Simpson Paradox in Covid-19 case fatality rates: a mediation analysis of age-related causal effect

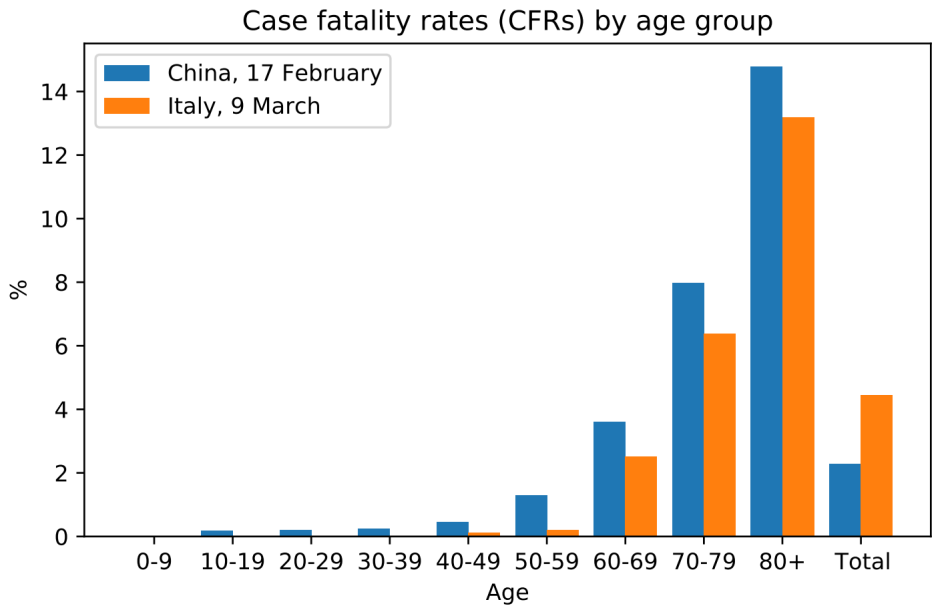
Stefano Gualtieri

Julius von Kugelgen, Luigi Gresele, Bernhard Scholkopf

Introduction

What is the paradox?

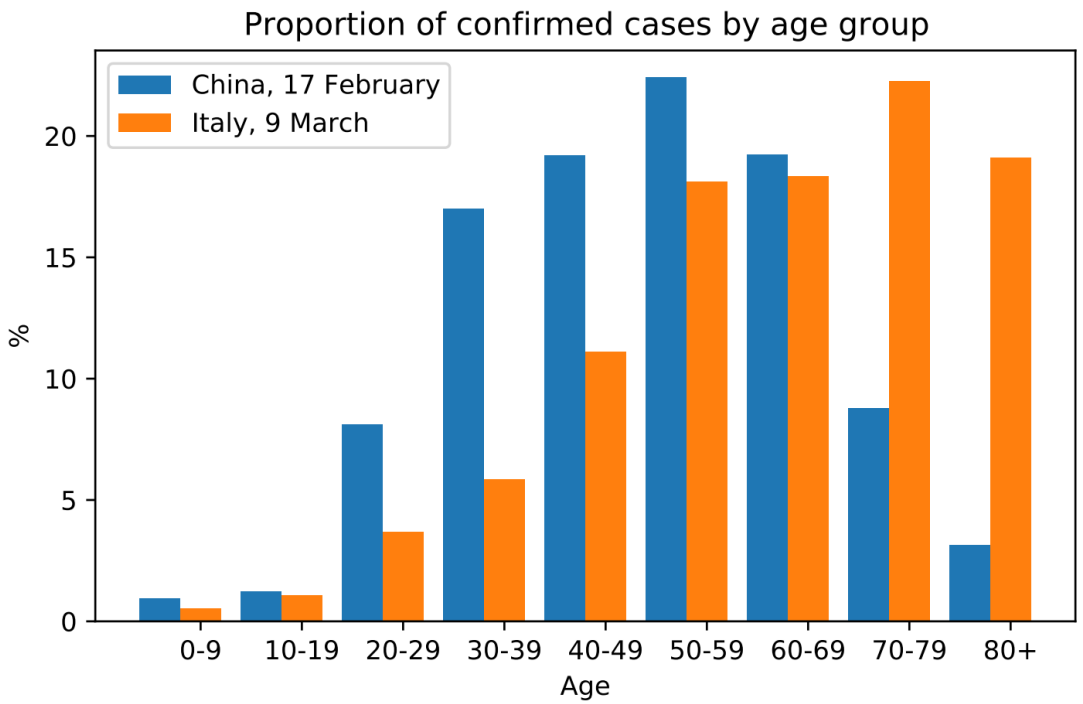
- CFRs are lower in *Italy* than in *China* for all age groups
- total CFR is higher in Italy



Higher CFRs for older people in both countries

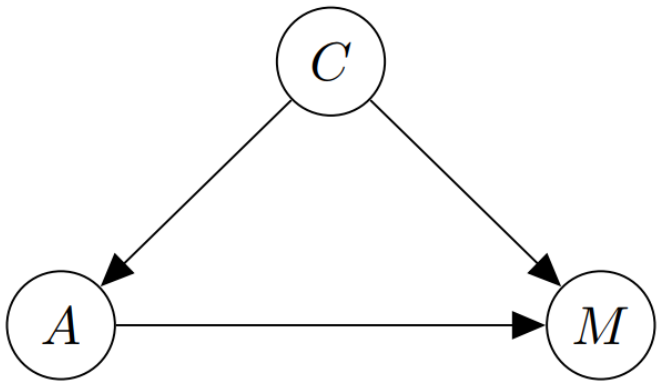
A first statistical explanation

- Statistical correlation between the country of reporting and the number of confirmed cases per age group
- Italy recorded a much higher proportion of confirmed cases in older patients



Assumptions

- Variables:
  - **C**: *country*, categorical variable
  - **A**: *age group*, ordinal variable
  - **M**: *mortality*, binary variable
- *Causal sufficiency*



TCE

"What would be the effect on mortality of changing country from China to Italy"

Total Causal Effect

$$TCE_{Ch \to It} = \mathbb{E}[M|do(C = Italy)] - \mathbb{E}[M|do(C = China)]$$

$$TCE_{0 \to 1} = \sum_a [P(M = 1|A = a, C = 1)P(A = a|C = 1)$$

$$-P(M = 1|A = a, C = 0)P(A = a|C = 0)] = 2\%$$

TCE

The results correspond to the difference of the total CFRs reported in the last column of this table:

Table 1: Comparison of case fatality rates (CFRs) by age group for Italy and China (deaths/confirmed cases in brackets). Lower CFRs are highlighted in bold. Sources: Wu and McGoogan (2020) and Istituto Superiore di Sanità (2020).

Age	0–9	10–19	20–29	30–39	40–49	50–59	60–69	70–79	≥ 80	Total
Italy	**0%** (0/43)	**0%** (0/85)	**0%** (0/296)	**0%** (0/470)	**0.1%** (1/891)	**0.2%** (3/1,453)	**2.5%** (37/1,471)	**6.4%** (114/1,785)	**13.2%** (202/1,532)	4.3% (357/8,342)
China	**0%** (0/0)	0.2% (1/549)	0.2% (7/3,619)	0.2% (18/7,600)	0.4% (38/8,571)	1.3% (130/10,008)	3.6% (309/8,583)	8% (312/3,918)	14.8% (208/1,408)	**2.3%** (1,023/44,672)

We would like to isolate the **direct** effect of *country* on *mortality*

Mediation Analysis: CDE

"For 50-59-year-olds, is it safer to get the disease in China or Italy?"

Controlled direct effect

$$CDE_{0\to 1}(a) = \mathbb{E}[M|do(C = 1, A = a)] - \mathbb{E}[M|do(C = 0, A = a)]$$

$$CDE_{0\to 1}(a) = P(M = 1|C = 1, A = a) - P(M = 1|C = 0, A = a)$$

We would like to measure the direct effect at the *population level*

Mediation Analysis: NDE

"For the Chinese case demographic, would it have been better to take the Italian approach instead"

Natural direct effect

$$NDE_{0\to 1} = \mathbb{E}[M_{A(0)}|do(C=1)] - \mathbb{E}[M_{A(0)}|do(C=0)]$$

A(0) is the counterfactual distribution of Age had Country been equal to China

$$NDE_{0\to 1} = \sum_a P(A=a|C=0)[P(M=1|A=a,C=1) - P(M=1|A=a,C=0)] = -0.8\%$$

Mediation Analysis: NIE

"How would the overall CFR in China change if the case demographic had instead been that from Italy while keeping all else (the CFRs) the same?"

Natural indirect effect

$$NIE_{0\to1} = \mathbb{E}[M_{A(1)}|do(C=0)] - \mathbb{E}[M_{A(0)}|do(C=0)]$$

Where A(1) refers to the counterfactual distribution of Age had Country been Italy

$$NIE_{0\to1} = \sum_{a} P(M=1|A=a, C=0)\cdot$$

$$[P(A=a|C=1) - P(A=a|C=0)] = 3.3\%$$

Final Considerations

When switching the Country from China to Italy we obatined:

- $NDE < 0$
- $NIE > 0$

This reflects the effect of the mediator (Age) on Mortality and it is consistent with the lower CFRs for age group in Italy