



資料迷思1：新冠只對老年人有害嗎？



曹天元 Capo

科普作家

431 人贊同了該文章

在這個巨量資料時代，各種統計資料可謂鋪天蓋地。不管是新聞機構、大眾自媒體，還是正規的學術研究，所有人似乎都正在引用越來越多的“資料”，以顯得自己“有理有據，嚴謹公正”。乍看起來，這似乎是一種很好的風氣。

然而遺憾的是，由於大部分人對機率和統計缺乏深入理解，更有很多人連相關的基本概念也是模糊不清，這就導致他們對引用的“資料”缺乏基本的邏輯分析能力，由此得出的結論當然也是南轅北轍的。

時至今日，哪怕在學界，對於機率統計的誤解誤用，情況也已經變得愈加嚴重。濫用p值，胡亂應用各種統計檢驗，倒果為因，把相關性解釋為因果性.....越來越多的例子正在被不斷報導出來。也許有史以來，從來沒有哪個時代，被資料“欺騙”的人會像今天這麼多。正如馬克吐溫引用的那句著名的話：世界上有三大謊言：謊言、該死的謊言，以及統計數字。

在這裡，我並不打算大量引用複雜的統計術語，更不打算長篇大論地去列公式、做計算。因為在很多話題上面，其實只要釐清一些基本的概念，再運用一些簡單的邏輯思維，就可以分辨絕大部分的“資料陷阱”，所需的數學水平基本不會超過中學。在這個系列裡，就讓我們用儘可能簡單的語言來聊一聊有關資料、機率和統計方面的話題。

首先，基於最近疫情的話題又開始火熱，我們不妨還是從新冠相關的問題開始吧。在媒體上，經常看到的說法是：在所有感染了新冠的人群當中，絕大多數死者都是老年人，所以，這個病只對老年人有嚴重危害，而對於年輕人則不算什麼大事。

乍看上去，好像很有道理，但是停，讓我們仔細想一想。如果你的思維比較縝密，你會很快意識到這裡有些不對勁。雖然最後的結論我們不知道到底對不對，但很明顯：這個論證方式是有問題的。哪怕是在自然的情況下，在每天正常死亡的人當中，難道本來不就應該是老年人居多嗎？如果我們把新冠換成“喝水”，那麼，在“喝過水後死去的人”當中，肯定也是老年人居多，年輕人很少。所以是不是可以說：喝水對老年人危害嚴重，對年輕人則無關痛癢？這個結論肯定是荒謬的。

當然了，大家憑日常經驗就知道，喝水是一件無害的事情，所以我們可以直覺式地排斥以上結論。但如果是某種陌生的事物呢？或許判斷就沒有那麼容易了。

讓我們想像這樣的場景：有一天，突然一個惡魔來到了地球上。這個惡魔對所有人都施加了某種邪術，然後突然間，大批人紛紛死去。悲劇過後，我們檢查死者的年齡分佈，發現老年人佔其中的絕大多數。而且年齡越高的人，死亡比例就越高。比方說，90歲以上的老人有99%都死了，而80-89歲的老人有90%死了，70-79歲的老人則死了75%.....當然，年輕人也有死的，但是絕對數量很少，而且佔同齡總人數的比例也很低。

那麼，我們可以從中得出什麼樣的結論呢？是不是這個惡魔特別討厭老年人，所以必要致他們於死地而後快？然而，這個惡魔很快又回來了，他驚訝地看著我們，說：沒有啊，其實我對所有人都是“一視同仁”的啊。我的法術，其作用不過是讓每個人都減少了十年陽壽而已。實際上，所有人的遭遇都是平等的，並沒有特殊針對任何群體。



知乎 @曹天元 Capo

惡魔究竟對地球人做了什麼？

現在，大家可以看到，在這個故事裡，其實表面上的“老人死得多”只是一種假象。真正的情況是，因為老人本來就余壽無多，所以哪怕遭遇到同樣的壽命折損，他們也會“當場”死得更多。而年輕人死得少，也並不代表他們一點事沒有。只不過年輕人本來餘下的壽命就很長，哪怕減去十年，其效果也不會立刻顯現出來。

所以，回到新冠的問題上，如果我們想要證明新冠確實對老年人危害更大，至少需要考慮以下幾種情況：

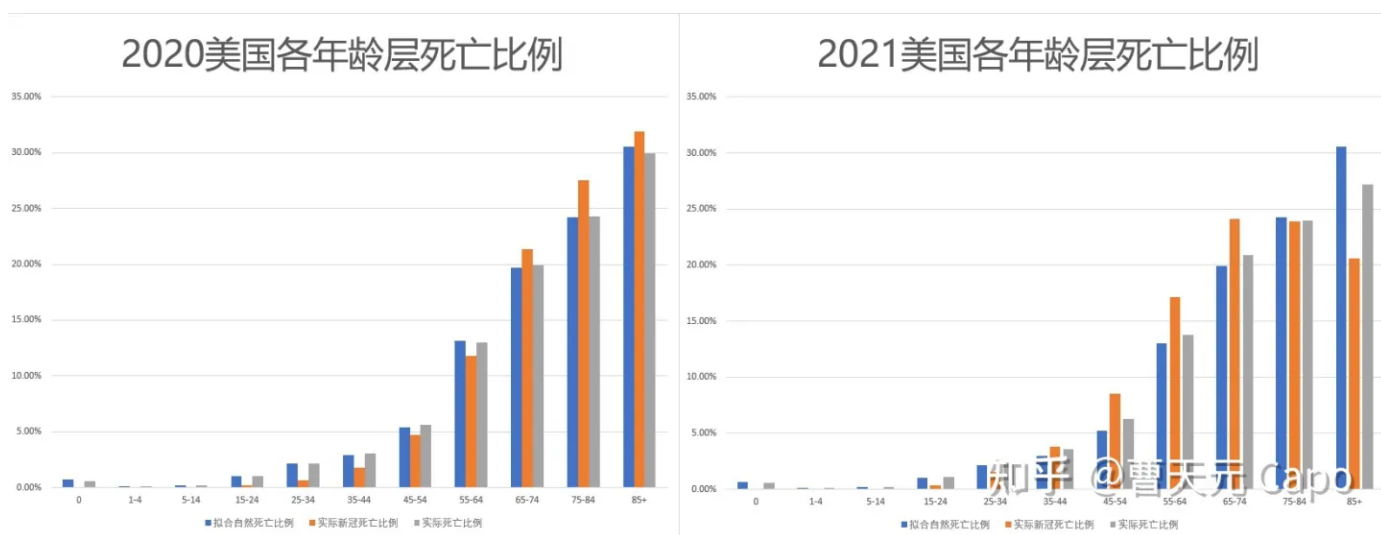
第一，不能單純地只說“老年人的絕對死亡人數更多”，而需要考察他們佔總死亡人數的比例。這個比例相比於“自然”狀態下，是大幅升高了，還是變化不大？

第二，即使老年人佔死亡人數的比例提升了，考慮到他們的“死亡風險”本來就更高。那麼，哪怕所有人的風險都提升同樣的比例，老年人的“新增死亡”也將會因此變得更多。如果是這種情況，那這個臨時的數字其實意義也不大，因為年輕人其實承擔了同樣的額外風險，只不過這個風險“目前”還沒有顯現出來而已。

現在，讓我們先來考察在沒有疫苗的情況下，原始新冠毒株以及後來的Delta變種對“自然狀態”下人群的影響。縱觀全球，在疫情問題方面，只有美國CDC發佈了最詳盡、最豐富的相關資料，因此

是最理想的考察對象。

首先，我們統計疫情前五年，也就是2015-2019年美國的人口資料，並且擬合出2020和2021年“本來應有”的分年齡死亡比例。事實上，因為CDC剛剛發佈了最新的2019年美國人口生命表，兩相參照之下，這個數字可以擬合得相當準確。接下來，我們再對照2020和2021年美國實際公佈的新冠死亡人數，同樣按照年齡層給出相應的比例。最後，考慮到美國“上報”的新冠數字未必全面精準，可能有沒查到的，或者被隱藏的。幸好，2020-2021年的全因死亡人數統計目前也已經出爐，所以我們可以獲取到這兩年精準的“美國總人口實際死亡數字”。先不管其中究竟有多少是“因為新冠導致”，我們把這個實際數字也按照年齡分層，得出相應的比例。將三個數字擺在一起對比，結果如下：



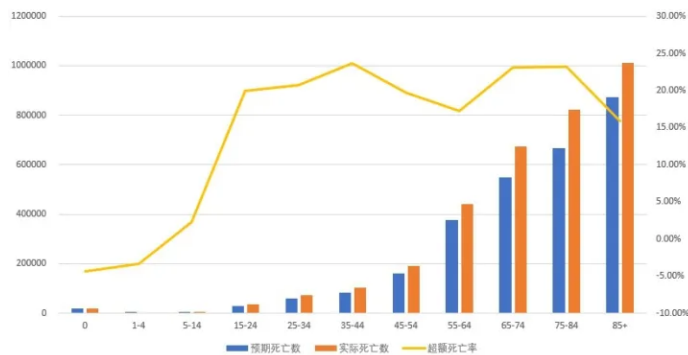
可以看出，在過去兩年，美國各年齡層的實際死亡比例其實跟預期差不多，並沒有出現劇烈的大幅波動。在新冠初起的2020年，美國85歲以上老人佔了實際死亡人口的29.93%，在當年報告的38萬個“新冠死亡”人數當中，他們則佔了31.91%，而我們本來“預計”他們應該佔自然死亡總數的30.51%。考慮到“新冠死亡統計”有一定的不可靠性，實際上，2020年高齡老人佔總死亡人數的比例甚至要比“預期”略少。

更誇張的是2021年，可能是因為極端脆弱的高齡老人在上一年基本都已去世，2021年，85歲以上老人僅佔美國實際死亡人口的27.18%。在報告“新冠死亡”的46萬人中，更是僅佔20.6%。而我們“預期”當年他們本來應該佔美國“自然死亡人數”的30.57%。可以說，在這一年，美國的超高齡老人相對其他年齡層，反而算是“死得少”的。

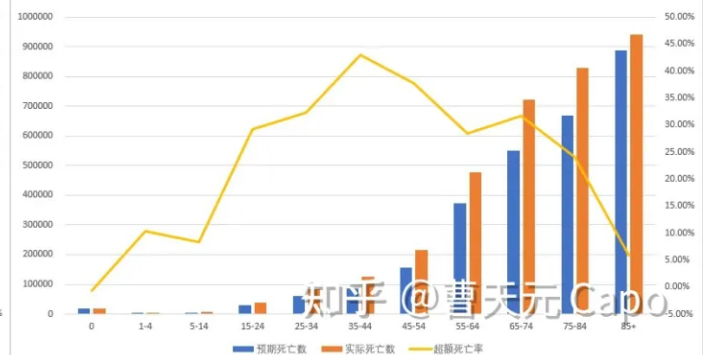
以上是按照“佔比”而言，接下來，我們再按照每年的絕對死亡率，求一個“實際”的風險出來。比方說，如果按照2019年的“自然死亡率”，某年齡段人群“應該”在2020年死亡50萬，而實際上死亡了55萬，那麼我們就容易得出結論：該年齡段在2020年“超額”死亡了10%。注意，這個數字只考慮全部人口，跟“報告”了多少新冠死亡無關，換句話說，不管這些人是“因為”什麼而死，總之在這一年全部加起來多死了10%。

同樣，我們把2020-2021年美國各年齡段的預期死亡人數，實際死亡人數，還有超額死亡率都畫出來，如下圖：

2020年美国各年龄层死亡总数
以及超额死亡率



2021年美国各年龄层死亡总数
以及超额死亡率



從這個圖中，我們可以看出，新冠對於15歲以下的兒童，其殺傷力確實很低，第一年的超额死亡率最多在2%，第二年最多在10%，而對於剛出生不久的嬰兒，風險甚至是負數（有些研究指出，這可能是因為嬰兒剛從母體出生時會“自帶抗體”的緣故）。而隨著年齡增加，新冠的殺傷力逐漸升高，但也並不是年齡“越大越危險”。實際上，資料表明：35-44歲的人風險最大。本來我們預計他們在2020年會死亡84527人，而實際卻死了104490人，增加了23.6%。在2021年，本來我們預計他們會死亡87089人，實際卻死了124577人，增加了43.05%。

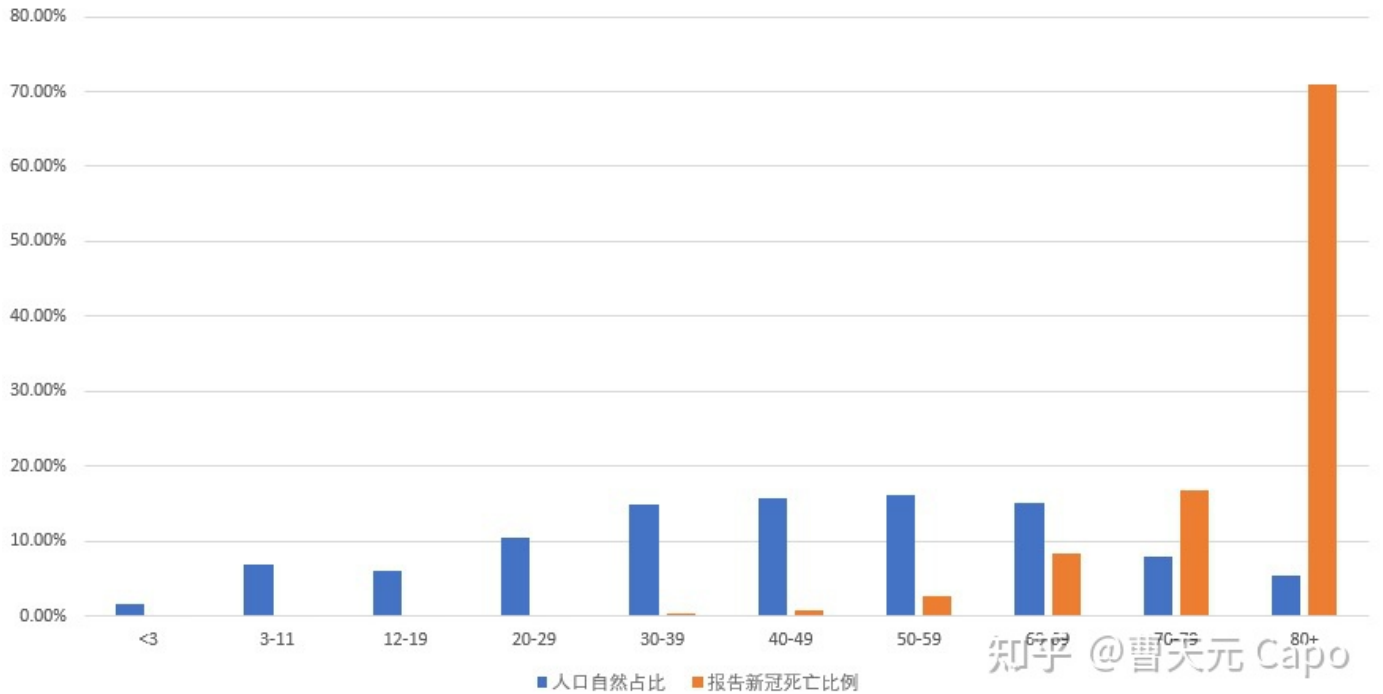
相比之下，對於85歲以上的超高齡人群來說，2020年我們預計他們死亡874281人，實際死亡1012805人，增加了15.8%。而2021年，我們預計將死亡888266人，實際死亡卻只有939942人，僅增加5.82%。雖然“絕對數字”看上去很多，但相比之下，高齡人群風險提升的“比例”卻反而更小。

事實上，從上圖中我們可以發現，“超额死亡”的比例大概在35-65歲之間是最高的，而到了特別高齡的階段，反而會有所降低。從這個角度上講，在大家都沒有接種疫苗的情況下，原始新冠毒株實際上對中年人和“次老”的老年人才是殺傷力最大的，而“極老”的老年人雖然看起來死得很多，但實際上，風險提升的比例反而沒有前者那麼高。

但是，以上說的是還沒有疫苗的時候（美國雖然從2020年底就開始接種疫苗，但考慮到全程接種三針需要大半年多的週期，加上很多美國人不肯打疫苗，他們在2021年的保護也可以說是很不充分的）。有了疫苗之後，加上新的Omicron變種的出現，情況變得稍微有點複雜起來。以下，我們來考察一下香港從今年1月1號起至今（5月13號）的數字。

按港府給出的官方人口數，香港在2021年底共有7403100人，自今年疫情爆發以來，總共報告死亡9142人（截止5月11日），如果按年齡分層，畫出對應的比例，我們可以得到下圖：

香港人口年齡自然比例 和新冠死亡比例



這個圖乍看上去，乖乖不得了，本來80歲以上的人群在香港總人口當中，僅佔比5.38%，但在9142個新冠死亡的人當中，80歲以上竟然有6495人，佔比高達71.05%！很多人肯定要說了，現在我們大可拋開美國的數字不管，因為當時還是新冠的原始株和Delta變種，已經是老皇曆了。至少目前來看，香港的資料說明Omicron變種對老年人的傷害更大，這總歸是毫無疑問的吧？

哎，還真不一定，因為這裡出現了一個新的變數，就是在不同的年齡層之間，存在著不同的疫苗接種率。事實上，如果把該變數考慮進去，我們會發現：這裡出現的是一個極其經典的統計學謬誤，也就是所謂的“辛普森悖論”。大家不妨先自行思考一下，我們在下一篇裡再來繼續詳談。