Size matters: little vs large

Tom Jemmett

Mohammed A Mohammed

Scenario

- A busy centralised GP clinic provides flu jabs during the winter on a walk-in basis.
- The clinic is staffed with **4 nurses**.
- On average **28 patients** arrive per hour.
- Administering the jab takes **8 minutes**.
- It was observed that on a typical day queues were forming and patients were complaining about long waiting times.
- It was suggested that a solution to the problem would be to split the resources such that there are two clinics with **2 nurses** at each clinic.
- It is assumed that patients will choose the more conveniently located of the two clinics for themselves such that the total patient arrival rate will be split equally between the two clinics (**14 patients** arrive per hour at each clinic).

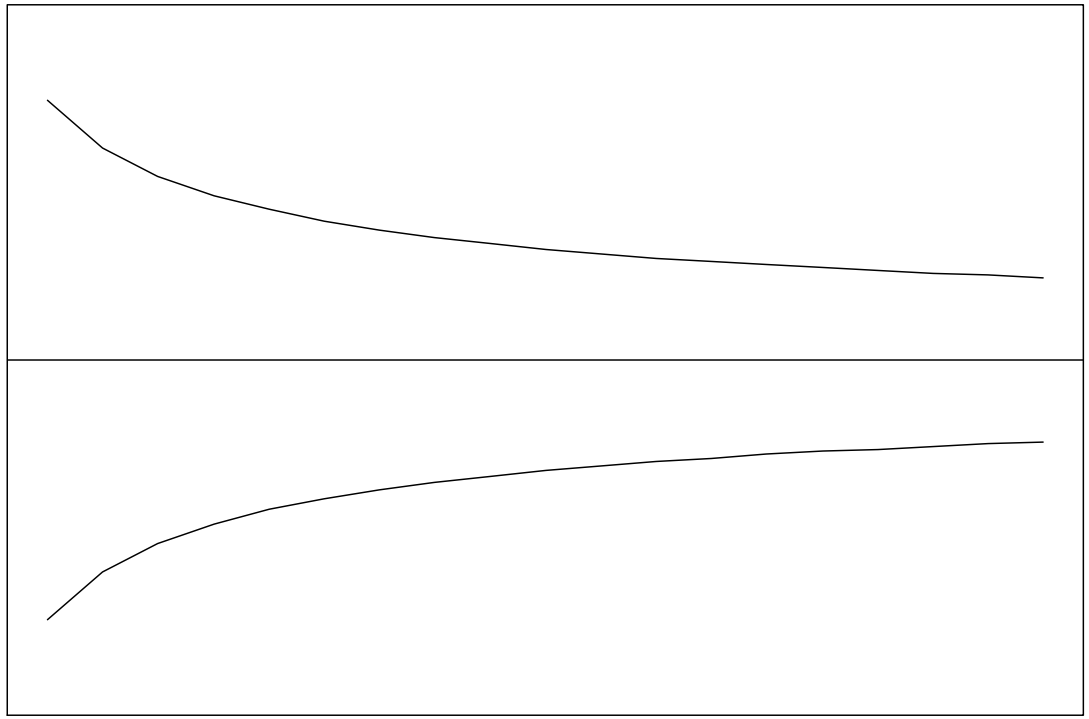
- **Will this change lead to shorter waiting times?**

Modelling the two scenarios (using Queuing Theory)

Metric	One large clinic	Two smaller clinics
Avg Arrival rate	28 patients per hour	14 patients per hour (per clinic)
Avg number in queue	11.9 patients	12.6 patients (per clinic)
Avg waiting time	26 mins	54 mins

With a single clinic the nurses are able to all work together to deal with the random arrivals and see the patients.

When the clinics are split, if one clinic becomes busy, then the second clinic's nurses are unable to help see patients, so the queue becomes longer and waiting times increase.



N	10	100	1000	10000	100000
p	10%	10%	10%	10%	10%
low	0%	2%	7%	9%	10%
high	50%	20%	13%	11%	10%

Little vs Large

- In a survey of 1662 schools in Pennsylvania, 6 of the top 50 were small, which is an over representation by a factor of 4.
- These data encouraged the Gates Foundation (and others) to invest $1.7 billion into setting up smaller schools, sometimes by splitting larger schools into smaller ones.

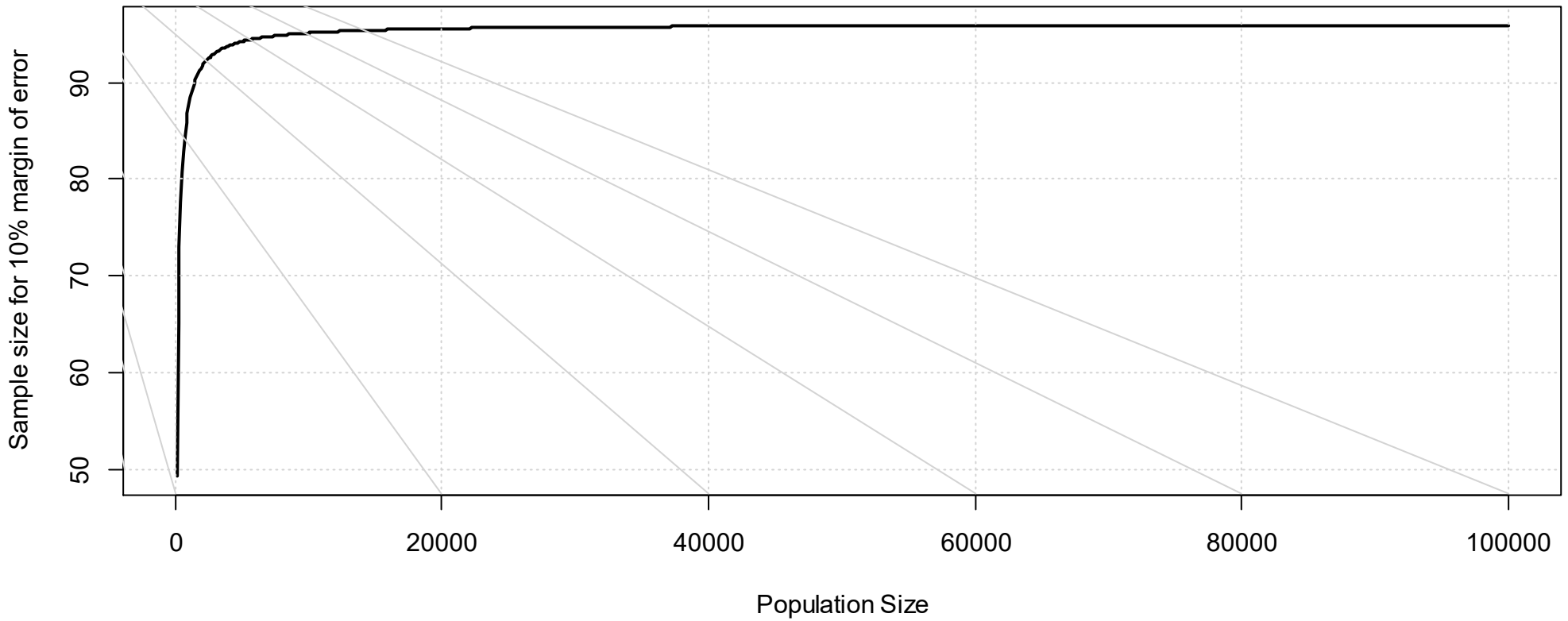
Kahneman Thinking, Fast and Slow pg 177

Survey sampling: little vs large

- There are 10,000 junior nursing staff
- We want to learn about vaccine uptake/hesitancy with a 10% margin of error
- Three identical random surveys are on the table, differing only in sample size – which do think will be more accurate?

    - Company A: sample size=8000
    - Company B: sample size=500
    - Company C: sample size=100

	Size of population					
Margin of error	>5000	5000	2500	1000	500	200
±10%	96	94	93	88	81	65



Article

Unrepresentative big surveys significantly overestimated US vaccine uptake

https://doi.org/10.1038/s41586-021-04198-4

Received: 18 June 2021

Accepted: 29 October 2021

Published online: 8 December 2021

Check for updates

Valerie C. Bradley[1,6], Shiro Kuriwaki[2,6], Michael Isakov[3], Dino Sejdinovic[1], Xiao-Li Meng[4] & Seth Flaxman[5]✉

Surveys are a crucial tool for understanding public opinion and behaviour, and their accuracy depends on maintaining statistical representativeness of their target populations by minimizing biases from all sources. Increasing data size shrinks confidence intervals but magnifies the effect of survey bias: an instance of the Big Data Paradox[1]. Here we demonstrate this paradox in estimates of first-dose COVID-19 vaccine uptake in US adults from 9 January to 19 May 2021 from two large surveys: Delphi–Facebook[2,3] (about 250,000 responses per week) and Census Household Pulse[4] (about 75,000 every two weeks). In May 2021, Delphi–Facebook overestimated uptake by 17 percentage points (14–20 percentage points with 5% benchmark imprecision) and Census Household Pulse by 14 (11–17 percentage points with 5% benchmark imprecision), compared to a retroactively updated benchmark the Centers for Disease Control and Prevention published on 26 May 2021. Moreover, their large sample sizes led to miniscule margins of error on the incorrect estimates. By contrast, an Axios–Ipsos online panel[5] with about 1,000 responses per week following survey research best practices[6] provided reliable estimates and uncertainty quantification. We decompose observed error using a recent analytic framework[1] to explain the inaccuracy in the three surveys. We then analyse the implications for vaccine hesitancy and willingness. We show how a survey of 250,000 respondents can produce an estimate of the population mean that is no more accurate than an estimate from a simple random sample of size 10. Our central message is that data quality matters more than data quantity, and that compensating the former with the latter is a mathematically provable losing proposition.

Hubbard's points about small samples

- Expand with above…

Reflections 1-2-4(All)

The Strategy Unit.

Stay connected with us!

https://www.strategyunitwm.nhs.uk/

@Strategy_Unit

0121 612 1538

strategy.unit@nhs.net

NHS
Midlands and Lancashire
Commissioning Support Unit