

Covid-19 Fatality Rate Between 0.1 and 0.3%, Gangelt and Santa Clara Combined

Dieter Mergel

University of Duisburg-Essen, Faculty of Physics

47048 Duisburg, Germany

Dieter.Mergel@uni-due.de

Abstract

Covid-19 fatality rates of 0.37% (Gangelt, Germany) and 0.17% (Santa Clara, USA) have been reported, estimated from serological studies. We show that the two confidence intervals strongly overlap when the uncertainty in the number of deaths is taken into account, so that the two investigations may be regarded as representative of the same population (tentatively: “Western society with no overload of the Health Care System during the pandemic”). Combining the results, the Covid-19 fatality rate is estimated to be found with 95% confidence in the range [0.1%; 0.3%].

Keywords: Infectious disease modelling, Infection fatality rate, Confidence interval

1 Introduction

In this paper, we combine the infection fatality rates for Covid-19 reported in two studies (Streeck and others 2020; and Bendavid and others 2020) where the infection rate in the respective county is estimated from the fraction of people with antibodies against Sars-CoV2 in a representative sample.

We shall neglect all possible systematic deficiencies of the studies, take the reported numbers at face value, and simplify the statistical calculations by approximating the uncertainties with standard errors. This simplification is expected to yield an essentially correct estimation and to change only slightly the numerical values.

2 Results and discussion

2.1 Infection fatality rates

In the German community Gangelt (12,597 inhabitants) with a superspreading event, a representative sample of 919 individuals was tested and 15.5% (CI95 [12.3; 19.0%] were found to have been infected (Streeck and others 2020). This leads to an estimated 1946 infections in the county and, with the officially reported 7 deaths related to Covid-19, to an infection fatality rate (IFR) of 0.36% with a CI95 of [0.29%; 0.45%] estimated from the uncertainty of the number of infected alone.

In the Californian county Santa Clara (1.9 million inhabitants), a sample of 3,300 was tested for antibodies with 50 positive results (Bendavid and others 2020). The sample itself is not representative of Santa Clara so that the data had to be re-weighted to reflect the demographic structure of that county. The infection rate was then estimated to be 2.8% (CI95 [1.3%; 4.7%])

corresponding to 54,000 individuals (CI95 [25,000; 91,000]). The IFR is calculated as $0.17\% = (94 \text{ deaths in the region})/(54,000 \text{ infected})$. The authors do not specify a confidence interval.

In the two studies, the infection fatality rates IFR are calculated as:

$$f = \frac{d}{i}$$

Equation 1

with f = fatality rate, d = number of Covid-19 fatalities reported for a county, i = number of infected persons in the county estimated from a sample. In both studies, it is taken care of explicitly and assumed

- that the samples are representative of the respective population.

Furthermore, it is implicitly assumed

- that the demographic distribution of actually infected persons is similar to that of the entire county.
- The IFR of Santa Clara (0.17 %) does not fall into the confidence interval ([0.29%; 0.45%]) of the IFR claimed for Gangelt. However, this does not mean that the two studies are not compatible with one another as will be argued in the next section.

2.2 Confidence intervals

The authors of the Gangelt study claim that their IFR may be used to estimate the percentage of infections in other places with similar population statistics. However, as the number of deaths is regarded a true value without uncertainty, the confidence interval claimed for Gangelt is valid only for this single specific infection process. If the infection process could be repeated, resulting again in about 1950 infected persons, we would not expect again exactly 7 fatalities.

In order to make the estimation valid for a larger population, e.g. for Germany, Gangelt has to be considered a sample and the number of fatalities must be considered the result of a random process where 1949 infected individuals die each with a probability 0.36%. This is a typical Poisson-process resulting in 2 to 12 deaths (94% confidence level). Consequently, in Equation 1, both nominator and denominator exhibit uncertainty.

For products and quotients, the relative variances of the two components add (Mergel 2017) leading to an equation for the uncertainty of the fatality rate:

$$\left(\frac{f_{2se}}{f}\right)^2 = \left(\frac{d_{2se}}{d}\right)^2 + \left(\frac{i_{2se}}{i}\right)^2$$

Equation 2

where the subscript “2se” indicates the error margin for a confidence level 0.95, approximately twice the standard error. This gives a CI95 of [0.04%; 0.68%] for the IFR in Gangelt.

Applying the same reasoning to Santa Clara, the number of deaths ranges between 74 and 103, yielding, together with the uncertainty of the number of infected, a CI95 of [0.06%; 0.29%].

The results are represented graphically in Figure 1. The numerical calculations are reported in Figure 2 in the annex as spreadsheet calculation.

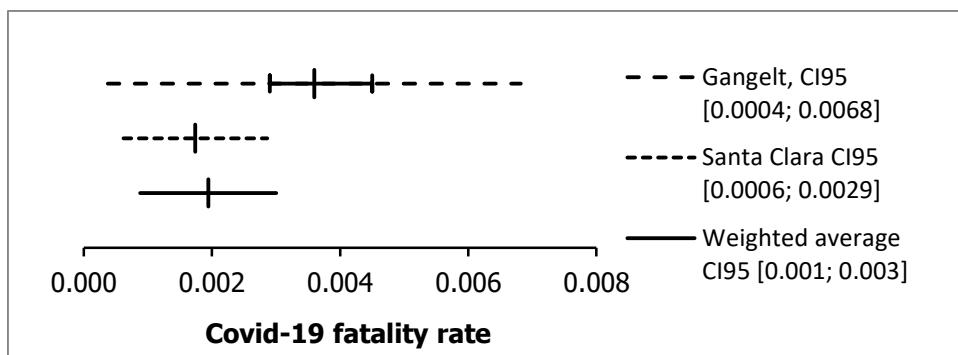


Figure 1: Covid-19 fatality rates, top: Gangelt, IFR and CI95 reported (bold line) and extended (dashed line) calculated with uncertainty in the number of deaths, middle: Santa Clara IFR and CI95, bottom: combined result

2.3 Combining the two results

In Figure 1, the top line represents the results of Gangelt and the middle line those of Santa Clara. The extended CI95 of Gangelt includes that of Santa Clara so that, from a pure statistical point of view, there is no objection against considering the samples as arising from the same population although Gangelt is a small town in Germany with some medieval buildings and surrounded by villages whereas Santa Clara is a metropolitain area in the USA. Demographically, the population could be characterized as “Western society with no overload of the health system during the epidemic”.

We may then combine the two IFRs (f_G for Gangelt and f_{SC} for Santa Clara) in a weighted average (Mergel 2017):

$$f = \frac{f_G \cdot w_G + f_{SC} \cdot w_{SC}}{w_G + w_{SC}}$$

Equation 3

with the weights being the respective values of $(f_{2se}^2)^{-1}$, proportional to the inverse of the variances. The IFR of Santa Clara gets a higher weight than that of Gangelt because its uncertainty is smaller, with the combined IFR being 0.2%, shifted only slightly from the IFR of Santa Clara towards the IFR of Gangelt.

The uncertainty of the combined IFR is calculated also with the inverse of the variances (Mergel 2017):

$$\frac{1}{f_{2se}(\text{combined})^2} = \frac{1}{f_{2se}(\text{Gangelt})^2} + \frac{1}{f_{2se}(\text{Santa Clara})^2}$$

Equation 4

The result is CI95 = [0.1%; 0.3%] represented by the bottom line in Figure 1. It is smaller than the CI95 of Santa Clara alone because more data are combined. The numerical calculations are also reported in Figure 2.

The confidence interval is not compatible with the data from New York City (Goodman, Rothfeld 2020) where the fatality rate is 0.14% of the entire population and about 20% of the population is seropositive. The reason is unknown. We suppose that the assumptions of the calculation may not hold:

- set if infected persons representative of the entire population
- health care system not overwhelmed.

3 Conclusions

In order to compare estimates of the IFR from different studies, the number of fatalities in a region has to be regarded the result of a random process. This enlarges the confidence interval.

The serological studies in Gangelt (small German town) and in Santa Clara (metropolitan area in the USA) may be considered statistically to be samples of the same population because the extended confidence intervals overlap so that the estimations of the Covid-19 fatality rate may be combined. The true IFR for the population (tentatively: "Western society with no overload of the health system during the Covid-19 pandemic") is still unknown but may be found with 95% confidence between 0.1% and 0.3%.

This is not compatible with data from New York. The assumption, implicitly underlying the estimation of IFR from samples, that the sociodemographic distribution of the whole population is also representative of the set of actually infected people has still to be checked.

4 References

(Bendavid and others, 2020): Eran Bendavid, Bianca Mulaney, Neeraj Sood, Soleil Shah, Emilia Ling, Rebecca Bromley-Dulfano, Cara Lai, Zoe Weissberg, Rodrigo Saavedra-Walker, James Tedrow, Dona Tversky, Andrew Bogan, Thomas Kupiec, Daniel Eichner, Ribhav Gupta, John Ioannidis, Jay Bhattacharya

COVID-19 Antibody Seroprevalence in Santa Clara County, California

doi: <https://doi.org/10.1101/2020.04.14.20062463>

posted April 30, 2020 on medRxiv

<https://www.medrxiv.org/content/10.1101/2020.04.14.20062463v2>

(Goodman, Rothfeld 2020) David Goodman, Michael Rothfeld, 1 in 5 New Yorkers May Have Had Covid-19, Antibody Tests Suggest, The New York Times, April 23, 2020

<https://www.nytimes.com/2020/04/23/nyregion/coronavirus-antibodies-test-ny.html>

(Mergel 2017) Mergel, D., Physik mit Excel und Visual Basic, Springer Spektrum 2017, ISBN 978-3-642-37856-0, Chapter 8, Auswertung von Messungen (Evaluation of measurements).

(Streeck and others, 2020): Hendrik Streeck, Bianca Schulte, Beate Kuemmerer, Enrico Richter, Tobias Hoeller, Christine Fuhrmann, Eva Bartok, Ramona Dolscheid, Moritz Berger, Lukas Wessendorf, Monika Eschbach-Bludau, Angelika Kellings, Astrid Schwaiger, Martin Coenen, Per Hoffmann, Markus Noethen, Anna-Maria Eis-Huebinger, Martin Exner, Ricarda Schmithausen, Matthias Schmid, Gunther Hartmann, Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event; doi: <https://doi.org/10.1101/2020.05.04.20090076>

posted June 02, 2020 on medRxiv;

<https://www.medrxiv.org/content/10.1101/2020.05.04.20090076v2>

5 Annex

	A	B	C	D	E	
1	Calculating the confidence interval for the Covid-19 fatality rate combined from Gangelt and Santa Clara					
2						
3						
4		Gangelt, CI95 [0.0004; 0.0068]				
5	infected Gangelt	iG	1946 =142/919*12597			
6	2 x standard error	iG2se	421 calculated from CI95 = [0.123%; 0.19%]			
7	deaths Gangelt	dG	7 reported			
8	2 x standard error	dG2se	5.3 =2*SQRT(dG) (Poisson)			
9						
10	relative variance infected	rViG	0.05 =(iG2se/iG)^2			
11	relative variance deaths	rVdG	0.76 =(dG2se/dG)			
12						
13	fatality Gangelt	fG	0.0036 =dG/iG			
14	2 x standard error	fG2se	0.0032 =SQRT(rViG+rVdG)*fG			
15	CI95		0.04% =fG-fG2se			
16			0.68% =fG+fG2se			
17	weight Gangelt	wG	96322 =1/fG2se^2			
18						
19						
20	Santa Clara CI95 [0.0006; 0.0029]					
21	infected Santa Clara	iSC	54000			
22	2 x standard error	iSC2se	33000 calculated from CI95 = [25,000 to 91,000]			
23	deaths Santa Clara	dSC	94 reported			
24	2 x standard error	dSC2se	19.4 =2*SQRT(dSC) (Poisson)			
25						
26	relative variance infected	rViSC	0.37 =(iSC2se/iSC)^2			
27	relative variance deaths	rVdSC	0.04 =(dSC2se/dSC)^2			
28						
29	fatality Santa Clara	fSC	0.0017 =dSC/iSC			
30	2 x standard error	fSC2se	0.0011 =SQRT(rViSC+rVdSC)*fSC			
31	CI95		0.06% =fSC-fSC2se			
32			0.29% =fSC+fSC2se			
33	weight Santa Clara	wSC	793283 =1/fSC2se^2			
34						
35	Weighted average CI95 [0.001; 0.003]					
36	fatality rate	fWt	0.0019 =(fG*wG+fSC*wSC)/(wG+wSC)			
37	Internally consistent error		0.11% =SQRT(wG+wSC)^-1			
38	2 x standard error					
39			0.0009 =C36-C37	0.1%		
40			0.0030 =C36+C37	0.3%		
41						
42	Externally consistent error		0.06% =SQRT((wG*(fG-fWt)^2 +wSC*(fSC-fWt)^2)/(wG+wSC))			
43						

Figure 2: Spreadsheet calculations (Excel 2010)