Selective Isolation

Wouter van Amsterdam

3/30/2020

## Introduction

The COVID-19 virus is infecting people at an exponential pace, overwhelming hospital resources and causing many fatalities. Across multiple countries, there is a pronounced age difference in the hazard of dying. We try to model the consequences of a selective isolation approach, where people with higher risk of dying (specifically >65 years old) are put in strict isolation, while the rest of the population is asked to behave and move as normal. This way, the low-risk population may build up the required herd-immunity, possibly preventing many deaths and intensive care usage.

## Materials and Methods

We base our modelling on the Dutch situation. Our primary data-sources are:

* the daily report by the Dutch institute for health (RIVM) on the number of known infections, hospitalizations and deaths, due to COVID-19.
* the Dutch demographic build up (source: <https://opendata.cbs.nl/statline/>)

These enter our analysis as two data-files:

# get the (curated) data  
agetable <- fread(here::here('data', 'agetable.csv'), na.strings = '')  
dems <- fread(here::here('data', 'demographics.csv'), na.strings='')  
print(head(agetable))

## Leeftijdsgroep Totaal % Ziekenhuisopname % Overleden %  
## 1: 0-4 38 ( 0.3) 23 ( 0.6) 0 ( 0.0)  
## 2: 5-9 8 ( 0.1) 1 ( 0.0) 0 ( 0.0)  
## 3: 10-14 33 ( 0.3) 4 ( 0.1) 0 ( 0.0)  
## 4: 15-19 60 ( 0.5) 8 ( 0.2) 0 ( 0.0)  
## 5: 20-24 229 ( 1.9) 10 ( 0.3) 0 ( 0.0)  
## 6: 25-29 488 ( 4.2) 36 ( 0.9) 0 ( 0.0)  
## agegrp\_min agegrp\_max  
## 1: 0 4  
## 2: 5 9  
## 3: 10 14  
## 4: 15 19  
## 5: 20 24  
## 6: 25 29

print(head(dems))

## Geslacht Leeftijd Burgerlijke staat Perioden Bevolking  
## 1: Mannen 0 tot 5 jaar Totaal burgerlijke staat 2019 443658  
## 2: Mannen 5 tot 10 jaar Totaal burgerlijke staat 2019 470233  
## 3: Mannen 10 tot 15 jaar Totaal burgerlijke staat 2019 489531  
## 4: Mannen 15 tot 20 jaar Totaal burgerlijke staat 2019 538304  
## 5: Mannen 20 tot 25 jaar Totaal burgerlijke staat 2019 548282  
## 6: Mannen 25 tot 30 jaar Totaal burgerlijke staat 2019 569071  
## agegrp\_min agegrp\_max Leeftijdsgroep  
## 1: 0 5 0-4  
## 2: 5 10 5-9  
## 3: 10 15 10-14  
## 4: 15 20 15-19  
## 5: 20 25 20-24  
## 6: 25 30 25-29

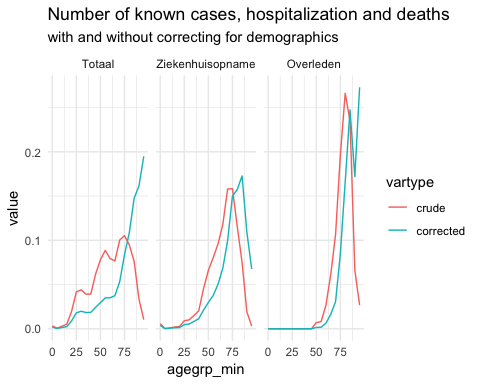
Where agetable contains the age-specific cases, hopitalizations and deaths. dems has the count of people

First we calculate the proportion of each variable (population, cases, etc) in each age group

## normalize   
dems[, Bevolking\_prop:=Bevolking / sum(Bevolking)]  
### make one with both sexes  
grpvars <- c('agegrp\_min', 'agegrp\_max', 'Leeftijdsgroep')  
demsmf <- dems[,list(  
 Bevolking=sum(Bevolking),  
 Bevolking\_prop=sum(Bevolking\_prop)  
), by = grpvars]  
# demsmf[, plot(agegrp\_min, Bevolking)]  
  
# make a long version of the data  
## remove missing age group (is tiny and not relevant for relative rates)  
grpvars <- c('agegrp\_min', 'agegrp\_max', 'Leeftijdsgroep')  
agem <- melt(agetable[!is.na(Leeftijdsgroep)], id.vars=grpvars,  
 measure.vars=c('Totaal', 'Ziekenhuisopname', 'Overleden'),  
 value.name='casecount')  
  
## calculate proportions  
agem[, prop:=casecount / sum(casecount), by = c('variable')]

There are no reliable sources on the age-specific risk of getting infected with the COVID-19 virus, as all testing schemes are biased where worse symptoms increase the likelihood of being tested. For our analysis we assume a uniform risk of getting infected. This means every person of every age has the same risk of getting the infection. It is likely that this overestimates the risk of infection for older people, as they are generally less mobile. We then compare the number of deaths per age group, with and without standardizing for the demographic situation.

## standardize per demographics  
mvars <- c('Bevolking', 'Bevolking\_prop')  
### make incidencerates by merging data on age group  
agem[demsmf, incidencerate:=casecount / i.Bevolking, on='Leeftijdsgroep']  
### make normalized incidence rate (as prop of total)  
agem[, ir\_norm:=incidencerate / sum(incidencerate), by = 'variable']  
  
## plots  
  
### ecdf of age  
demsmf[order(agegrp\_min), Bevolking\_cum:=cumsum(Bevolking)]  
demsmf[, Bevolking\_cdf:=Bevolking\_cum / sum(Bevolking)]  
# demsmf[, plot(agegrp\_min, Bevolking\_cdf, type='l')]  
  
agem %>%  
 melt(id.vars=c('variable', 'agegrp\_min'),  
 measure.var = c('prop', 'ir\_norm'),  
 variable.name='vartype', value.name='value') %>%  
 mutate(vartype=factor(vartype, levels=c('prop', 'ir\_norm'),   
 labels=c('crude', 'corrected'))) %>%  
 ggplot(aes(x=agegrp\_min,y=value,group=vartype,col=vartype)) +   
 geom\_line() +   
 theme\_minimal() +   
 facet\_grid(~variable) +   
 ggtitle('Number of known cases, hospitalization and deaths',  
 'with and without correcting for demographics')



This graph shows that, when correcting for the demographic situation, the age effect is even more pronounced.

### Modeling

Next, we model a range of scenarios to account for uncertainty. Due to the problem of underreported infections, and possibly underreported deaths, the true marginal mortality rate is unknown. We consider situations for the true marginal mortality rate between 0.5% and 3.5%. The required proportion of the population required for herd-immunity is also unknown, though it depends on the R0 of the virus. We consider it to be between 60% and 80%. Moverover, we assume that everyone who comes in contact with the virus will get infected and will either develop lasting immunity, or die.

<<

We will consider two strategies: non-selective measures, or selective isolation, where we assume everyone under the age-limit (65) will get infected, and no one older will get infected.

As a first step, we need to check if the required herd-immunity can even be reached when only people under 65 can get infected.

## make some scenarios  
marginaldeathrates <- seq(0.005, 0.035, by=0.005)  
herdimmunityprops <- seq(0.6, 0.8, by=0.05)  
freeages <- c(65, 130)  
   
### look at minimum age to get at least this amount of the people infected  
Bevolking\_cdf\_fn <- approxfun(x=demsmf$agegrp\_max, y=demsmf$Bevolking\_cdf)  
Bevolking\_icdf\_fn <- approxfun(x=demsmf$Bevolking\_cdf, y=demsmf$agegrp\_max)  
maxfreeages <- Bevolking\_icdf\_fn(herdimmunityprops)  
print(data.table(herdimmunityprop=herdimmunityprops,  
 maxfreeage=maxfreeages))

## herdimmunityprop maxfreeage  
## 1: 0.60 50.02493  
## 2: 0.65 53.40866  
## 3: 0.70 56.85836  
## 4: 0.75 60.41165  
## 5: 0.80 64.35073

This seems to be just enough, even in the most extreme case of 80% required herd-immunity

Now we can continue with modeling the number of deaths for the different scenarios.

#### define settings  
settings <- CJ(  
 marginaldeathrate = marginaldeathrates,  
 herdimmunityprop = herdimmunityprops,  
 freeage = freeages,  
 dumkey = 1  
)  
settings[, setting\_idx:=1:.N]  
settings[, herd\_needed:=herdimmunityprop\*demsmf[, sum(Bevolking)]]  
settings[, herd\_avail:=demsmf[agegrp\_max<freeage, max(Bevolking\_cum)], by = 'setting\_idx']  
settings[, herd\_prop:=herd\_needed/herd\_avail]

We do not know the age-specific death rate. However, using the different marginal death rates and the observed death rates, we can infer the age specific death rates using a few simple calculations.

Where the proportion of the population in age group and the death rate in that age group We introduce a single factor R such that , and that . Then:

So now we have that

We procede to calculate and the corresponding for each marginal death rate

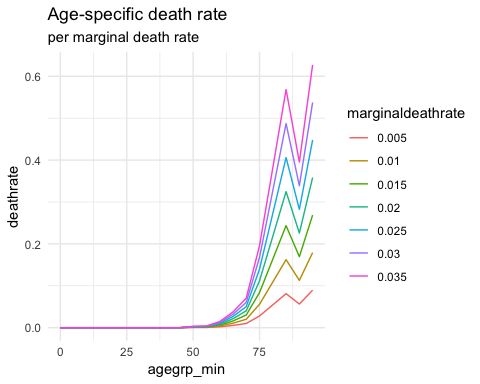
ir\_mu = agem[variable=='Overleden'][demsmf, sum(ir\_norm\*i.Bevolking\_prop), on='Leeftijdsgroep']  
settings[, Rfactor:=marginaldeathrate / ir\_mu]

Next we combine each possible scenario with the observed data and calculate the number infections and deaths per age group.

#### create outer tables for each setting  
agem[, dumkey:=1]  
dfs <- agem[variable=='Overleden'][settings, on='dumkey', allow.cartesian=T]  
  
#### calculate number infected  
dfs[, inherd:=agegrp\_max < freeage]  
dfs[demsmf, Bevolking:=i.Bevolking, on='Leeftijdsgroep']  
dfs[demsmf, Bevolking\_prop:=i.Bevolking\_prop, on='Leeftijdsgroep']  
dfs[, num\_infected:=herd\_prop\*Bevolking\*as.numeric(agegrp\_max<freeage)]  
  
##### calculate total number of deaths  
dfs[, deathrate:=Rfactor\*ir\_norm]  
###### make sure the marginal death rates are ok  
drs <- dfs[, list(obs\_marginaldeathrate=sum(Bevolking\_prop\*deathrate)), by=c('setting\_idx', 'marginaldeathrate')]  
stopifnot(drs[, all.equal(obs\_marginaldeathrate, marginaldeathrate)])  
dfs[, deaths:=deathrate\*num\_infected]  
  
### define a baseline (freeage==130)  
baseline <- dfs[freeage==130]  
dfs[baseline, deathdifference:=deaths - i.deaths,   
 on=c('marginaldeathrate', 'herdimmunityprop', 'Leeftijdsgroep')]  
dfs[, excessdeaths:=ifelse(deathdifference>0, deathdifference, 0)]  
dfs[, preventeddeaths:=ifelse(deathdifference<0, -deathdifference, 0)]  
dfs[, totaldeaths:=sum(deaths), by='setting\_idx']

We now plot the age-specific death rate for diffenent marginal death rates

### deathrate  
dfs %>%  
 .[!duplicated(paste0(marginaldeathrate, Leeftijdsgroep))] %>%  
 mutate(marginaldeathrate=factor(marginaldeathrate)) %>%  
 ggplot(aes(x=agegrp\_min, y=deathrate, grp=marginaldeathrate,  
 col=marginaldeathrate)) +   
 geom\_line() +   
 theme\_minimal() +   
 ggtitle('Age-specific death rate', 'per marginal death rate')



Now we can calculate the number of lives saved for the different situations

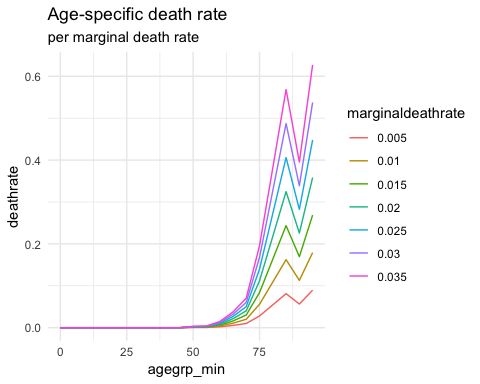
## tables  
totaldeaths <- dfs[, list(deaths=sum(deaths),   
 deathdifference=sum(deathdifference),  
 num\_infected=sum(num\_infected)),  
 by=c('marginaldeathrate', 'herdimmunityprop', 'freeage', 'setting\_idx')]  
# totaldeaths[order(deaths)]  
# dcast(totaldeaths, herdimmunityprop~freeage+marginaldeathrate, value.var='deaths')  
  
livessaved <- dcast(totaldeaths[freeage==max(freeage)], herdimmunityprop~marginaldeathrate, value.var='deaths') -   
 dcast(totaldeaths[freeage==min(freeage)], herdimmunityprop~marginaldeathrate, value.var='deaths')  
livessaved

## herdimmunityprop 0.005 0.01 0.015 0.02 0.025 0.03  
## 1: 0 49023.06 98046.12 147069.2 196092.2 245115.3 294138.4  
## 2: 0 53108.32 106216.63 159324.9 212433.3 265541.6 318649.9  
## 3: 0 57193.57 114387.14 171580.7 228774.3 285967.9 343161.4  
## 4: 0 61278.83 122557.65 183836.5 245115.3 306394.1 367673.0  
## 5: 0 65364.08 130728.16 196092.2 261456.3 326820.4 392184.5  
## 0.035  
## 1: 343161.4  
## 2: 371758.2  
## 3: 400355.0  
## 4: 428951.8  
## 5: 457548.6

deathranks <- totaldeaths[freeage==130]  
deathranks <- deathranks[order(deaths), deathrank:=1:.N]  
totaldeaths[deathranks, deathrank:=i.deathrank, on=c('marginaldeathrate', 'herdimmunityprop')]

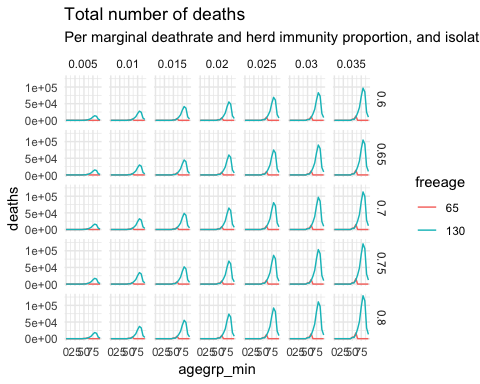
As we see, this ranges from 4.902310^{4} in the most optimistic scenario to 4.5754910^{5} in the most pessimistic scenario.

## plots  
### deathrate  
dfs %>%  
 .[!duplicated(paste0(marginaldeathrate, Leeftijdsgroep))] %>%  
 mutate(marginaldeathrate=factor(marginaldeathrate)) %>%  
 ggplot(aes(x=agegrp\_min, y=deathrate, grp=marginaldeathrate,  
 col=marginaldeathrate)) +   
 geom\_line() +   
 theme\_minimal() +   
 ggtitle('Age-specific death rate', 'per marginal death rate')



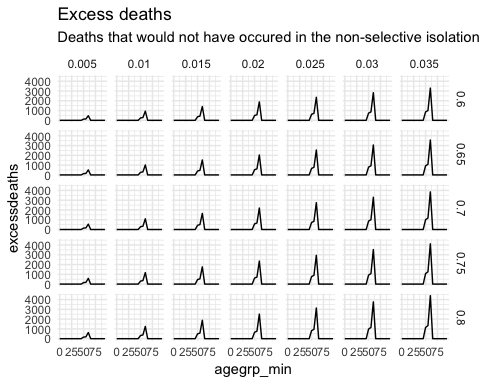
Our main result is the comparison between the number of deaths with and without selective isolation

dfs %>%   
 mutate(freeage=factor(freeage)) %>%  
 ggplot(aes(x=agegrp\_min, y=deaths, group=setting\_idx, col=freeage)) +   
 geom\_line() +   
 facet\_grid(herdimmunityprop~marginaldeathrate) +   
 theme\_minimal() +   
 ggtitle('Total number of deaths', 'Per marginal deathrate and herd immunity proportion, and isolation type (freeage)')



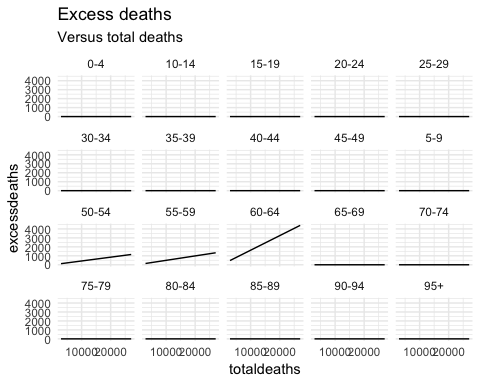
To model the possible hazards of the selective isolation approach, we specifically look at excess deaths in age groups <65 years, that would not occur in case of a non-selective approach.

dfs %>%   
 .[freeage<130] %>%  
 .[excessdeaths>=0,] %>%  
 mutate(freeage=factor(freeage)) %>%  
 ggplot(aes(x=agegrp\_min, y=excessdeaths, group=setting\_idx)) +   
 geom\_line() +   
 facet\_grid(herdimmunityprop~marginaldeathrate) +   
 theme\_minimal() +   
 ggtitle('Excess deaths', 'Deaths that would not have occured in the non-selective isolation method')



To put these excess deaths in perspective, we plot them against the total number of deaths in that scenario.

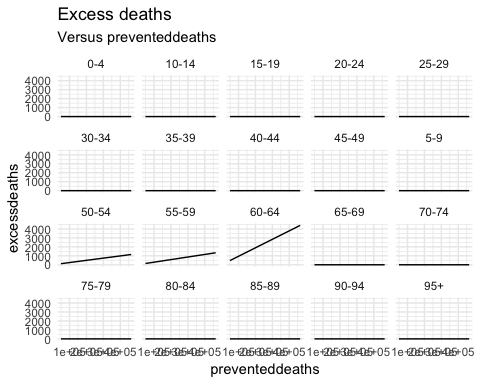
dfs %>%  
 .[freeage<130] %>%  
 ggplot(aes(x=totaldeaths, y=excessdeaths, grp=Leeftijdsgroep)) +   
 geom\_line() +   
 facet\_wrap(~Leeftijdsgroep) +   
 theme\_minimal() +   
 ggtitle('Excess deaths', 'Versus total deaths')



As expected, the 60-64 age group has the most excess deaths, with 471 excess deaths in the most optimistic scenario, and 4395 deaths in the most pessimistic scenario.

Next, we constrast these numbers with the number of lives saved.

dfs %>%  
 .[freeage<130] %>%  
 .[, preventeddeaths:=sum(preventeddeaths), by='setting\_idx'] %>%  
 ggplot(aes(x=preventeddeaths, y=excessdeaths, grp=Leeftijdsgroep)) +   
 geom\_line() +   
 facet\_wrap(~Leeftijdsgroep) +   
 theme\_minimal() +   
 ggtitle('Excess deaths', 'Versus preventeddeaths')



## Discussion

In our simple modeling of a selective isolation approach, we find that in the Netherlands between 4.902310^{4} and 4.5754910^{5} can be saved.

Our approach relies on (many) simplifying assumptions:

* uniform infection rate
* all herd gets infected and becomes immune (adapt to allow for deaths)
* perfect intervention possible

Not modeled:

* improved outcomes for non-COVID-19 patients due to less strain on hospital system.