

Introduction

Coronavirus pandemic is changing our lifestyle from daily routine to near- and midterm plans, affecting relationships at home and work, adjusting our economical priorities and abilities, making us reassess value of goods and services, and arguably impacting all aspects of life. Better knowledge and understanding of the disease, its manifestations and dynamics must play critical role in assessment of current events and decisions we make. Below I compiled some useful facts about COVID-19 into 5 charts and included discussion of R and ggplot2 techniques used to create them.

At the end of 2019, a novel coronavirus was identified as the cause of a cluster of pneumonia cases in Wuhan, a city in the Hubei Province of China. It rapidly spread, resulting in an epidemic throughout China, followed by an increasing number of cases in other countries throughout the world. In February 2020, the World Health Organization designated the disease COVID-19, which stands for coronavirus disease 2019. The virus that causes COVID-19 is designated severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2); previously, it was referred to as 2019-nCoV.

Understanding of COVID-19 is evolving. This topic will discuss the epidemiology, clinical features, diagnosis, management, and prevention of COVID-19.

[Coronavirus disease 2019 \(COVID-19\)](#) by Kenneth McIntosh, MD

Though not all topics above are covered in this blog I reserve the right to publish more charts so stay tuned.

Clinical Features

Incubation Period

The incubation period for COVID-19 is thought to be within 14 days following exposure, with most cases occurring approximately four to five days after exposure [29-31].

Using data from 181 publicly reported, confirmed cases in China with identifiable exposure, one modeling study estimated that symptoms would develop in 2.5 percent of infected individuals within 2.2 days and in 97.5 percent of infected individuals within 11.5 days [32]. The median incubation period in this study was 5.1 days.

[Coronavirus disease 2019 \(COVID-19\)](#) by Kenneth McIntosh, MD

Common approach to display quartiles and extreme percentiles of continuous distribution is with [box plot](#). I chose against it for couple of reasons: a) research above had insufficient information about quartiles and b) box plots are less known outside of statistical community. Instead a gauge chart common in dashboard types of applications was used:

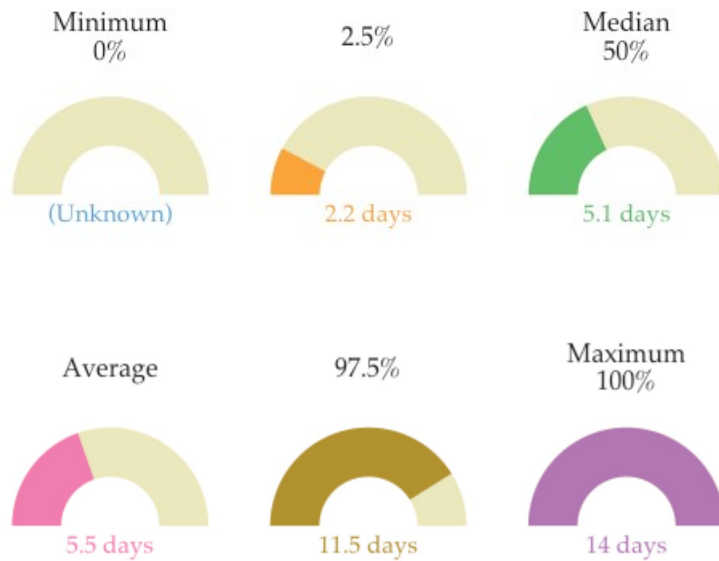
Implementation details in R

Dataset

Dataset consists of 6 rows corresponding to 5 percentiles - 0% (minimum), 2.5% and 97.5% (corresponding to 0.95 confidence interval), 50% (median), 100% (maximum) - and one row more for average:

COVID-19 Incubation Time

Time before infected person becomes symptomatic (percentiles and mean)



Source: COVID-19 Incubation Period: An Update, Stephen G. Baum, MD reviewing Lauer SA et al. Ann Intern Med 2020 Mar 10
<https://www.jwatch.org/na51083/2020/03/13/covid-19-incubation-period-update>

Using `factor()` will place gauges in order from least to greatest and additional column `stext` used to display a value in readable format for each gauge.

Graphics

First, let's load packages used for plotting: `ggplot2`, `ggthemes`, and `scales`:

```
library(ggplot2)
library(ggthemes)
library(scales)
```

Realization of gauge charts using `ggplot2` I borrowed from [this example](#) with a few changes explained next:

```
few_palette = "Medium"
ggplot(incubation_data, aes(ymax = svalue, ymin = 0, xmax = 2, xmin = 1, fill = sname)) +
  geom_rect(aes(ymax=14, ymin=0, xmax=2, xmin=1), fill = "#e6e8bd") +
  geom_rect() +
  coord_polar(theta = "y", start=-pi/2) +
  xlim(c(0, 2)) + ylim(c(0, 28)) +
  scale_fill_few(palette = few_palette) +
  scale_color_few(palette = few_palette) +
  geom_text(aes(x = 0, y = 0, label=stext, colour=sname), size=4, family="Palatino") +
  facet_wrap(~sname, ncol = 3) +
  guides(fill=FALSE, colour=FALSE) +
  labs(title="COVID-19 Incubation Time",
       subtitle = "Time before infected person becomes symptomatic (percentiles and mean) ",
       caption = "Source: COVID-19 Incubation Period: An Update. Stephen G. Baum, MD reviewing Lauer SA et al. Ann Intern Med 2020 Mar 10",
       x=NULL, y=NULL) +
  theme_tufte(ticks = FALSE, base_size = 12, base_family = "Palatino") +
  theme(axis.text = element_blank(),
        plot.caption = element_text(face = "italic", size=6),
        strip.text = element_text(size = 12))
```

Line by line explainer:

- 2-4: prepare rectangles for each value . Each gauge is a pair of overlapping rectangles - one displaying value `geom_rect()` with constant one `geom_rect(aes(ymax=14, ymin=0, xmax=2, xmin=1), fill = "#ece8bd")` as a background.
- 10: separate gauges by facets.
- 5, 6: transform coordinate system to polar, rotate it to start at 9 pm and trim to display only upper half of gauges.
- 9 places text label with value in the middle of each gauge.
- 7, 8: color schema from `few_pal()`.
- 11: removing guides from the chart.
- 12-15: title, subtitle, caption, and axis labels.
- 16-19: customization using `ggthemes` package and `theme()`.

Illness Severity

The spectrum of symptomatic infection ranges from mild to critical; most infections are not severe [33,35-40]. Specifically, in a report from the Chinese Center for Disease Control and Prevention that included approximately 44,500 confirmed infections with an estimation of disease severity [41]:

- Mild (no or mild pneumonia) was reported in 81 percent.
- Severe disease (eg, with dyspnea, hypoxia, or >50 percent lung involvement on imaging within 24 to 48 hours) was reported in 14 percent.
- Critical disease (eg, with respiratory failure, shock, or multiorgan dysfunction) was reported in 5 percent.
- The overall case fatality rate was 2.3 percent; no deaths were reported among noncritical cases.

Coronavirus disease 2019 (COVID-19) by Kenneth McIntosh, MD

Obvious choice is a **bar chart** consisting of 4 bars - 3 for illness severity spectrum plus case fatality rate reported in the same study:

Implementation details in R

Dataset

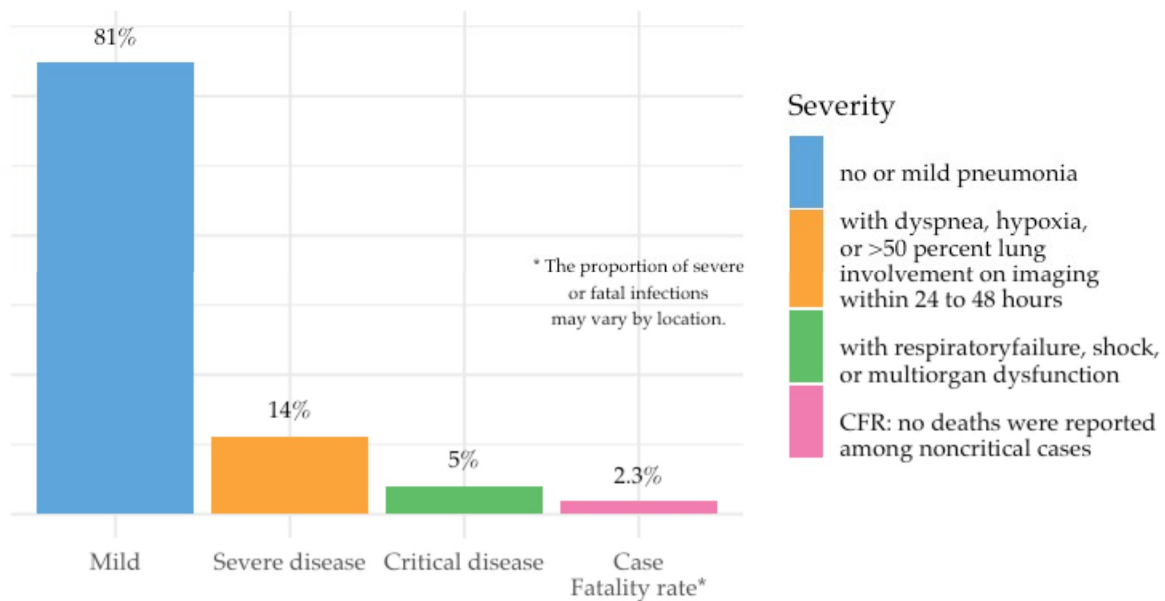
Dataset with 4 rows and 4 columns where `severity` is a `factor()` ordered by percent, `percent_label` used to display values above bars, and `severity_label` details illness severity:

```
severity_rates = c(0.81, 0.14, 0.05, 0.023)
spectrum_severity = data.frame(severity = c("Mild", "Severe disease", "Critical disease", "Case\nFatality rate*"),
                                percent = severity_rates,
                                percent_label = paste0(100 * severity_rates, "%"),
                                severity_label = c("\nno or mild pneumonia\n",
                                                    "with dyspnea, hypoxia,\nor >50 percent lung\ninvolvement on imaging\n",
                                                    "\nwith respiratory failure, shock,\nor multiorgan dysfunction",
                                                    "\n\nCFR: no deaths were reported\namong noncritical cases"))
spectrum_severity$severity = factor(spectrum_severity$severity,
                                    levels = spectrum_severity$severity[order(-spectrum_severity$percent)],
                                    ordered = TRUE)
```

Graphics

COVID-19 Spectrum of Illness Severity

The spectrum of symptomatic infection from mild to critical plus case fatality rate (CFR) as reported by the Chinese CDC from 72,314 Cases



Source: Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72,314 Cases From the Chinese Center for Disease Control and Prevention. Wu Z, McGoogan JM. JAMA. 2020 <https://jamanetwork.com/journals/jama/fullarticle/2762130>

This is the case of simple bar chart using `geom_bar()` with `state='identity'` enhanced just with a couple of artifacts: `geom_text()` and `annotate()`:

```
ggplot(spectrum_severity, aes(severity, percent)) +
  geom_bar(aes(fill=severity), stat = "identity") +
  geom_text(aes(label=percent_label), family = "Palatino", nudge_y = 0.05) +
  scale_y_continuous(labels = percent) +
  scale_fill_few(palette = few_palette, name = "Severity",
                 labels = spectrum_severity$severity_label) +
  annotate("text", 4, 0.4, family="Palatino", size=3,
          label="* The proportion of severe\nor fatal infections\nmay vary by location.") +
  labs(title="Spectrum of Illness Severity",
       subtitle = "The spectrum of symptomatic infection from mild to critical plus case fatality rate (CFR)\nas report",
       x=NULL, y=NULL,
       caption = "Source: Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbre",
  theme_minimal(base_size = 14, base_family = "Palatino") +
  theme(legend.position = "right",
        legend.direction = "vertical",
        axis.text.y = element_blank(),
        plot.caption = element_text(face = "italic", size=6))
```

Line by line explainer:

- 1-2: bar chart with `stat="identity"` displaying 4 bars.
- 3: placing percent labels above bars.
- 4: displaying y-axis labels in percent format.
- 5-6: color schema from `few_pal()` and custom labeling of the legend.
- 7-8: text annotation about CFR in the middle of the chart.
- 9-12: title, subtitle, caption, and axis labels.
- 13-17: customization using `ggthemes` package and `theme()`.

Clinical Manifestations

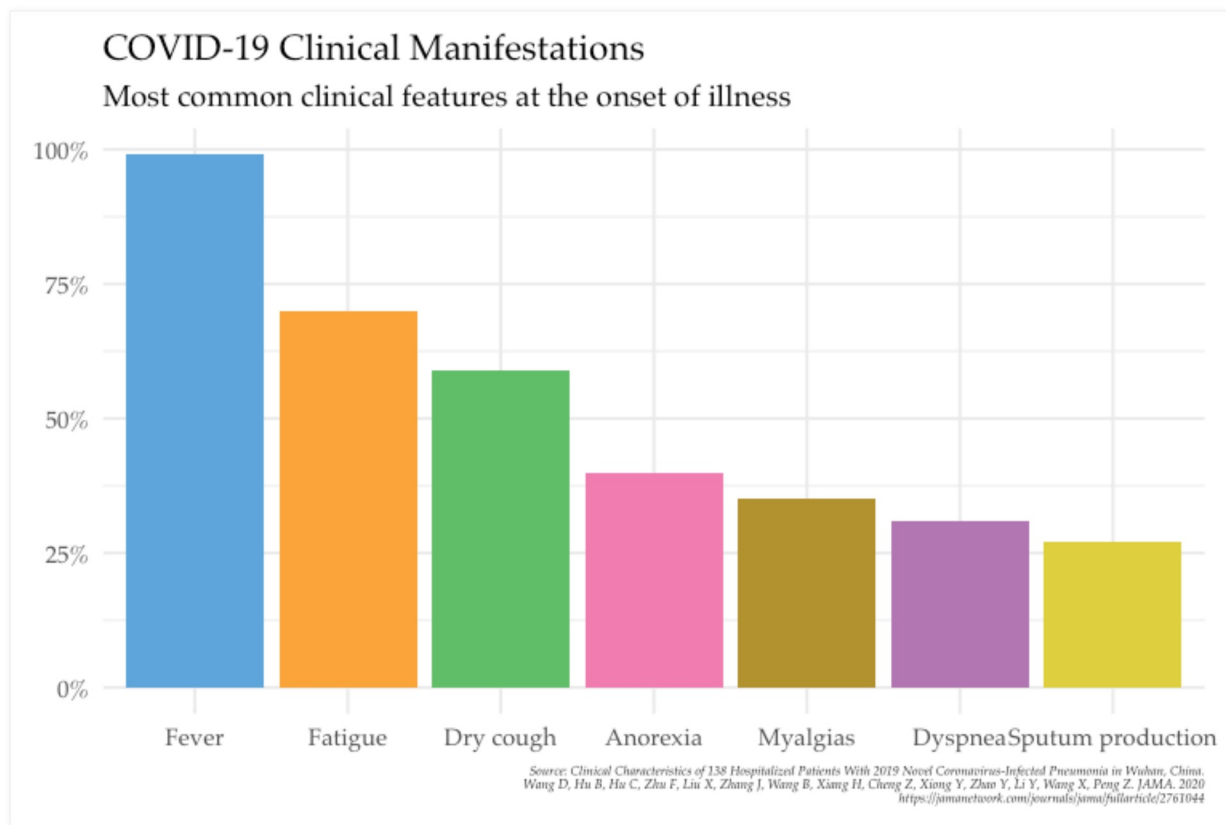
Pneumonia appears to be the most frequent serious manifestation of infection, characterized primarily by fever, cough, dyspnea, and bilateral infiltrates on chest imaging [32,36-38]. There are no specific clinical features that can yet reliably distinguish COVID-19 from other viral respiratory infections.

In a study describing 138 patients with COVID-19 pneumonia in Wuhan, the most common clinical features at the onset of illness were [38]:

- Fever in 99 percent
- Fatigue in 70 percent
- Dry cough in 59 percent
- Anorexia in 40 percent
- Myalgias in 35 percent
- Dyspnea in 31 percent
- Sputum production in 27 percent

Coronavirus disease 2019 (COVID-19) by Kenneth McIntosh, MD

Continuing using bar chart to display clinical manifestations of COVID-19 at the onset of illness:



Implementation Details in R

Dataset

This is example of a bar chart requiring a bare minimum of information - just 2 columns with name and percent to display 7 bars:

```
manifestation = data.frame(name =  
  factor(c("Fever", "Fatigue", "Dry cough", "Anorexia", "Myalgias", "Dyspnea", "Sputum produ  
    levels = c("Fever", "Fatigue", "Dry cough", "Anorexia", "Myalgias", "Dyspnea", "Spu  
    ordered = TRUE),  
  percent = c(0.99, 0.7, 0.59, 0.4, 0.35, 0.31, 0.27))  
manifestation$name = factor(manifestation$name,
```

```
levels = manifestation$name[order(-manifestation$percent)],
ordered = TRUE)
```

Graphics

Once again code below creates a bar chart using `stat = "identity"`:

```
ggplot(manifestation) +
  geom_bar(aes(name, percent, fill=name), stat="identity") +
  scale_y_continuous(labels = percent) +
  scale_fill_few(palette = few_palette, name = "Features") +
  labs(title = "Clinical Manifestations",
       subtitle = "Most common clinical features at the onset of illness",
       caption = "Source: Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Patients",
       x=NULL, y=NULL) +
  theme_minimal(base_size = 14, base_family = "Palatino") +
  theme(legend.position = "none",
       legend.direction = "vertical",
       plot.caption = element_text(face = "italic", size=6))
```

Line by Line explainer:

- 1-2: bar chart with `stat="identity"` displaying 4 bars.
- 3: displaying y-axis labels in percent format.
- 4: color schema from `few_pal()`.
- 5-8: title, subtitle, caption, and axis labels.
- 9-12: customization using `ggthemes` package and `theme()`.

Case Fatality Rate

According to a joint World Health Organization (WHO)-China fact-finding mission, the case-fatality rate ranged from 5.8 percent in Wuhan to 0.7 percent in the rest of China [17]. Most of the fatal cases occurred in patients with advanced age or underlying medical comorbidities [20,41]. (See ['Risk factors for severe illness'](#) below.)

The proportion of severe or fatal infections may vary by location. As an example, in Italy, 12 percent of all detected COVID-19 cases and 16 percent of all hospitalized patients were admitted to the intensive care unit; the estimated case fatality rate was 7.2 percent in mid-March [42,43]. In contrast, the estimated case fatality rate in mid-March in South Korea was 0.9 percent [44]. This may be related to distinct demographics of infection; in Italy, the median age of patients with infection was 64 years, whereas in Korea the median age was in the 40s.

[Coronavirus disease 2019 \(COVID-19\)](#) by Kenneth McIntosh, MD

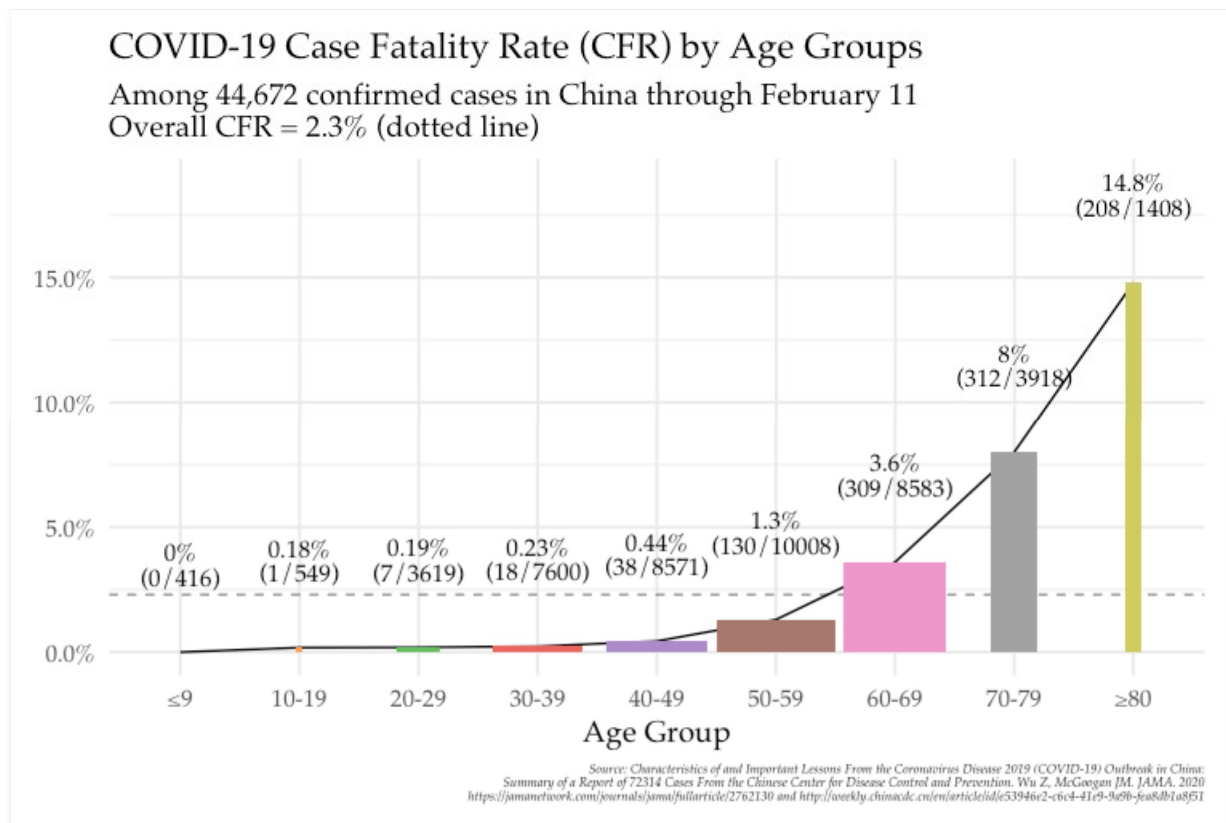
This chart displays CFR's by age groups based on 44672 confirmed cases in China through February 11 with overall CFR = 2.3%:

Implementation Details in R

Dataset

The data includes age, deaths, cases, and cfr computed as a ratio of last two:

```
age_groups = c("<=9", "10-19", "20-29", "30-39", "40-49", "50-59", "60-69", "70-79",
```



Graphics

This chart combines bar and line charts into single plot reflecting CFR rate dynamic over age groups and additionally reflects size of these groups using bar width:

```
ggplot(cfr_china, aes(x=age, y=cfr, group=1)) +
  geom_line() +
  geom_hline(yintercept = 0.01 * 2.3, linetype="dashed", alpha=0.5) +
  geom_bar(aes(width=cases/10000, fill=age), stat = "identity", position = "identity") +
  geom_text(aes(label=paste0(cfr*100,"%")), family="Palatino", size=4, nudge_y = 0.04) +
  geom_text(aes(label=paste0("(",deaths,"/",cases,")")), family="Palatino", size=4, nudge_y = 0.03) +
  scale_y_continuous(labels = percent) +
  scale_fill_tableau(palette = "Classic 10 Medium") +
  labs(title = "COVID-19 Case Fatality Rate (CFR) by Age Groups",
       subtitle = "Among 44,672 confirmed cases in China through February 11\nOverall CFR = 2.3% (dotted line)",
       x = "Age Group", y = NULL,
       caption = "Source: Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. Wu Z, McGoogan JM. JAMA. 2020",
       theme_minimal(base_size = 14, base_family = "Palatino") +
  theme(plot.caption = element_text(face = "italic", size=6),
       legend.position = "none")
```

Line by line explainer:

- 1,2: line chart over CFR by age groups.
- 3: horizontal dotted line representing overall case fatality rate.
- 1,4: bar chart with stat="identity" displaying CFR's for each age group with adjusted bar width based on number of cases in each group.
- 5,6: placing text labels with explicit value and calculation of CFR for each age group.
- 7: displaying y-axis labels in percent format.
- 8: color schema from few_tableau().
- 9-12: title, subtitle, caption, and axis labels.
- 13-15: customization using ggthemes package and theme().

Epidemiology

Period of infectivity

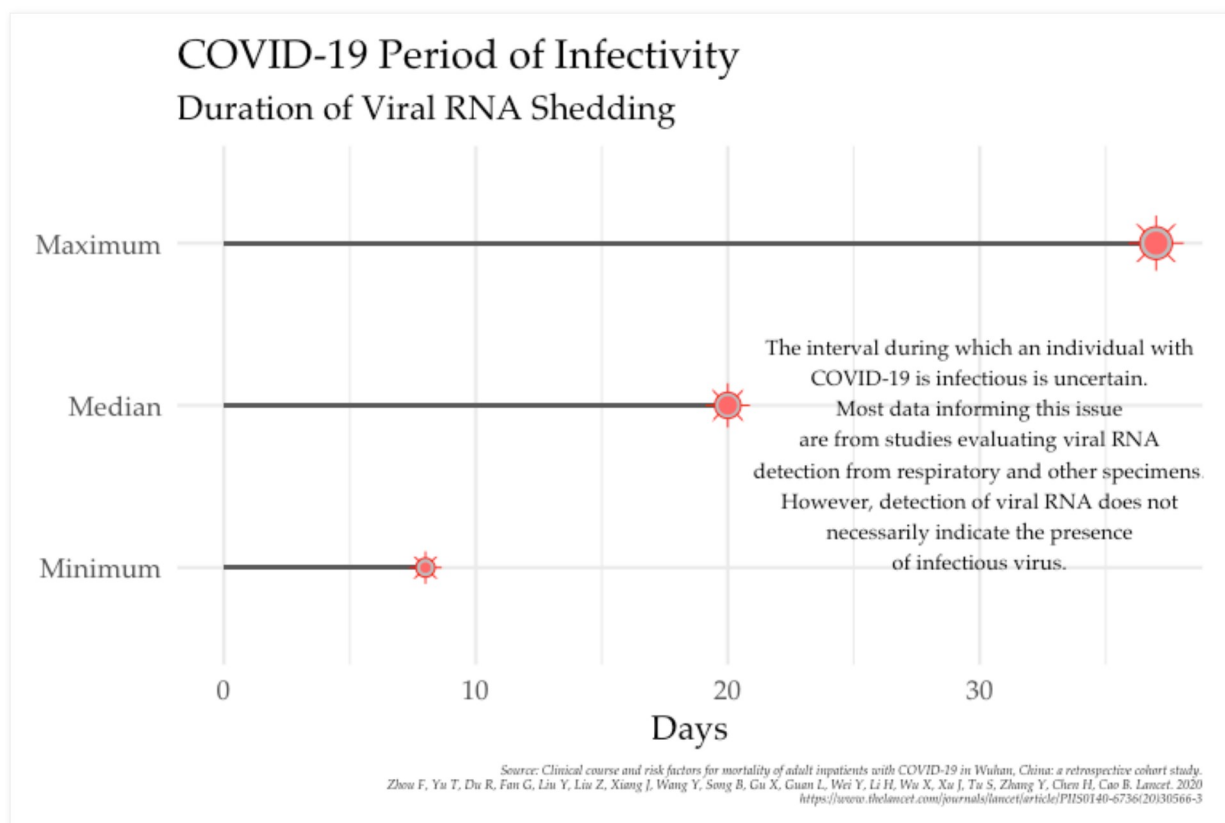
The interval during which an individual with COVID-19 is infectious is uncertain. Most data informing this issue are from studies evaluating viral RNA detection from respiratory and other specimens. However, detection of viral RNA does not necessarily indicate the presence of infectious virus.

Viral RNA levels appear to be higher soon after symptom onset compared with later in the illness [18]; this raises the possibility that transmission might be more likely in the earlier stage of infection, but additional data are needed to confirm this hypothesis.

The duration of viral shedding is also variable; there appears to be a wide range, which may depend on severity of illness. In one study of 21 patients with mild illness (no hypoxia), 90 percent had repeated negative viral RNA tests on nasopharyngeal swabs by 10 days after the onset of symptoms; tests were positive for longer in patients with more severe illness [19]. In another study of 137 patients who survived COVID-19, the median duration of viral RNA shedding from oropharyngeal specimens was 20 days (range of 8 to 37 days) [20].

Coronavirus disease 2019 (COVID-19) by Kenneth McIntosh, MD

This chart informs of minimum, median, and maximum duration of viral shedding by infected individuals by using bars resembling time lines:



Implementation Details in R

Dataset

This chart will use bars to imitate time lines of period of infectivity based on research of how long individuals shedded viral RNA that identified minimum, median and maximum times:

```
infectivity = data.frame(period = factor(c("Minimum", "Median", "Maximum"),  
                                         levels = c("Minimum", "Median", "Maximum"), ordered = TRUE),  
                         value = c(8, 20, 37))
```

Graphics

Yet another example of a bar chart with additional hack using `geom_point()`'s to display an improvised icon of SARS-CoV-2 virus:

```
ggplot(infectivity, aes(period, value)) +  
  geom_bar(stat = "identity", width = 0.03) +  
  geom_point(aes(size = value), shape = 8, color = "red") +  
  geom_point(aes(size = value/1.5), shape = 21, fill = "grey", color="red") +  
  geom_point(aes(size=value/4), shape = 16, color="indianred1") +  
  scale_size(range = c(2,8)) +  
  annotate("text", 1.7, 30, family="Palatino", size=3.5,  
          label="The interval during which an individual with\nCOVID-19 is infectious is uncertain.\nMost data informi",  
  coord_flip() +  
  labs(title = "COVID-19 Period of Infectivity",  
       subtitle = "Duration of Viral RNA Shedding",  
       caption = "Source: Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, Ch",  
       x = NULL, y = "Days") +  
  theme_minimal(base_size = 16, base_family = "Palatino") +  
  theme(plot.caption = element_text(face = "italic", size=6),  
        legend.position = "none")
```

Line by line explainer:

- 1,2: bar chart with `stat="identity"` displaying 3 very thin bars imitating time line.
- 3-6: overlaying 3 different point shapes with varying size to improvise virus icon
- 7,8: text annotation about the difference between being infectious and viral RNA shedding.
- 9: flipping x and y axis to display time line horizontally.
- 10-13: title, subtitle, caption, and axis labels.
- 14-16: customization using `ggthemes` package and `theme()`.

Conclusions

Most of the facts above are results of very young research of COVID-19 - just little over 3 months old. There are still many unknowns about both the virus SARS-CoV-2 and the disease. To emphasize this I compiled a few of unknowns in the bonus chart - some will seem surprising given the wealth of knowledge scientists accumulated about other similar diseases:

COVID-19 What We Still Don't Know

There are no specific clinical features that can yet reliably distinguish COVID-19 from other viral respiratory infections.

Transmission of SARS-CoV-2 from asymptomatic individuals (or individuals within the incubation period) has also been described ... However, the extent to which this occurs remains unknown.

The interval during which an individual with COVID-19 is infectious is uncertain

Although not highlighted in the initial cohort studies from China, smell and taste disorders (eg, anosmia and dysgeusia) have also been reported as common symptoms in patients... Whether this is a distinguishing feature of COVID-19 is uncertain.

Asymptomatic infections have also been described ... but their frequency is unknown.

Immunity — Antibodies to the virus are induced in those who have become infected.

Preliminary evidence suggests that some of these antibodies are protective, but this remains to be definitively established.

Moreover, it is unknown whether all infected patients mount a protective immune response and how long any protective effect will last.

The accuracy and predictive values of SARS-CoV-2 testing have not been systematically evaluated, and the sensitivity of testing likely depends on the precise test as well as the type of specimen obtained.

Certain investigational agents (a.k.a experimental drugs) have been described in observational series or are being used anecdotally based on in vitro or extrapolated evidence.

It is important to acknowledge that there are no controlled data supporting the use of any of these agents, and their efficacy for COVID-19 is unknown.

The optimal duration of home isolation is uncertain

Research Area



Clinical Features



Evaluation and Diagnosis



Epidemiology



Management

Source: Coronavirus disease 2019 (COVID-19) by Kenneth McIntos
<https://www.uptodate.com/contents/coronavirus-disease-2019-c>