\lambda(N-m) + \mu N} P_m(i-1, N-m) \\ &= \left(\frac{\mu N}{\lambda(N-m) + \mu N} \right)^i \end{aligned} \tag{2}$$

Algorithm 1 Final Size Distribution for Exponential Disease Time

1: *Set* $P_m(0, s) = 0$ *for* $s < N - m$
2: *Set* $P_m(0, N - m) = 1$
3: **for** i=1,2,. . . ,m **do**
4: *Compute* $P_m(i, N - m)$ *using*(2)
5: **end for**
6: *Set* $s = N - m + 1$ # Start of outer for-loop
7: **for** i=1,..N-s **do**
8: *Compute* $P_m(i, s)$ *using*(1)
9: **end for**
10: *Set* $s = s + 1$. If $s \leq N - 1$ go to step 7. Otherwise, stop.

Assumption that the infection process is memoryless: OK

recovery process memoryless $\iff$ disease length $\sim$ exponential
this assumption NOT OK !!

Better to use Erlang disease times with shape parameter k.

- Member of a versatile class of distributions
- Allows to select different coefficients of variation.
- Can be formulated as Markov process with $k+1$-dimensional state space.
- (similar approach available for mixture of two Erlangs)

Model Definition

- instead of state space (I, S) we need
- $\Omega = (\tilde{I}(t), S(t))$, where $\tilde{I}(t)$ is a vector holding k stages.
- $\tilde{I}(t) := \{(i_1(t), i_2(t), i_3(t), ...i_k(t)) : \sum i_n \leq N \quad for \quad n \leq k\}$
- Define $I(t)$ as

$$I(t) = \sum_{n=1}^{k} i_k$$

Computational Efficiency

- First step analysis again allows to find the final outbreak size distribution. (Black, Ross 2015)
- Computational effort and memory requirements proportional to $k\,N^{(k+1)}$
- However, possible to use as state variable only the sum of all infected stages.
  $\Rightarrow$ Computational effort and memory requirements proportional only to kN^2 (İşilier, Güllü, H)

Table: Time Required to Calculate Exact Final Outbreak Size Distribution

	time to perform calculation with population size N		
k	100	500	1000
2	2.02 s	240.52 s	1886.14 s
5	5.10 s	614.49 s	4828.98 s
10	10.45 s	1271.74 s	10228.08 s

Basic reproduction Number R_0 for SIR

- **Basic Reproduction Number** R_0**:** expected number of secondary cases that one case would produce in an entirely susceptible population
- $R_0 > \mathbf{1}$ $\rightarrow$ Positive probability of an outbreak
- $R_0 < \mathbf{1}$ $\rightarrow$ An outbreak not possible

For the above Markov model $R_0 = \lambda/\mu$

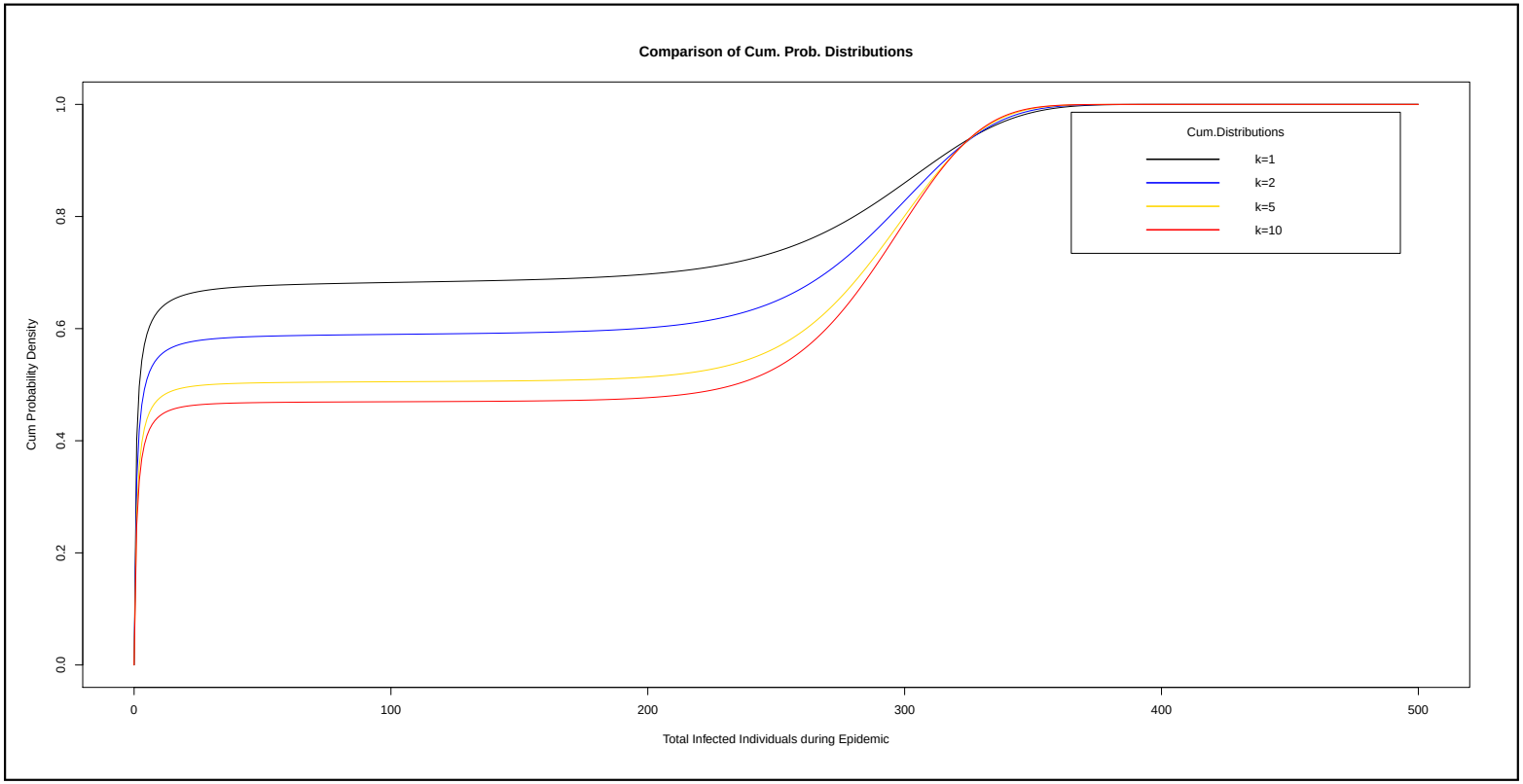


Figure: CDF of total number of infected for $R_0 = 1.5$, $N = 500$ and $k = 1, 2, 5, 10$

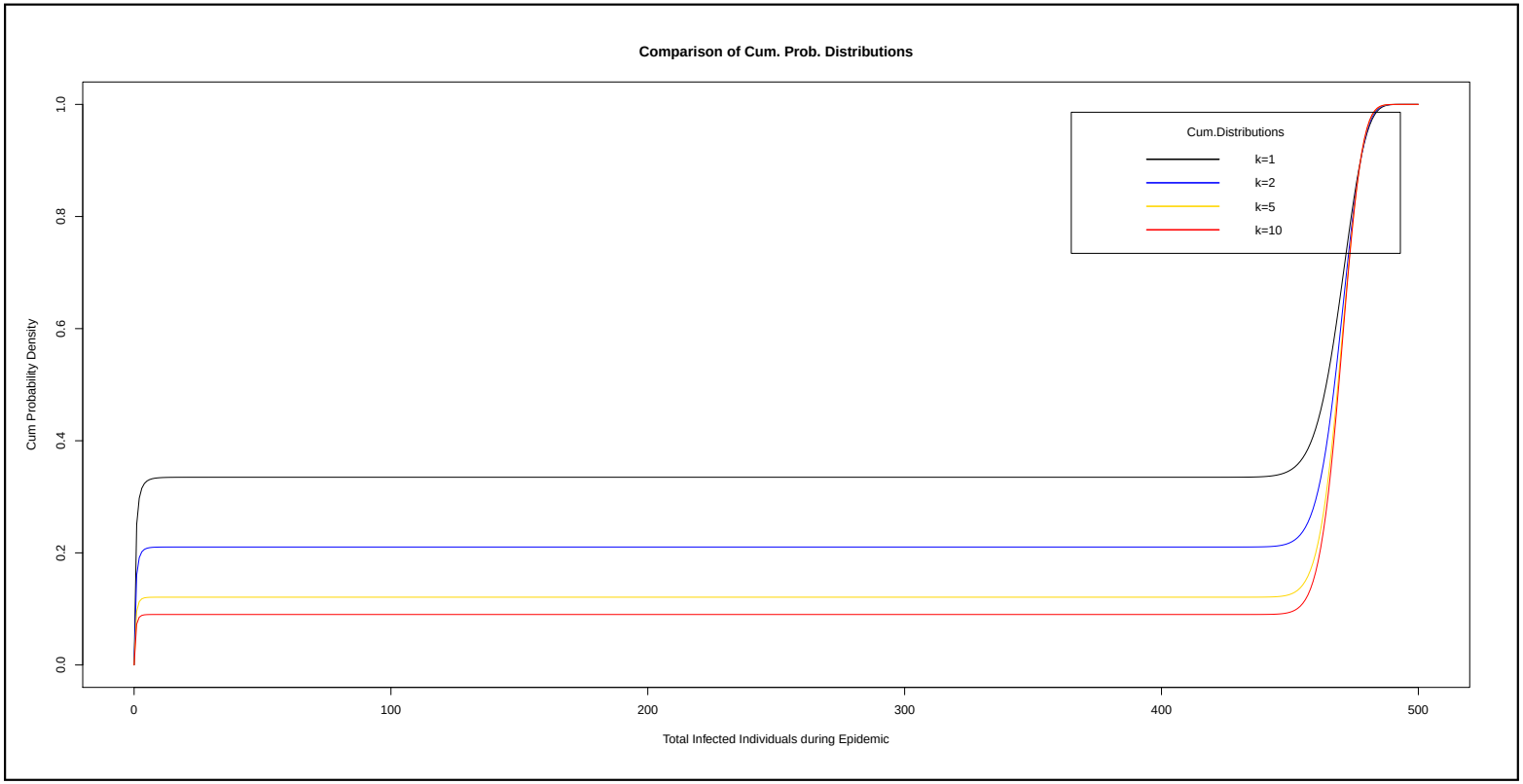


Figure: CDF of total number of infected for $R_0 = 3$, $N = 500$ and $k = 1, 2, 5, 10$

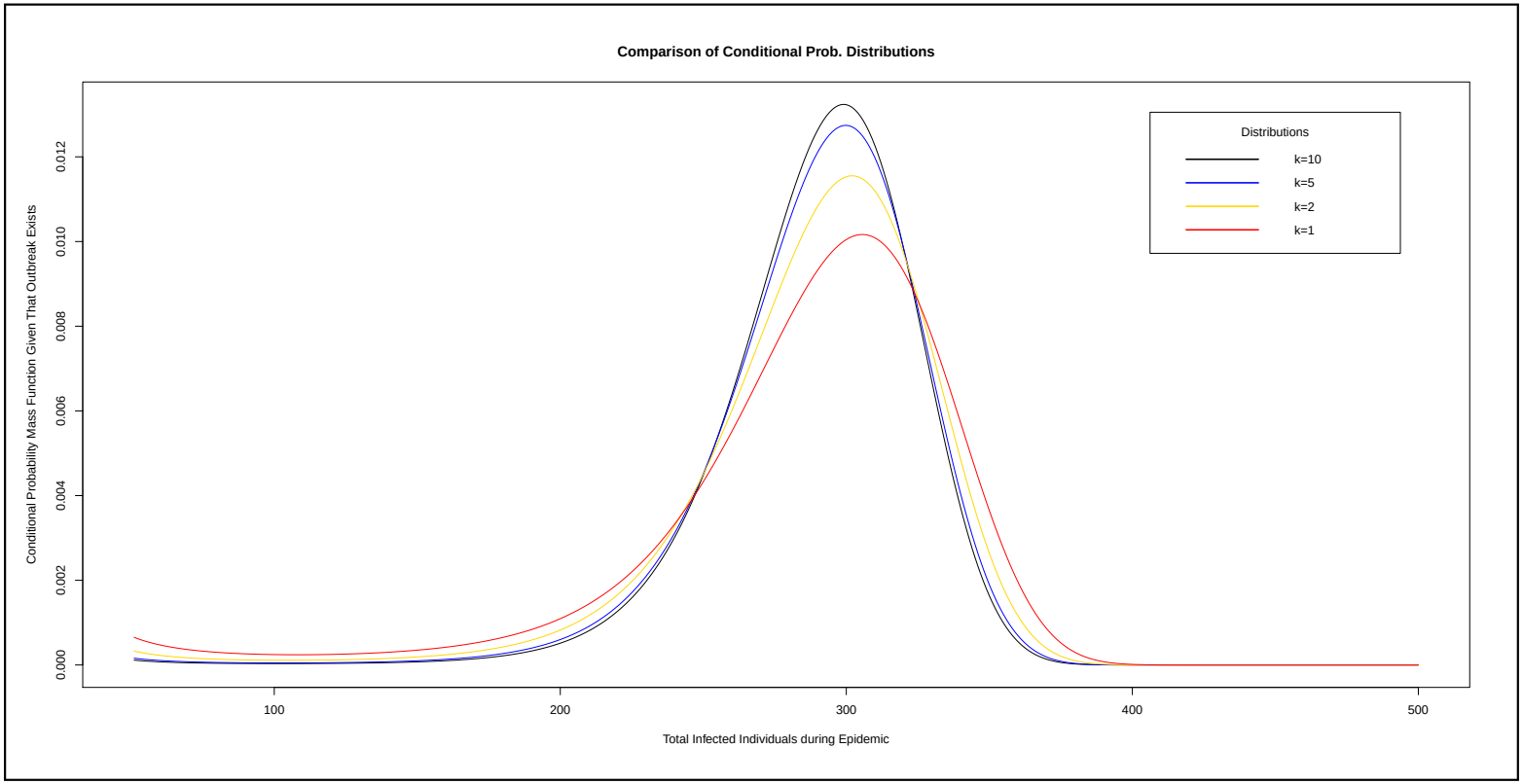


Figure: PDF of total number of infected for $R_0 = 1.5$, $N = 500$ and $k = 1, 2, 5, 10$

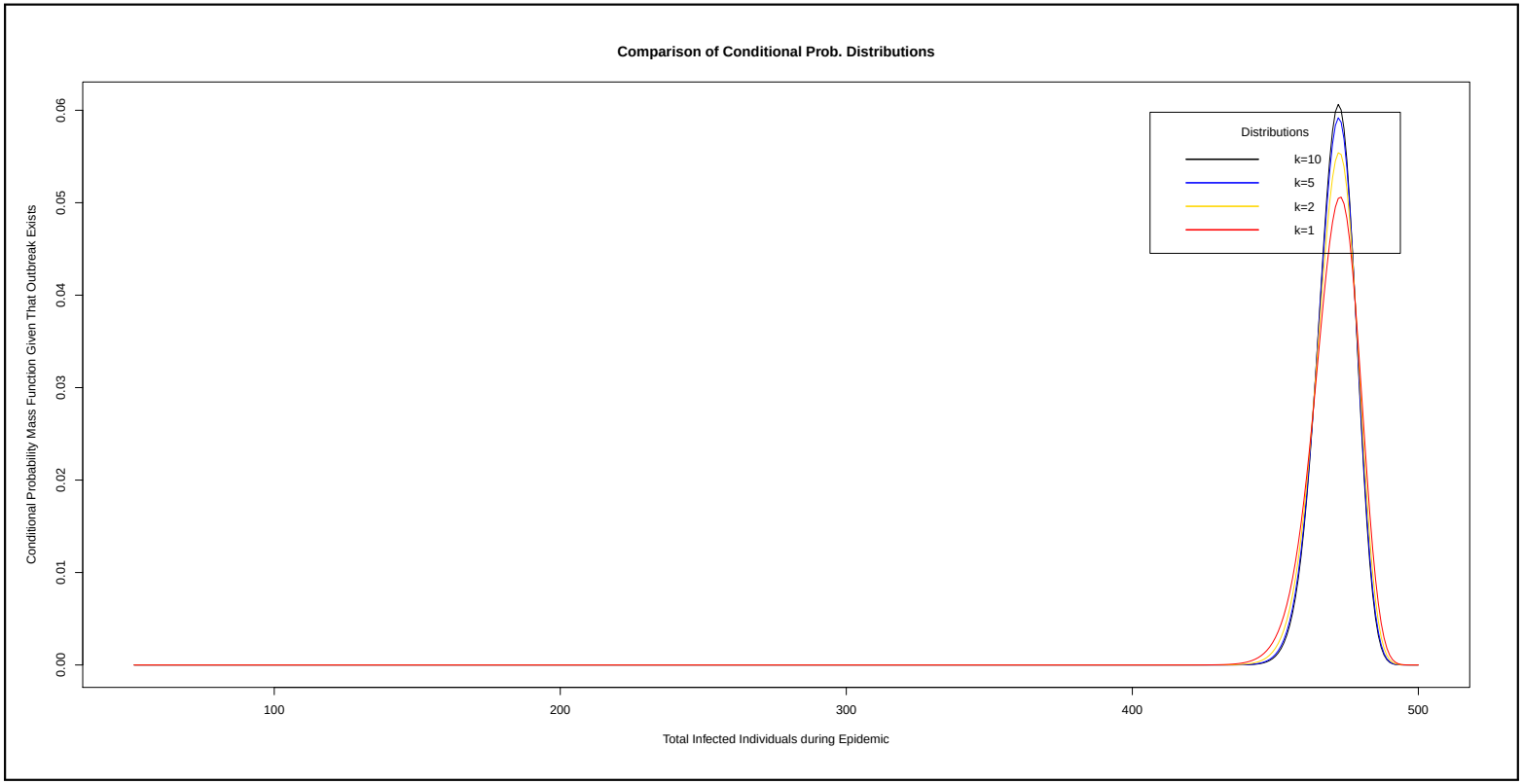


Figure: PDF of total number of infected for $R_0 = 3$, $N = 500$ and $k = 1, 2, 5, 10$

Discrete time SIR model with homogeneous mixing

Called "Reed-Frost" or "Binomial chain" model in epidemiology.

- Number of newly infected on day t,
  $NI_t \sim$Binomial$(n = S_t, \theta = 1 - (1 - p)^{I_t})$,
  where p denotes the probability that a single infected individual infects a single susceptible.
- Simplest model assumes deterministic disease time of one time step. This allows the recursive calculation of the joint distribution of (I_t, S_t).
- It is also possible to use a discrete disease length distribution. This increases the dimension of the state space and makes thus the use of recursions impracticable.

R_0 for homogeneous SIR model with random disease time

- R_0 expected number of secondary cases in a population with a single infected and $N-1$ susceptible
- probability p of a single susceptible to be infected is

$$p = \sum_d (1-(1-p_{day})^d) f_D(d) \quad \text{and} \quad R_0 = p\,(N-1)\,,$$

where $f_D(d)$ denotes the pmf of the (discrete) disease length and p_{day} the infection probablity for a single day.

Discrete time non-homogeneous model

For **disease spread**: SIR model for agent based simulation with overlapping mixing groups.
Usually a discrete disease time distribution is used.

For **information spread** (eg. in social networks):
Independent cascade models.
Constant "active" time is assumed to be equal to one time-step.

model parameters: infection probabilities p_{ij}.

Model: general infection structure; disease length = 1

parameters: infection probabilities p_{ij}.

- at $t = 0$: sets $\mathscr{I}_0$ (infected) and $\mathscr{S}_0$ (susceptible).
- Recursion:
  in set $\mathscr{S}_t$ new infections occur according to independent Bernoulli trials with infection probabilities

$$p_j = 1 - \prod_{\forall\, i \in \mathscr{I}_t} (1 - p_{ij}) \qquad \text{for } j \in \mathscr{S}_t$$

$\mathscr{I}_{t+1}$ is the set of all these newly infected.
$\mathscr{S}_{t+1} = \mathscr{S}_t \setminus \mathscr{I}_{t+1}$ *(remove newly infected from susceptible set)*
$\mathscr{R}_{t+1} = \mathscr{R}_t \cup \mathscr{I}_t$ *(add old infected to recovered)*

Agent based models with Overlapping Mixing Groups

- special case of the general model defined above
- assumes a structure of the matrix p_{ij}, that allows an easy calibration to real data, especially to census data.
- **Typical Mixing Groups:**
  Community, Neighborhood, Family, School/Work
- popular in the literature

- **Disadvantage**
  Literature claims: Even for R_0 no closed form formula available
  - simulation is necessary for all epidemiological quantities
  - flu spread models for huge populations (eg. for whole USA)

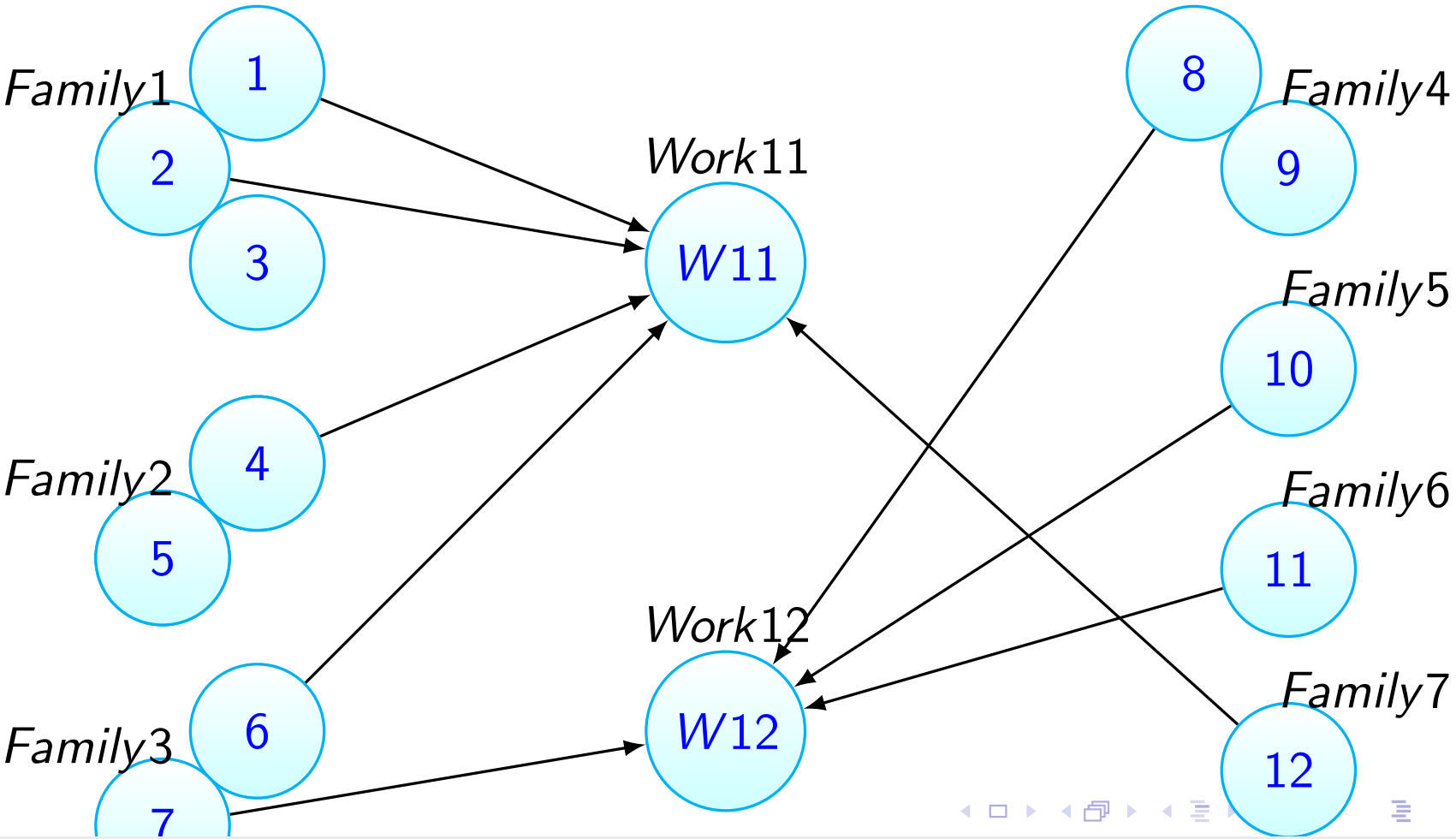
Overlapping Mixing Groups

- The infection probabilities p_j depend on the number of infected individuals in the respective mixing groups
- The probability that individual j with I_c, I_n, I_w, and I_f infected individuals in his community, neighborhood, work and family is infected:

$$p_j = 1 - (1 - p_c)^{I_c}(1 - p_n)^{I_n}(1 - p_w)^{I_w}(1 - p_f)^{I_f}$$

- The total number of newly infected is the sum of independent not identically distributed Bernoulli variates

Example A: a Small Population



Probabilities for Example A:

Probability in Family	Probability in Work	Probability in Community	Disease Duration
0.2	0.1	0.05	1

$$p_j = 1 - (1 - p_c)^{l_c}(1 - p_n)^{l_n}(1 - p_w)^{l_w}(1 - p_f)^{l_f} \qquad (3)$$

- Case: Only Individual 1 is initially infected.

$$p_{12} = 1 - (1 - 0.05)^1(1 - 0.1)^1(1 - 0.2)^1 = 0.316 \qquad (4)$$

p_{1j} for $i = 1$ is the only initially infected

j	Family Number	Work Number	Community	Mixing Groups	p_{1j}
2	1	11	100	F-W-C	0.316
3	1	-	100	F-C	0.240
4	2	11	100	W-C	0.145
5	2	-	100	C	0.05
6	3	11	100	W-C	0.145
7	3	12	100	C	0.05
8	4	12	100	C	0.05
9	4	-	100	C	0.05
10	5	12	100	C	0.05
11	6	12	100	C	0.05
12	7	11	100	W-C	0.145
					$\sum$ =1.291

$\Rightarrow$ expected number of people infected by $i = 1$ is $R_0(1) = 1.291$

For non-homogeneous case: "Individual $R_0(i)$"

$$R_0(i) = \sum_{j \neq i} p_{ij}$$

ID	Family Number	Work Number	Community	$R_0(i)$
1	1	11	100	1.291
2	1	11	100	1.291
3	1	-	100	0.93
4	2	11	100	1.12
5	2	-	100	0.74
6	3	11	100	1.12
7	3	12	100	1.025
8	4	12	100	1.025
9	4	-	100	0.74
10	5	12	100	0.835
11	6	12	100	0.835
12	7	11	100	0.93

How to define population-R_0?

Agent based simulation literature suggests to use as "population-R_0"
the average of all individual R_0-values.
Not sensible !!! can lead to outbreaks for $R_0 < 1$.

- necessary to interpret all individual R_0 values
- **outbreak impossible only for** $\max_i R_0(i) < 1$.

Intervention Strategies

- Vaccination
- Quarantine
- Antiviral Drugs

Literature: Assesses intervention strategies using simulation.

- The calculation of individual R_0-values allows to assess many intervention strategies by comparing the change the intervention implies for the individual R_0 values.
- Check if $\max_i R_0(i) < 1$ to identify interventions that guarantee that outbreaks are impossible.

Example B: Population for Intervention Analysis

Example of Longini et al. (2004)
four neighborhoods of 500 people; each neighborhood with:

- 100 singles
- 50 families of size two
- 34 families of size three
- 37 families of size four
- 5 families of size five
- 3 families of size six
- Single family of size seven
- Disease length 2, 3, 4, 5, and 6 with probabilities 0.21, 0.19, 0.18, 0.22, and 0.2.

Assessing Vaccination Strategies

- Vaccination before the start of the infection
- assumption: All individuals develop immunity

Sequential optimal vaccination strategy:
vaccinate person with **maximal individual R_0-value**.

We compare the distribution of the individual R_0 for

- no vaccination
- random vaccination of 50% of the population
- optimal vaccination of 50% of the population

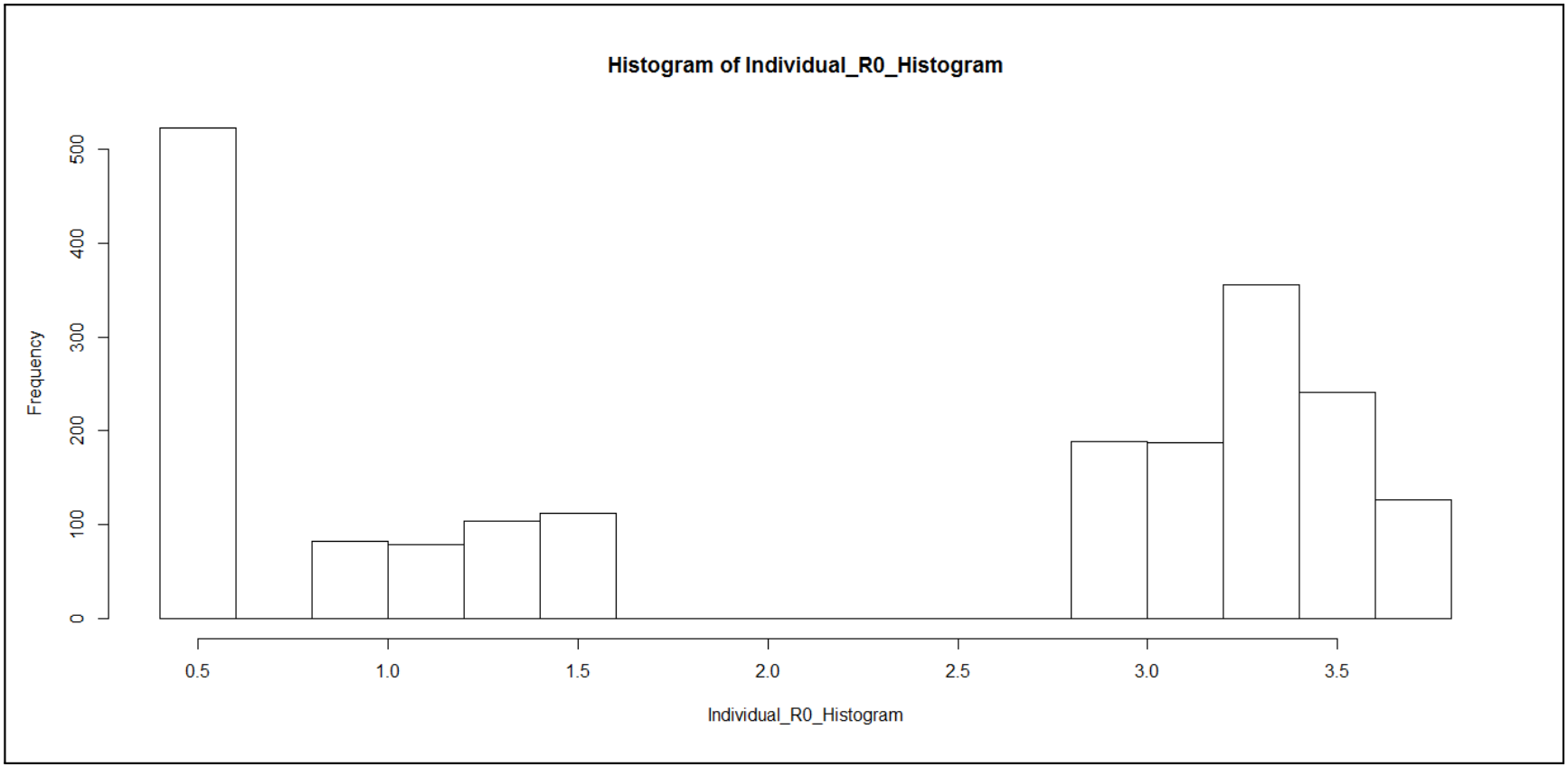


Figure: Individual R_0; case without vaccination.

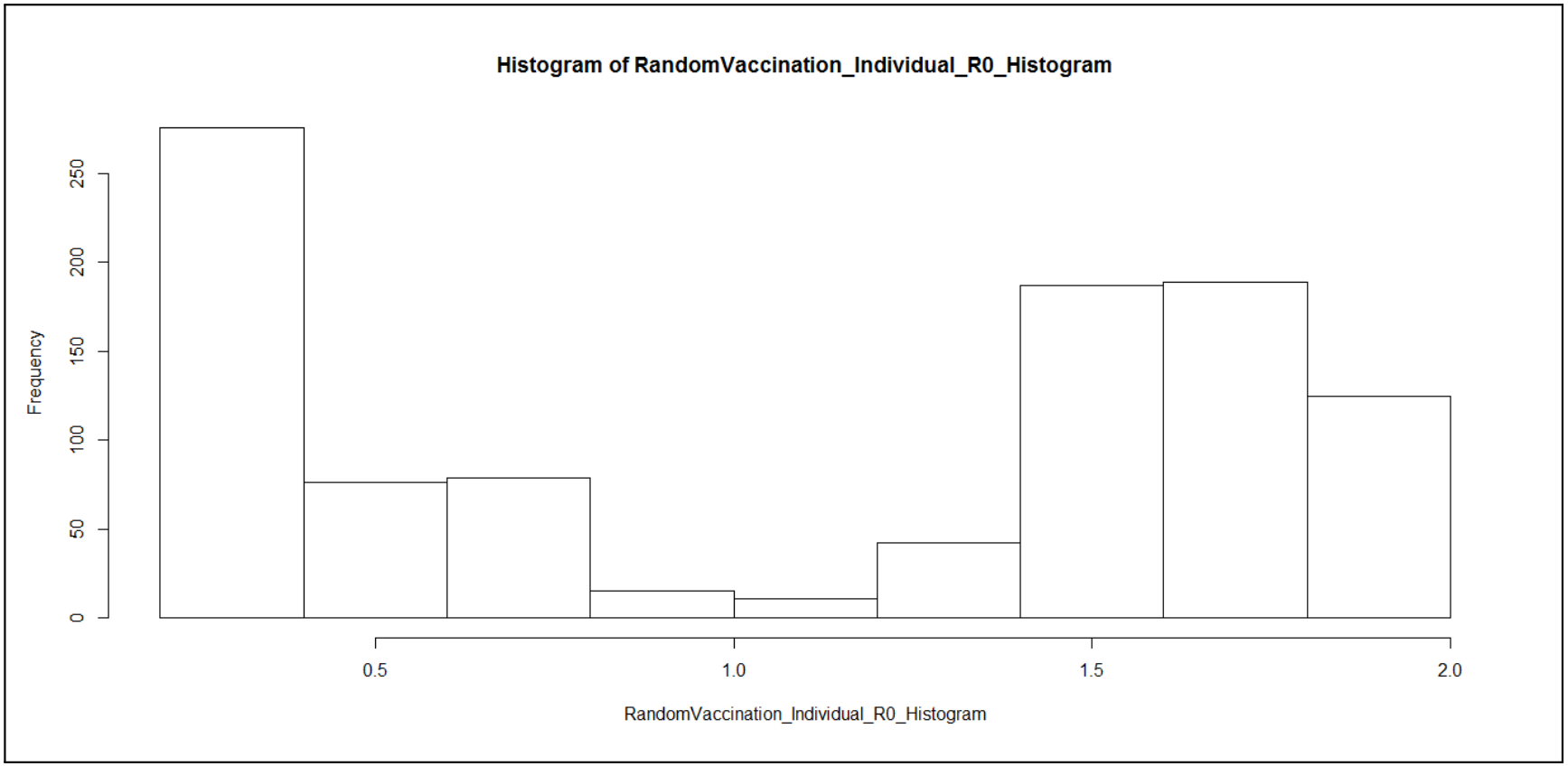


Figure: Individual R_0; case random vaccination of 50%.

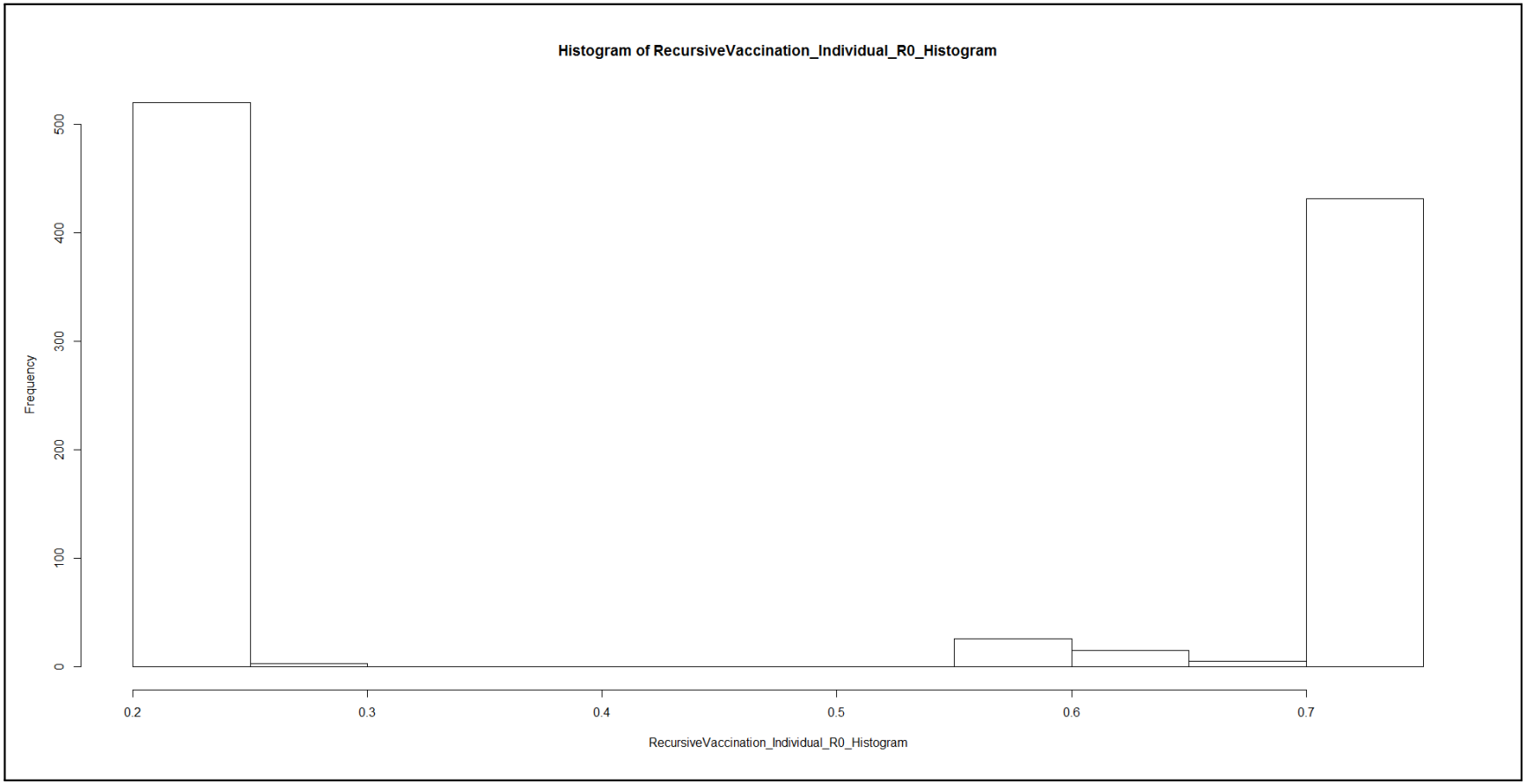


Figure: Individual R_0; case optimal vaccination of 50%.

Assessing Household Quarantine

- Infected individuals stay home after the first day; they have no contacts outside of their family.
- Not very effective because disease spread on the first day of infection is possible.
  In addition in the spread of influenza the disease is also spread by people who are infected but have no symptoms of flu.

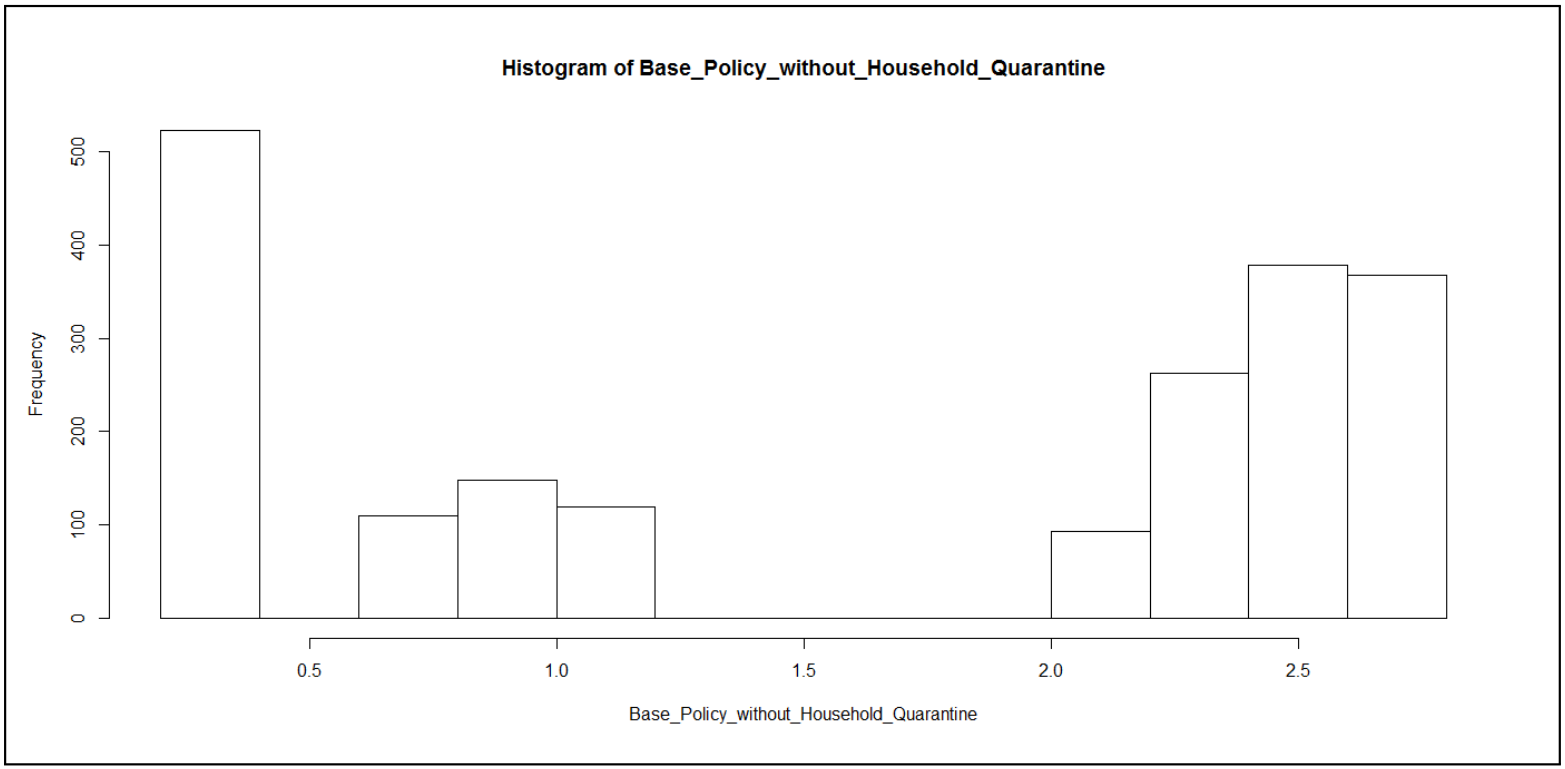


Figure: Household quarantine of 30% of all infected

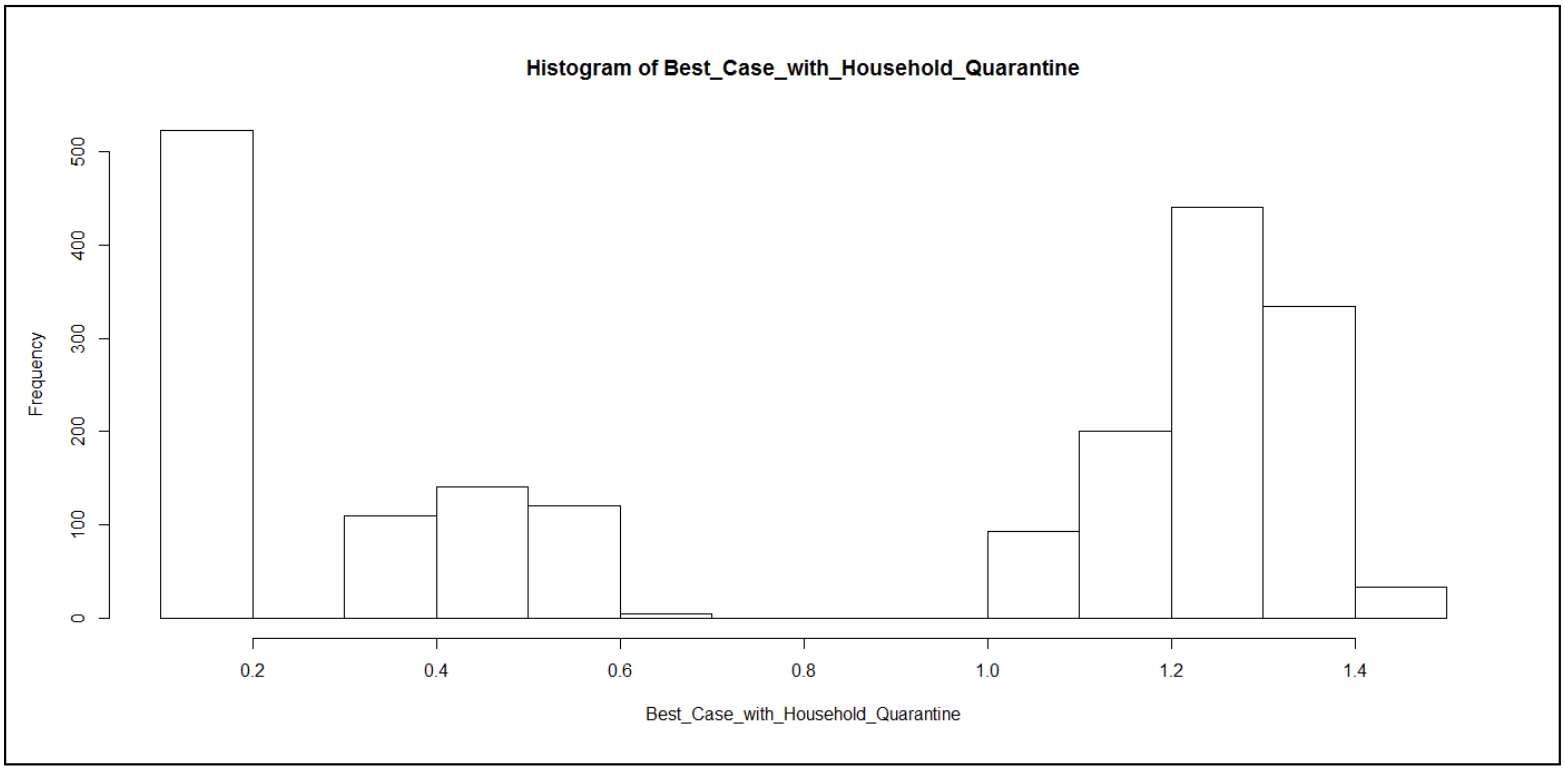


Figure: Household quarantine of 60% of all infected

Assessing the Use of Antiviral Drugs

- The use of antiviral prevents infection given exposure
- Reduces
  1) probability of transmission to others given infection
  2) probability of being infected given exposure
- Not very effective as they are not assumed to make infection impossible

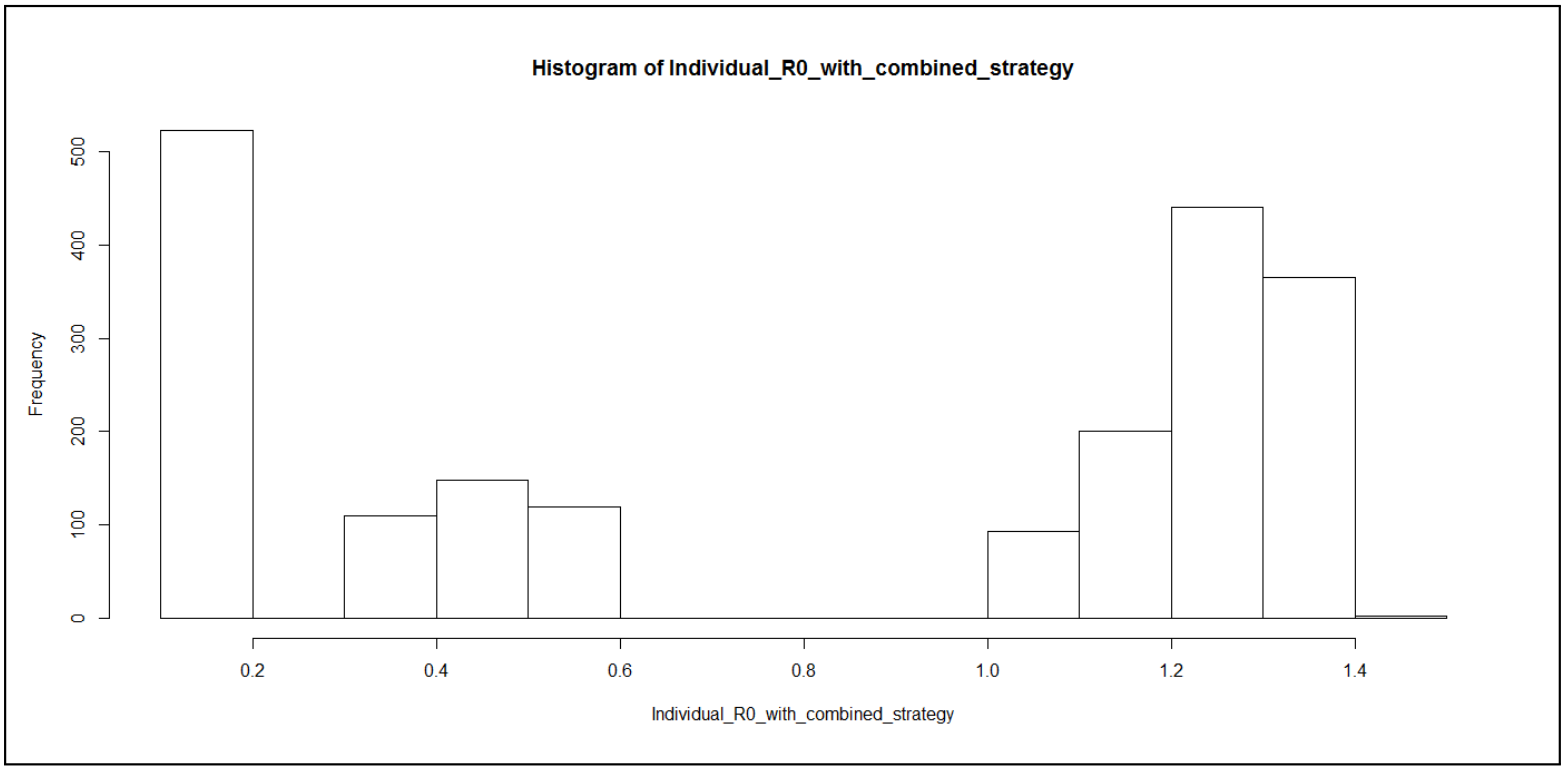


Figure: R_0 distribution for combining antiviral drugs and quarantine

THANK YOU..