Italy is routinely and correctly described as particularly vulnerable to COVID-19 because of its older age profile. I set out to understand for myself how important this factor is. What would happen if the case fatality rates observed in Italy were applied to demographic profiles of other countries?

**Fatality rates by age and sex so far in Italy**

It’s worth pointing out that the snapshots presented in these bulletins change fast, including the raw fatality rate (for both sexes) which has increased from 5.8% seven days earlier to 8.5% on 19 March. Further rapid change is to be expected, remembering that deaths lag beginning of the illness by days or weeks, and diagnoses lag infections by days

It’s also worth pointing out how much worse this disease seems to be for men. Of the deaths in Italy at the time of this bulletin, 71% were men. Of diagnosed cases, 59% were male (more than 200 Italian boys have the illness but none had died at the time of the bulletin). There were more male fatalities aged 80 and over than female of all ages. Also, it’s worth pointing out that while it is definitely *worse* for older people, fatality rates are pretty bad for middle-aged people – about 1% for those between 30 and 59. That’s bad for a disease expecting as many cases as this one.

**Population profiles in selected countries**

I took population breakdowns by age and sex from the United Nations’ World Population Prospects. To illustrate I chose nine countries representing a range of cultural and economic situations. I’ve chosen to present these as density charts, not population pyramids (which I find difficult to make comparisons with). We can readily see the contrast between Italy and (for an extreme example) economically poor Timor Leste:

**Applying fatality rates to population profiles**

It’s straightforward to take a country’s population and apply the Italian case fatality rates to it to get a weighted average fatality rate. In effect, this tells us what the fatality rate would be in a country, if the Italian rates applied to its whole population or a subpopulation that was representative of the overall age and sex balance. Here’s what we get for our nine ‘countries’ (including the World aggregate):

Two things stand out.

First, the different demographics of the different countries make a *huge* difference. On these sorts of age-based rates, Italy can expect twice the fatality rate of China (and nearly five times that of Timor Leste).

Second, the death rate for Italy from this method is much lower than the actual fatality rate in the 19 March bulletin – 3.9% compared to 8.5%. This isn’t a mistake – it comes about because the profile of Italians diagnosed with COVID-19 is older and more male than Italians in general.

Older people and men are not just more likely to die if they get COVID-19, they are also more likely to be diagnosed with it in the first place.

As I note on the graphic, this could be due to women and younger people of either sex being less likely to be diagnosed given they have the disease; or it might mean they are less likely to have the disease at all. There is no way to tell with this data.

We can adjust the fatality rates by scaling them up to match Italy’s 19 March observed level. This gives a more realistic but still very rough answer to the question “what would Italy’s case fatality rates mean, translated to other countries”. It’s very rough because doing this assumes away a whole bunch of possible complexities and interactions between variables, but it’s probably as thorough a method as is warranted at the moment with the fast changing data. Here’s those scaled results:

**What does it all mean?**

Well, the danger to people over 50, particularly but not only men, is very very real from this disease. And the age profiles of countries vary enough for this to make big differences to the overall impact.

But regardless of this, the necessary actions are clear. Work hard to avoid getting this horrible disease and to avoid passing it on. Work to help others do the same, and pull together to manage society through some difficult months ahead. Wash your hands and practice social distancing.

Here’s the code behind those charts. The Italian data is just entered by hand because it’s only 20 numbers, not worth trying to automate.

#------------setup---------------

# 59% cases male

20686 / (20686 + 14378)

# 71% deaths men (no boys)

2139 / (2139 + 890)

library(tidyverse)

library(scales)

library(wpp2019)

# colours for male and female used by Washington Post 2017; see <https://blog.datawrapper.de/gendercolor/>

sex\_cols <- c(Male = "#F4BA3B", Female = "#730B6D")

#---------------------Italian fatality rates---------

italy\_rates <-tibble(

age\_grp = rep(c('0-9', '10-19', '20-29', '30-39', '40-49', '50-59', '60-69', '70-79', '80-89', '90+'), 2),

sex = rep(c("Male", "Female"), each = 10),

cfr = c(0, 0, 0, 0.6, 0.7, 1.7, 6.0, 17.8, 26.4, 32.5,

0, 0, 0, 0.2, 0.4, 0.6, 2.8, 10.7, 19.1, 22.3) / 100,

age\_midpt = rep(c(5, 15, 25, 35, 45, 55, 65, 75, 85, 95), 2)

)

italy\_rates %>%

ggplot(aes(x = age\_midpt, y = cfr, colour = sex)) +

geom\_point() +

geom\_text(data = filter(italy\_rates, cfr > 0.01),

aes(label = percent(cfr), y = cfr + 0.012), size = 3) +

geom\_line() +

scale\_x\_continuous(breaks = italy\_rates$age\_midpt, labels = italy\_rates$age\_grp) +

scale\_y\_continuous(label = percent\_format(accuracy = 1)) +

scale\_colour\_manual(values = sex\_cols) +

theme(panel.grid.minor = element\_blank(),

panel.grid.major.x = element\_blank()) +

labs(x = "Age group", colour = "", y = "Observed case fatality rate",

title = "Observed fatality rate of diagnosed COVID-19 cases in Italy to 19 March 2020",

subtitle = "20,686 men and boys with case fatality rate of 10.3%; 14,378 women and girls with case fatality rate of 6.2%",

caption = "Source: Istituto Superiore di Sanità, Roma")

#----------------Population rates ------------------

data(popF)

data(popM)

selected\_countries <- c("Australia", "Italy", "Timor-Leste", "United States of America", "World",

"China", "Brazil", "Japan", "Germany")

age\_lu <- tibble(age = unique(popF$age),

age\_grp = c(rep(unique(italy\_rates$age\_grp), each = 2), "90+")) %>%

mutate(age\_grp = factor(age\_grp, levels = unique(age\_grp)))

# Visual check that this shorthand worked ok

# View(age\_lu)

pop\_2020 <- popF %>%

mutate(sex = "Female") %>%

rbind(mutate(popM, sex = "Male")) %>%

select(country = name, age, pop = `2020`, sex) %>%

left\_join(age\_lu, by = "age") %>%

group\_by(country, age\_grp, sex) %>%

summarise(pop = sum(pop)) %>%

ungroup() %>%

filter(country %in% selected\_countries) %>%

mutate(country = fct\_drop(country)) %>%

group\_by(country) %>%

mutate(prop = pop / sum(pop)) %>%

ungroup()

# check no misspellings in countries

stopifnot(sum(!selected\_countries %in% unique(pop\_2020$country)) == 0)

pop\_2020 %>%

ggplot(aes(x = as.numeric(age\_grp), y = prop, colour = sex)) +

geom\_line() +

facet\_wrap(~country) +

scale\_y\_continuous(label = percent\_format(accuracy = 1)) +

scale\_x\_continuous(breaks = 1:10, labels = levels(pop\_2020$age\_grp)) +

scale\_colour\_manual(values = sex\_cols) +

theme(panel.grid.minor = element\_blank(),

panel.grid.major.x = element\_blank(),

axis.text.x = element\_text(angle = 45, hjust = 1)) +

labs(x = "Age group",

y = "",

colour = "",

title = "Estimated proportion of the population in 2020",

subtitle = "By age group and sex",

caption = "Source: UN World Population Prospects 2019")

#----------Combine fatality rate with population--------------------

the\_caption = "Source: Italian case fatality rates to 19 March 2020 from Istituto Superiore di Sanità, Roma, combined with UN World Population Prospects 2019"

projected\_cfr <- pop\_2020 %>%

mutate(age\_grp = as.character(age\_grp)) %>%

left\_join(italy\_rates, by = c("age\_grp", "sex")) %>%

group\_by(country) %>%

summarise(cfr = sum(cfr \* prop) / sum(prop)) %>%

ungroup() %>%

mutate(country = fct\_reorder(country, -cfr))

xlabel <- "Case fatality rate if rates observed in Italy applied to each country's total age and sex profile.\n

Do not treat these as forecasts of actual case fatality rate."

# Version 1:

projected\_cfr %>%

ggplot(aes(y = country, x = cfr)) +

geom\_point(colour = "steelblue") +

geom\_text(aes(label = percent(cfr, accuracy = 0.1)), nudge\_x = 0.001, size = 3) +

geom\_segment(aes(yend = country, xend = 0), colour = "steelblue") +

scale\_x\_continuous(label = percent\_format(accuracy = 0.1)) +

theme(panel.grid.minor = element\_blank(),

panel.grid.major.y = element\_blank()) +

labs(subtitle = xlabel,

y = "",

title = "Different age profiles can make a big difference to overall fatality rates, based on Italian data",

x = "Note that in observed situations (eg Italy 8.5% to 19 March 2020), raw case fatality rates are more than double

those shown here, suggesting younger cases are either not diagnosed or not occurring.",

caption = the\_caption)

# Version 2, calibrated to actual Italy case fatality rate so far

projected\_cfr %>%

mutate(cfr\_adj = cfr / cfr[country == "Italy"] \* 0.085) %>%

ggplot(aes(y = country, x = cfr\_adj)) +

geom\_point(colour = "steelblue") +

geom\_text(aes(label = percent(cfr\_adj, accuracy = 0.1)), nudge\_x = 0.002, size = 3) +

geom\_segment(aes(yend = country, xend = 0), colour = "steelblue") +

scale\_x\_continuous(label = percent\_format(accuracy = 0.1)) +

theme(panel.grid.minor = element\_blank(),

panel.grid.major.y = element\_blank()) +

labs(subtitle = xlabel,

y = "",

title = "Different age profiles can make a big difference to overall fatality rates, based on Italian data",

x = "Estimates have been scaled to match Italy's raw case fatality rate to 19 March, to

reflect likely patterns in younger people's case rate and diagnosis.",

caption = the\_caption)