
Missing data in global health and why it matters

SALONI DATTANI

- RESEARCHER, OUR WORLD IN DATA
- CO-FOUNDER, WORKS IN PROGRESS

SALONI@OURWORLDINDATA.ORG

MARCH 2023, EAG X CAMBRIDGE

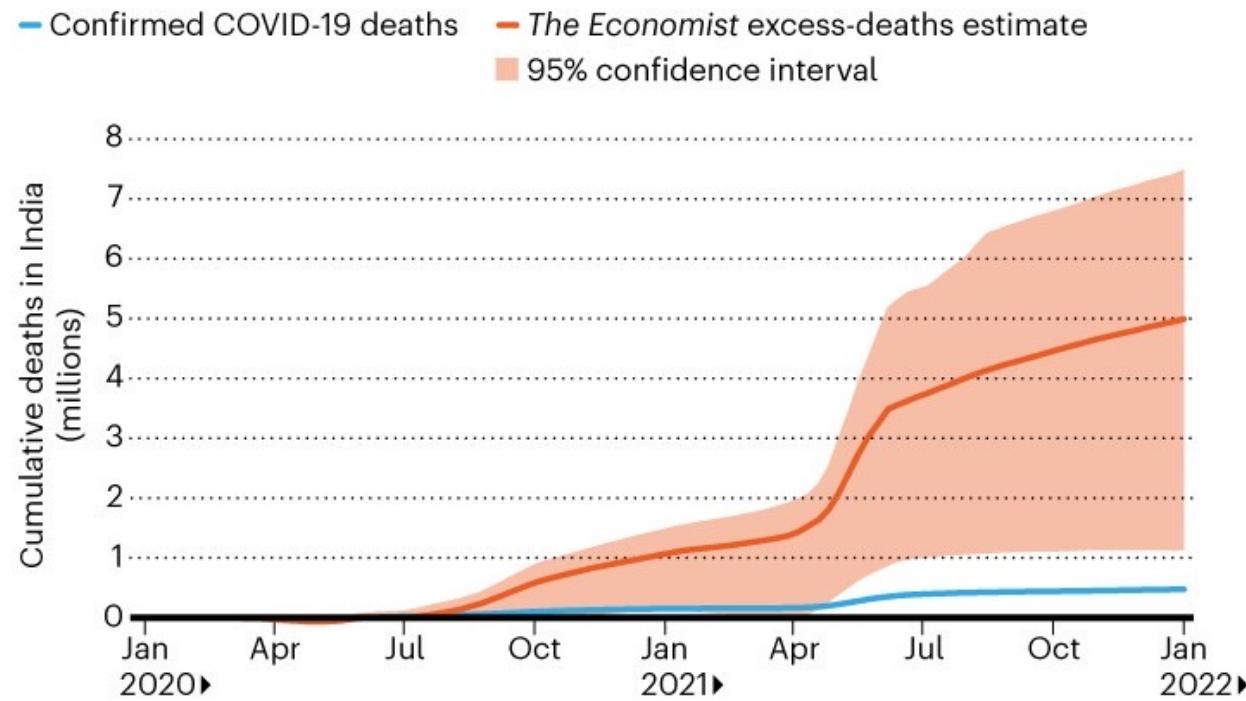


Two key messages:

1. There's wide uncertainty around how many Covid deaths actually occurred.
2. Even the lower bound estimate of excess deaths is almost **triple** the number of reported Covid deaths.

MILLIONS OF MISSING DEATHS

India has reported fewer than 500,000 COVID-19 deaths — but both *The Economist*'s model and household surveys suggest that real pandemic toll is in the millions.



©nature

The Institute for Health Metrics and Evaluation's model not shown, for clarity: it gives similar, overlapping, results.

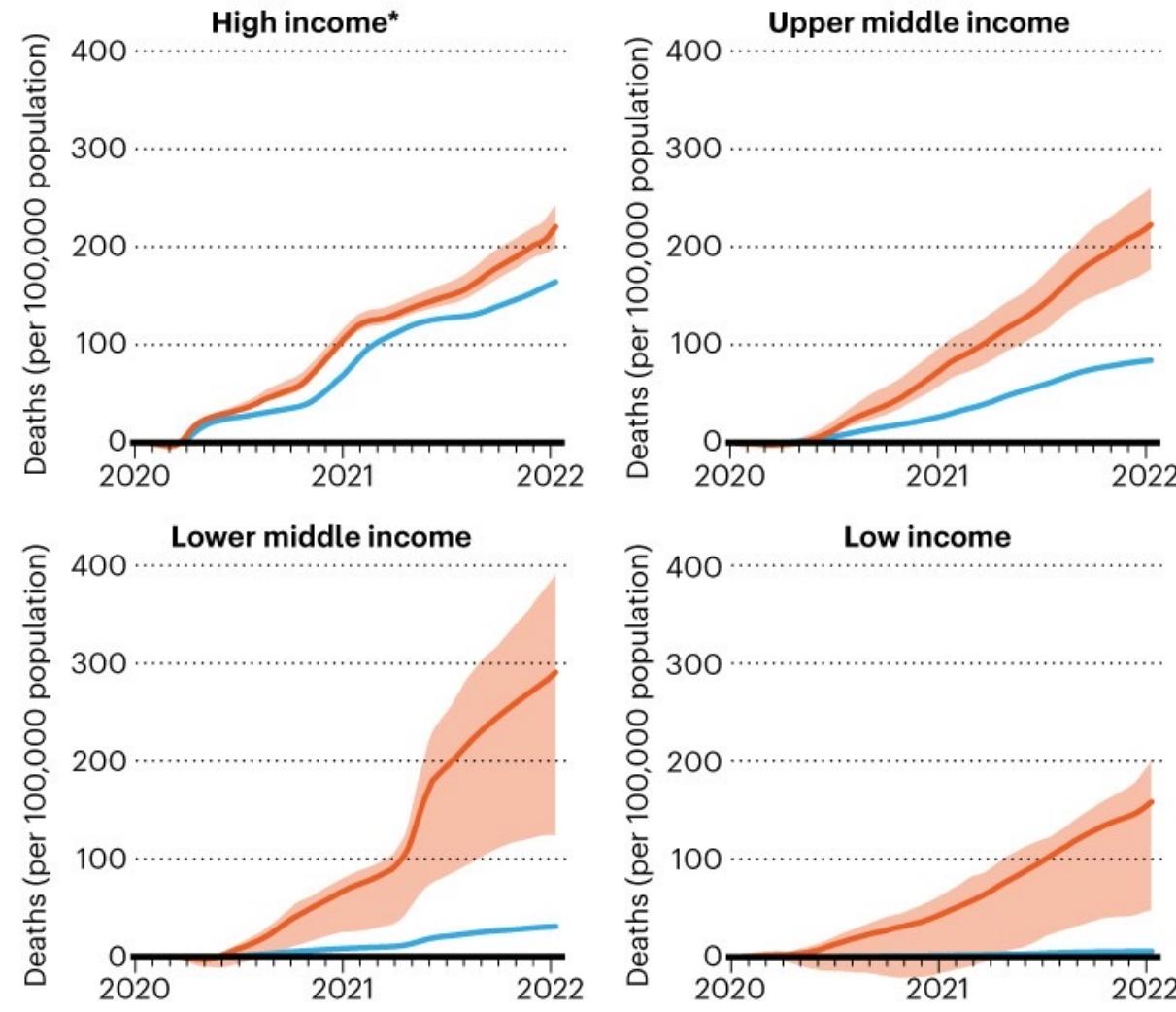
RICH AND POOR

Official figures suggest that wealthy countries had the highest number of deaths per capita during the pandemic. But a model that estimates excess deaths suggests that is false: lower middle-income countries might have been hit hardest.

— Confirmed COVID-19 deaths — *The Economist* excess deaths estimate
— 95% confidence interval

Two key messages:

1. There's wide uncertainty around how many Covid deaths actually occurred.
2. Even the lower bound estimate of excess deaths is almost **triple** the number of reported Covid deaths.

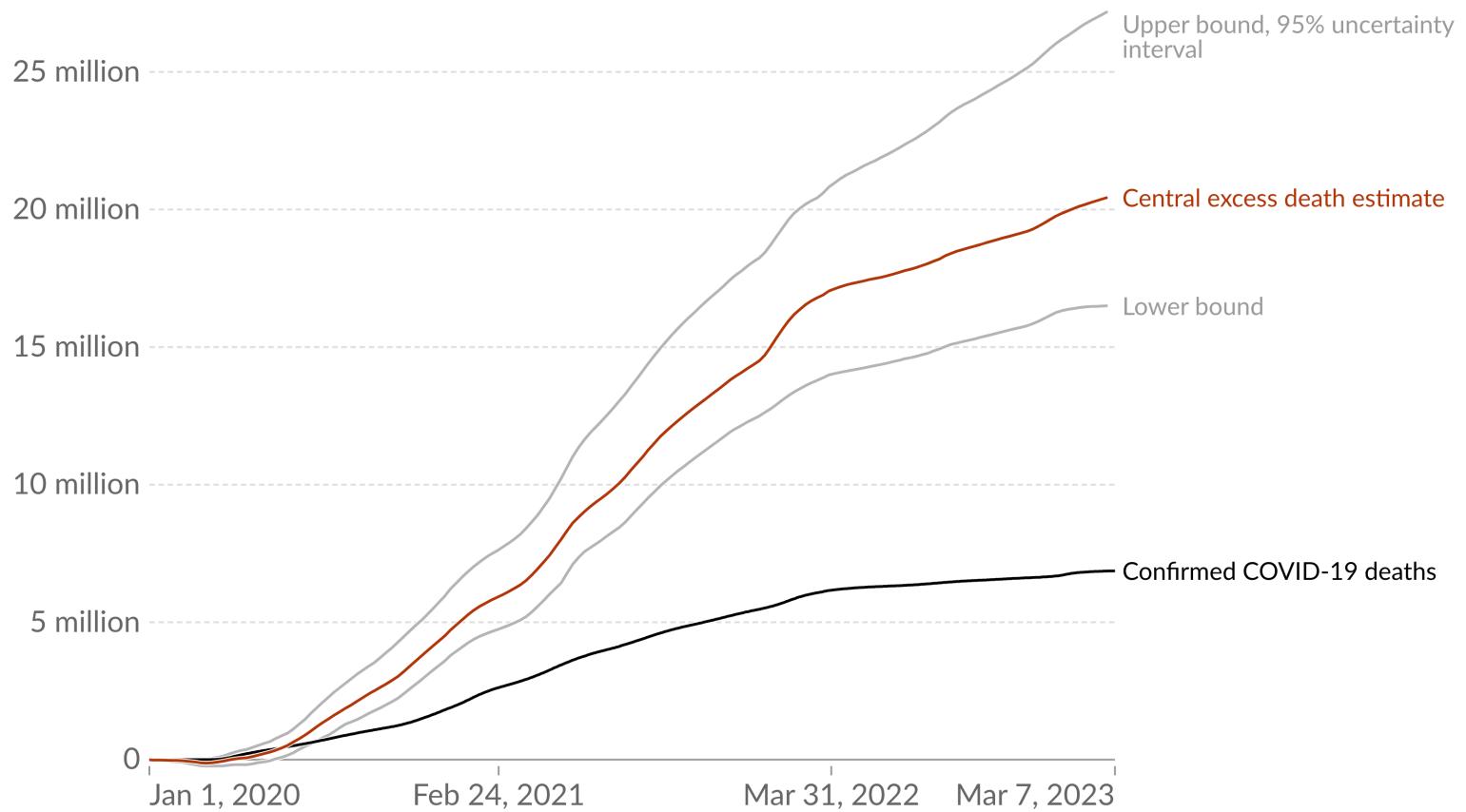


Two key messages:

1. There's wide uncertainty around how many Covid deaths actually occurred.
2. Even the lower bound estimate of excess deaths is almost **triple** the number of reported Covid deaths.

Estimated cumulative excess deaths during COVID, World

For countries that have not reported all-cause mortality data for a given week, an estimate is shown, with uncertainty interval. If reported data is available, that value only is shown. For comparison, cumulative confirmed COVID-19 deaths are shown.



Source: The Economist (2023); WHO COVID-19 Dashboard

OurWorldInData.org/coronavirus • CC BY

Let's take a step
back...

Why are we missing
data in the first place?

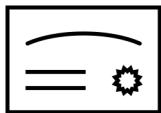
How do we know what people are dying from?



When someone dies,



A healthcare worker (doctor or nurse) who took care of them fills in a death certificate. Otherwise, a coroner investigates the cause of their death.



In the certificate, they describe the steps that led to the person's death and their 'underlying cause of death'.



These certificates are collected by the country's civil registry, and the 'underlying cause of death' is turned into an ICD death code. For example, someone who dies of tuberculosis is given the code A15.



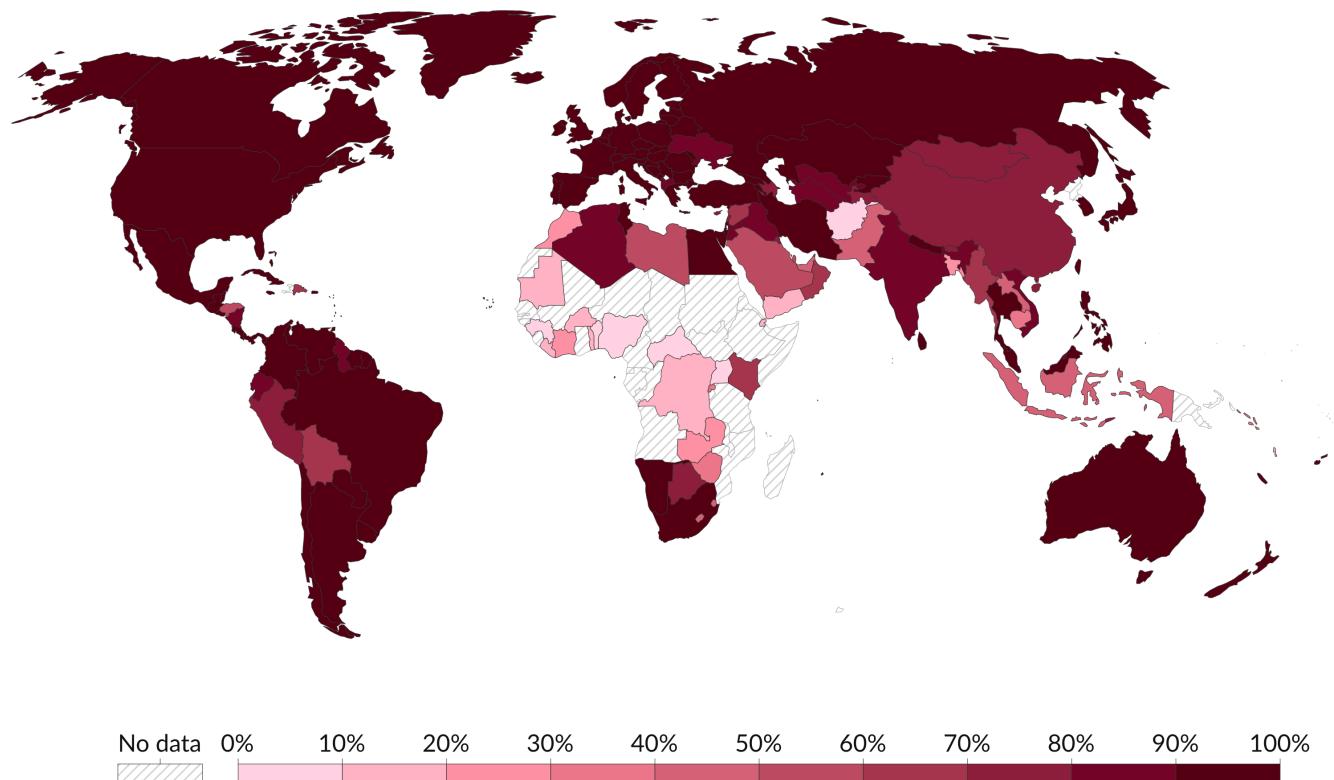
Then, these death codes are sent to the WHO by each country, once a year.

But...

Share of deaths that are registered, 2019

The number of deaths reported in a country's vital registration system as a share of total expected deaths. Expected deaths are taken as the average of estimates from three international sources: the UN, WHO, and IHME.

Our World
in Data



1. Many deaths aren't registered at all

Even in countries with a civil registry, not all deaths are registered.

2. Many deaths are registered without a cause of death

Even in countries with death registration, people may not be given a cause of death.

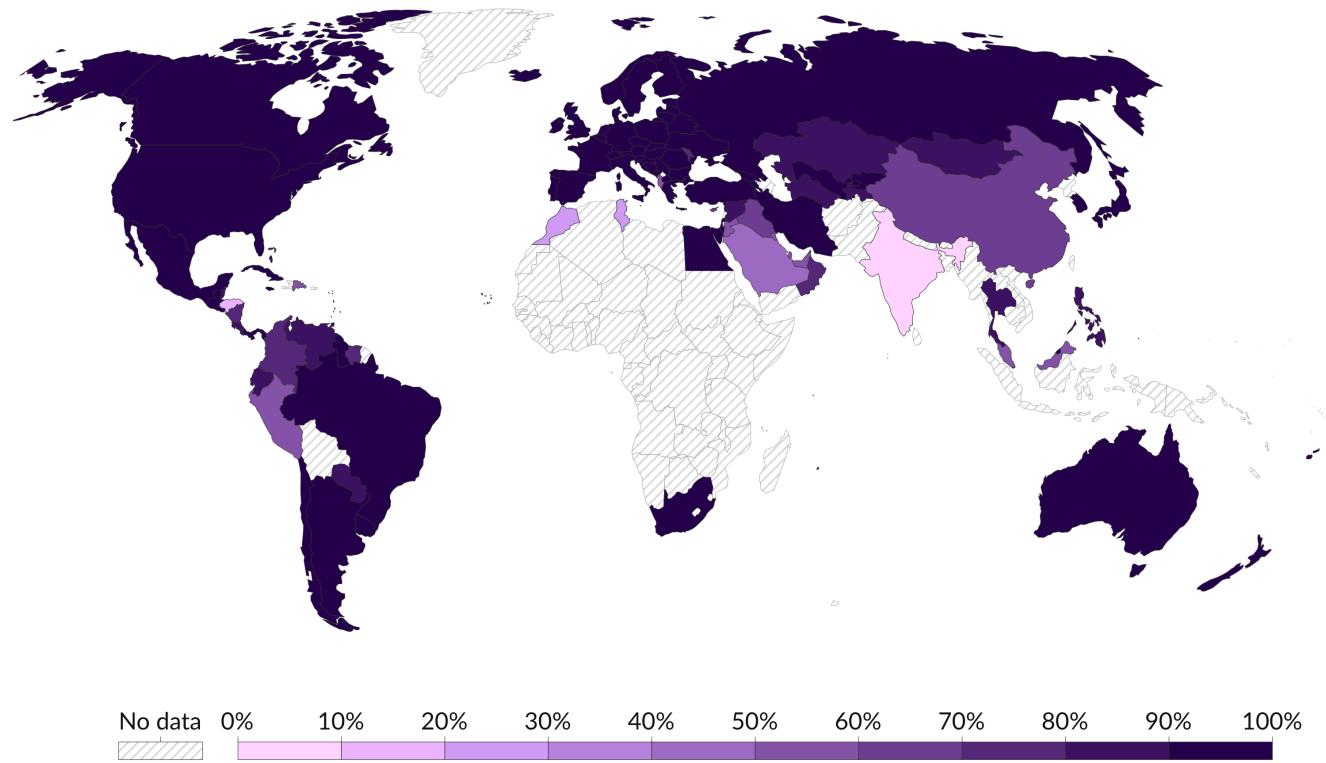
In India:

- around 80% of deaths are officially registered.
- around 10% are registered with a cause of death

Share of deaths for which the cause is registered

The number of deaths that have been registered with cause-of-death information in a country's vital registration system as a share of total estimated deaths.

Our World
in Data



Source: WHO, Global Health Observatory (2022)

Note: Data points are taken as single-year observations between 2007 and 2016, depending on the country.

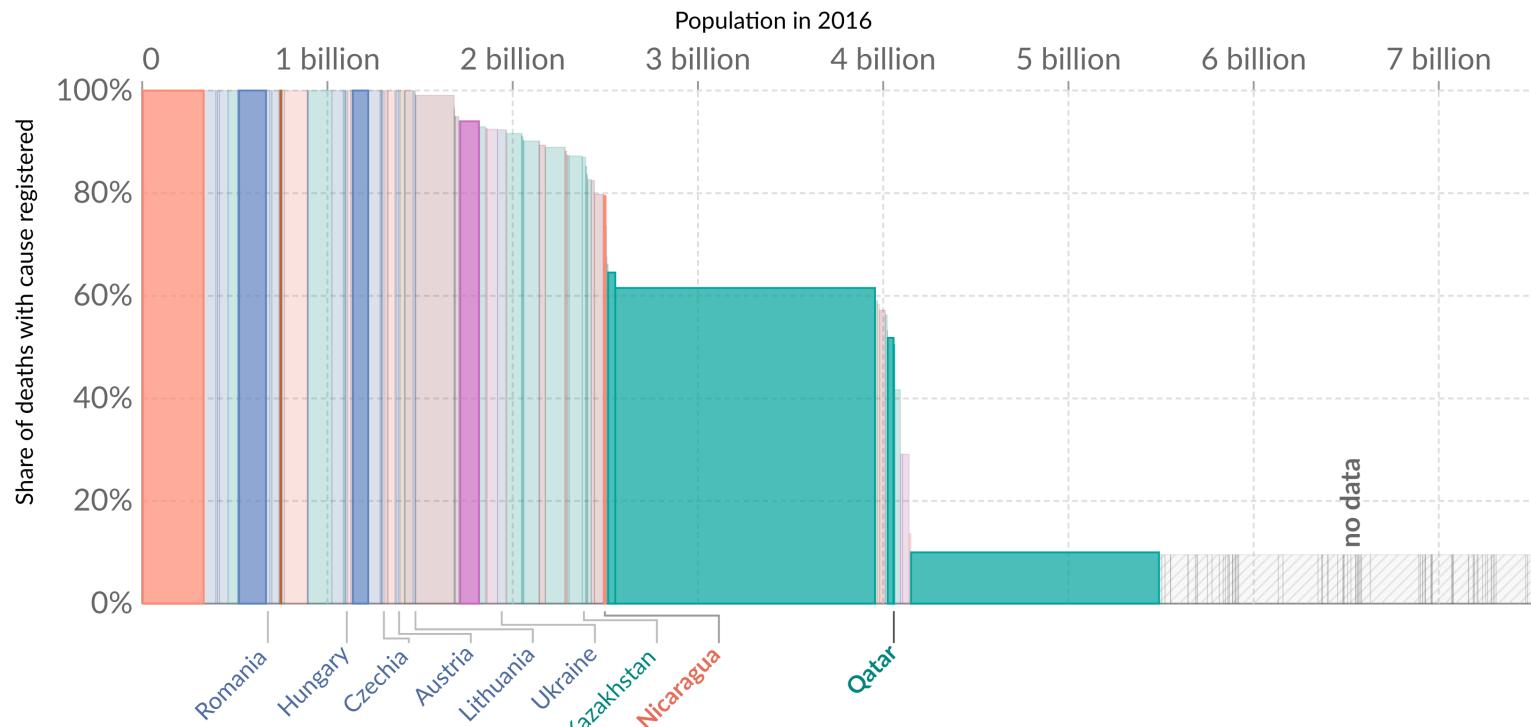
OurWorldInData.org/causes-of-death • CC BY

World population living in countries with cause-of-death registration

Our World
in Data

This shows the share of deaths that have been registered with cause-of-death information in a country's vital registration system.

Africa Antarctica Asia Europe North America Oceania South America



Most of the world lives in countries without cause-of-death data.

Source: WHO, Global Health Observatory (2022), Gapminder (v6); UN (2022); HYDE (v3.2); Gapminder (Systema Globalis)

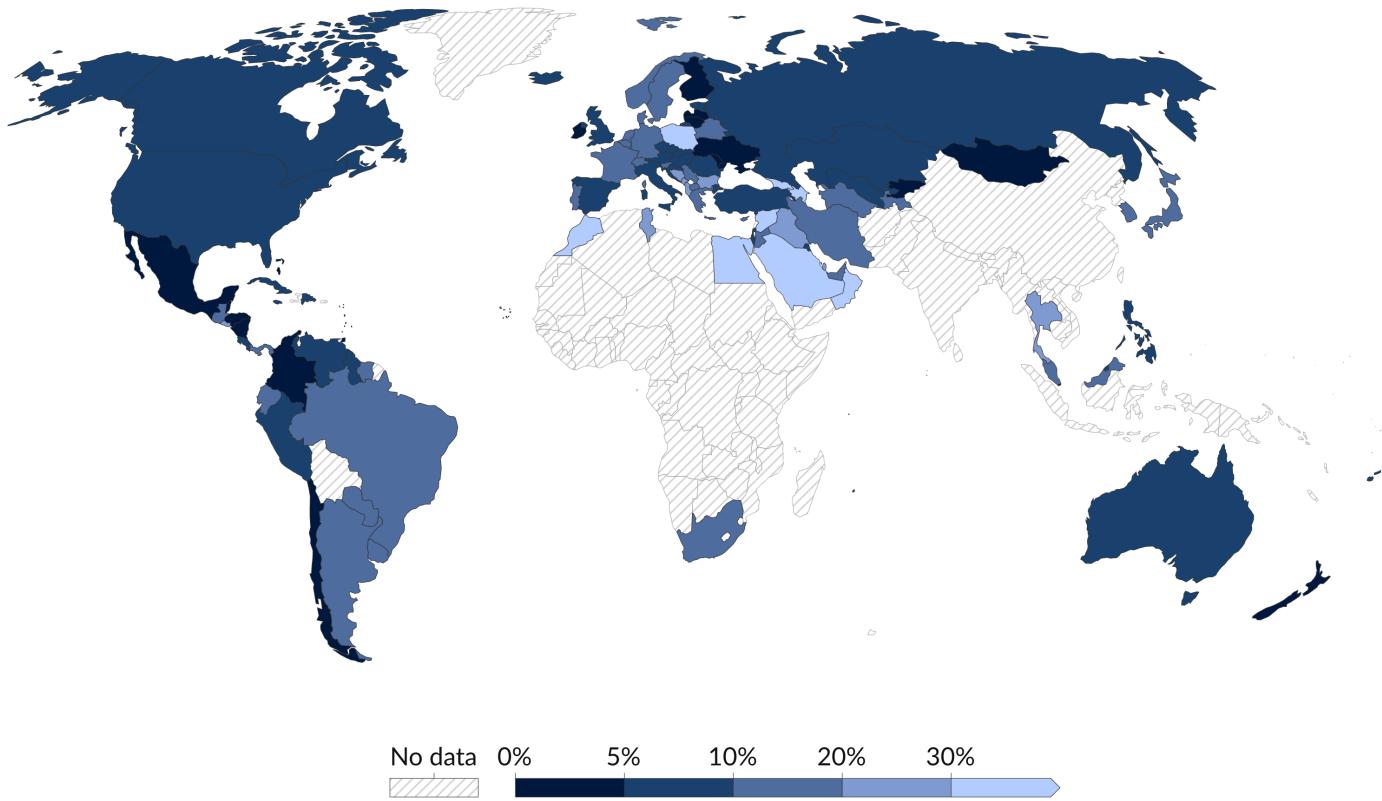
Note: Data points are taken as single-year observations between 2007 and 2016, depending on the country.

OurWorldInData.org/causes-of-death • CC BY

Deaths registered with an ill-defined cause of death

The share of total deaths that were registered with an ill-defined cause of death. This includes causes that are not specific, or when they refer to symptoms rather than an underlying cause.

Our World
in Data



3. Many deaths aren't given a precise cause of death

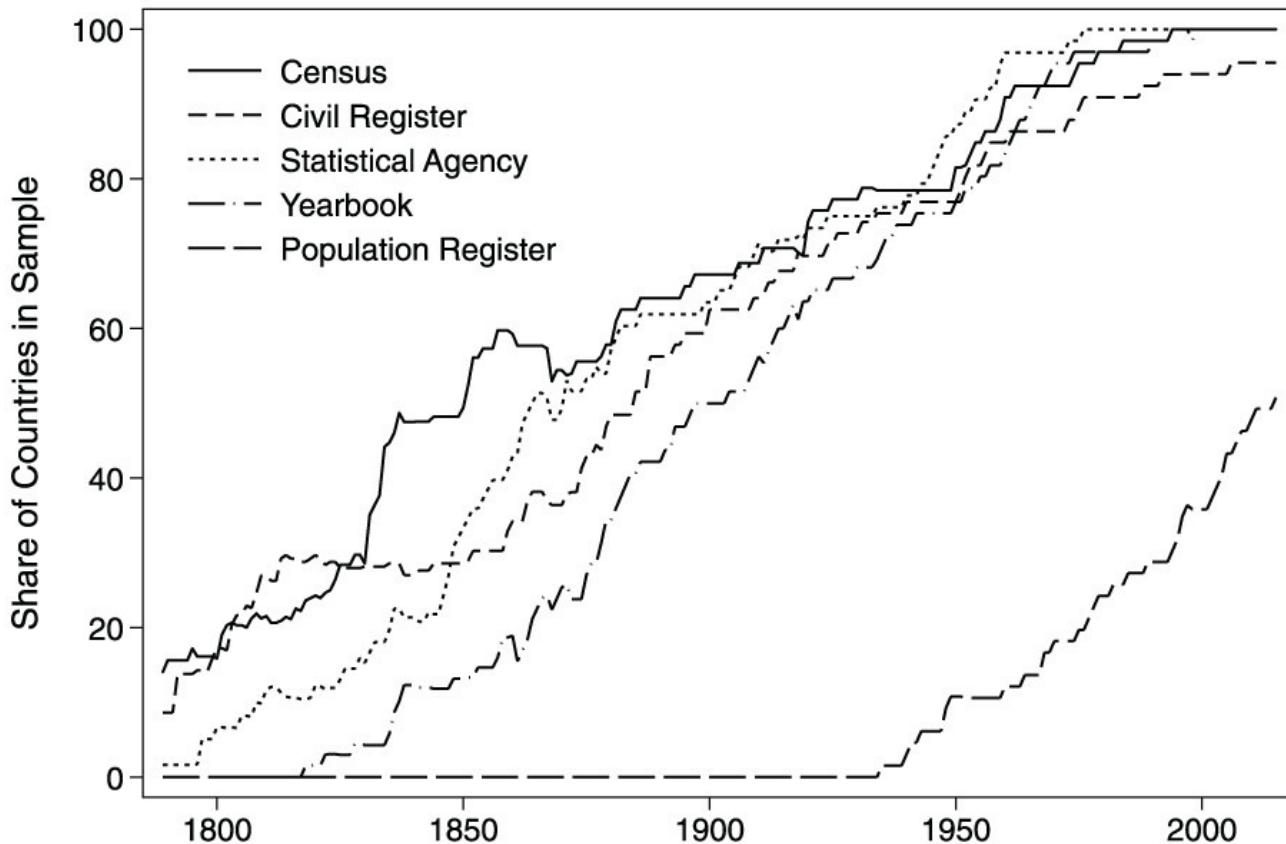
In many countries, people are given causes of death which are 'ill-defined'.

e.g. 'chest pain', 'sudden death', 'cancer of unknown primary site', 'event of undetermined intent'

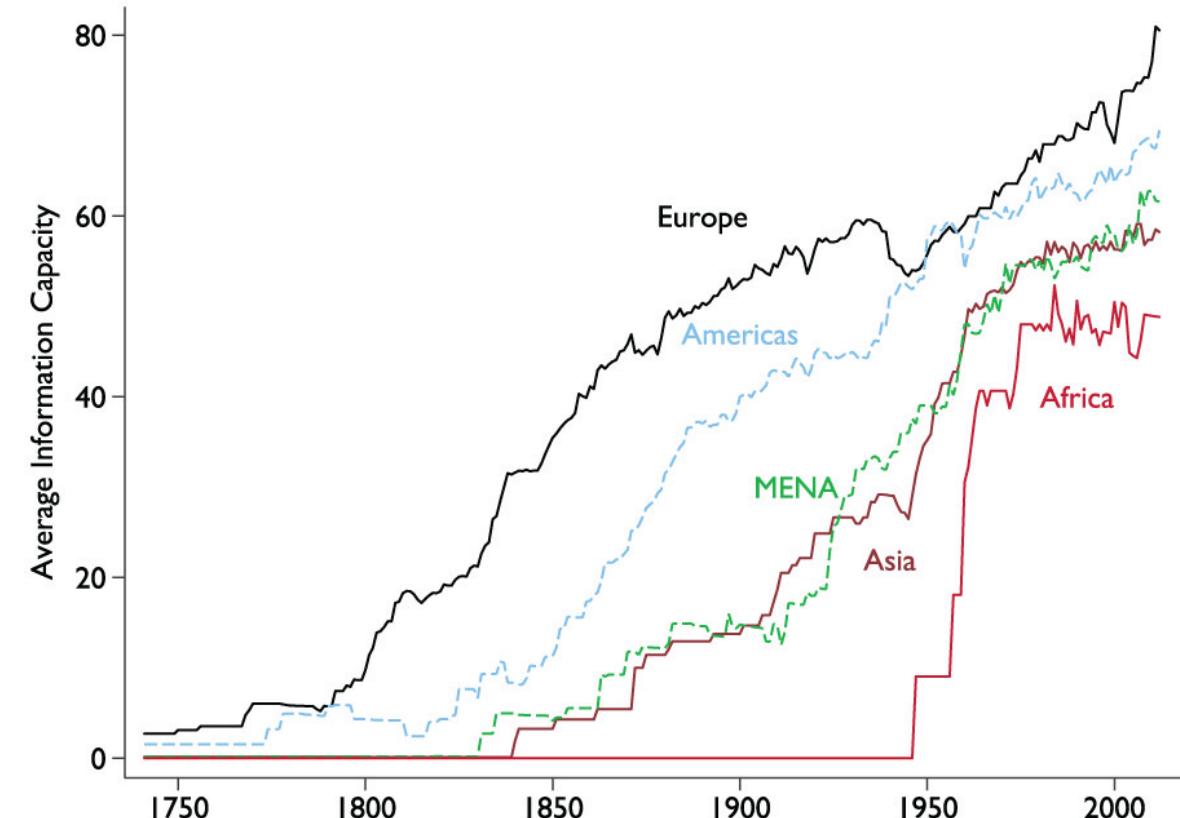
Source: WHO, Global Health Observatory (2022)

Note: Data points are taken as single-year observations between 2007 and 2016, depending on the country.

OurWorldInData.org/causes-of-death • CC BY



Share of countries with population data
(Brambor et al., 2020)



Information capacity across world regions.
(Brambor et al., 2020)

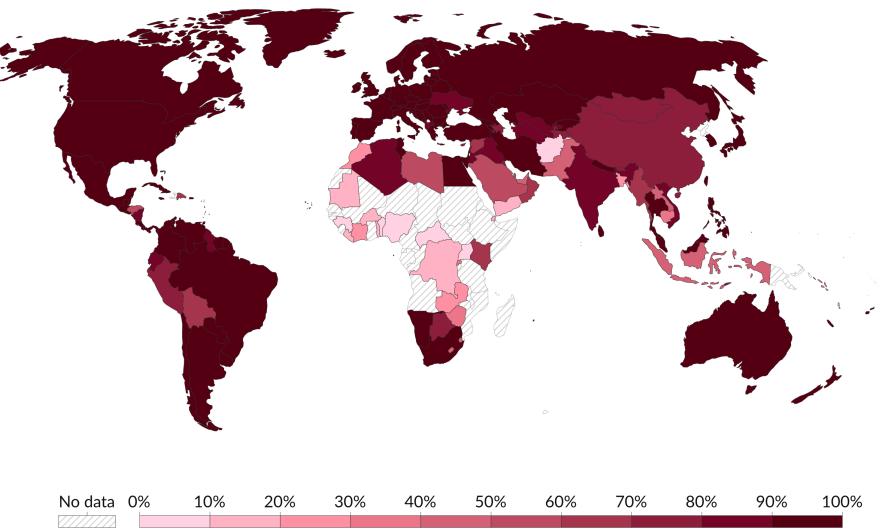
Civil registries – where births and deaths are recorded – are modern.

Many countries still don't have them, especially in Africa and parts of Asia.

Censuses don't collect data on causes of death, and only take place once every ~10 years.

Share of deaths that are registered, 2019

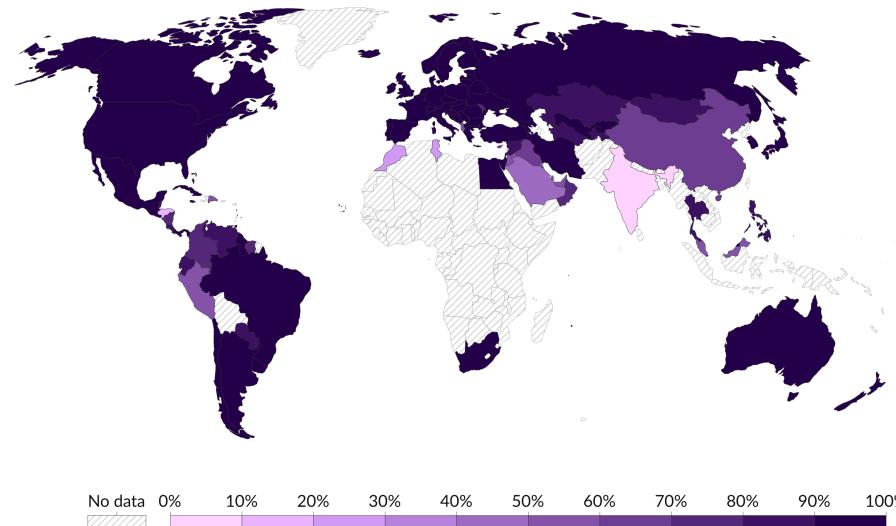
The number of deaths reported in a country's vital registration system as a share of total expected deaths. Expected deaths are taken as the average of estimates from three international sources: the UN, WHO, and IHME.



Source: Karlinsky, A. (2021). International Completeness of Death Registration 2015-2019. OurWorldInData.org/world-population-growth • CC BY

Share of deaths for which the cause is registered

The number of deaths that have been registered with cause-of-death information in a country's vital registration system as a share of total estimated deaths.

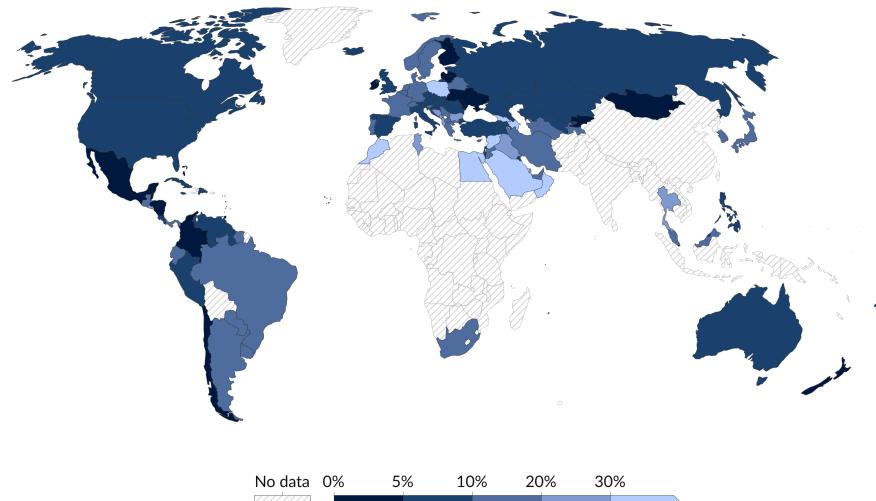


Source: WHO, Global Health Observatory (2022) OurWorldInData.org/causes-of-death • CC BY

Note: Data points are taken as single-year observations between 2007 and 2016, depending on the country.

Deaths registered with an ill-defined cause of death

The share of total deaths that were registered with an ill-defined cause of death. This includes causes that are not specific, or when they refer to symptoms rather than an underlying cause.



Source: WHO, Global Health Observatory (2022)

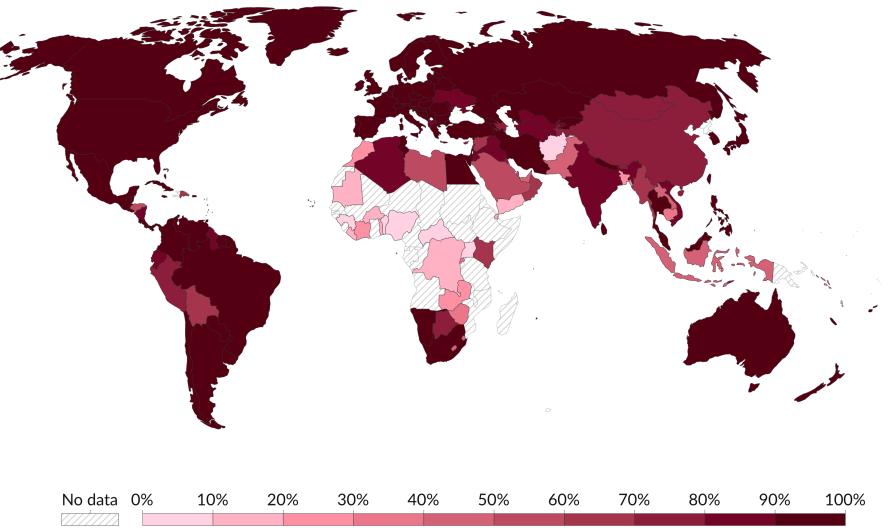
Note: Data points are taken as single-year observations between 2007 and 2016, depending on the country.

In total,

- Many deaths aren't registered at all
- Even among those that are, many aren't registered with a cause
- Even among those that are, many lack a precise cause of death

Share of deaths that are registered, 2019

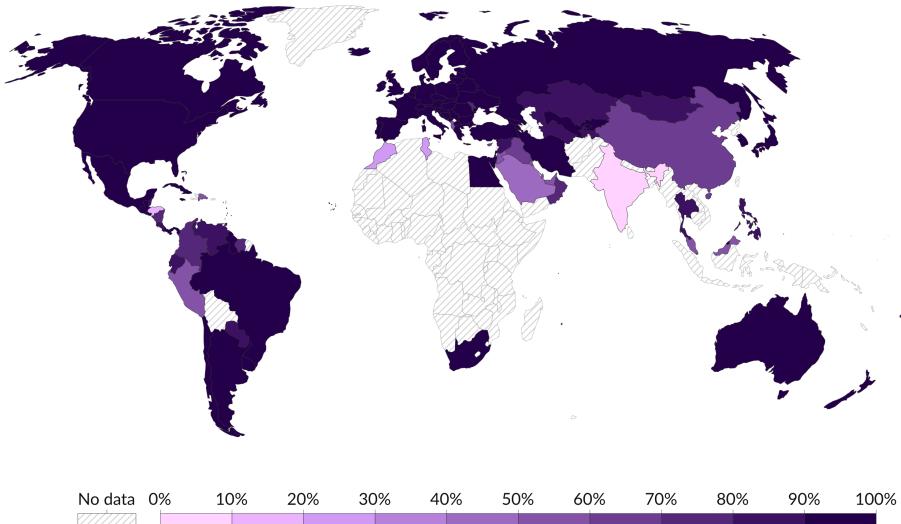
The number of deaths reported in a country's vital registration system as a share of total expected deaths. Expected deaths are taken as the average of estimates from three international sources: the UN, WHO, and IHME.



Source: Karlinsky, A. (2021). International Completeness of Death Registration 2015-2019. OurWorldInData.org/world-population-growth • CC BY

Share of deaths for which the cause is registered

The number of deaths that have been registered with cause-of-death information in a country's vital registration system as a share of total estimated deaths.



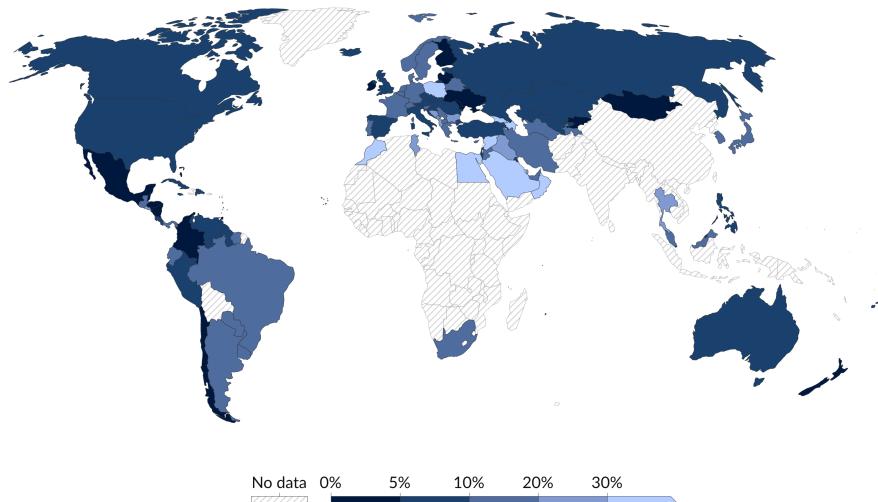
Source: WHO, Global Health Observatory (2022)

Note: Data points are taken as single-year observations between 2007 and 2016, depending on the country.

OurWorldInData.org/causes-of-death • CC BY

Deaths registered with an ill-defined cause of death

The share of total deaths that were registered with an ill-defined cause of death. This includes causes that are not specific, or when they refer to symptoms rather than an underlying cause.

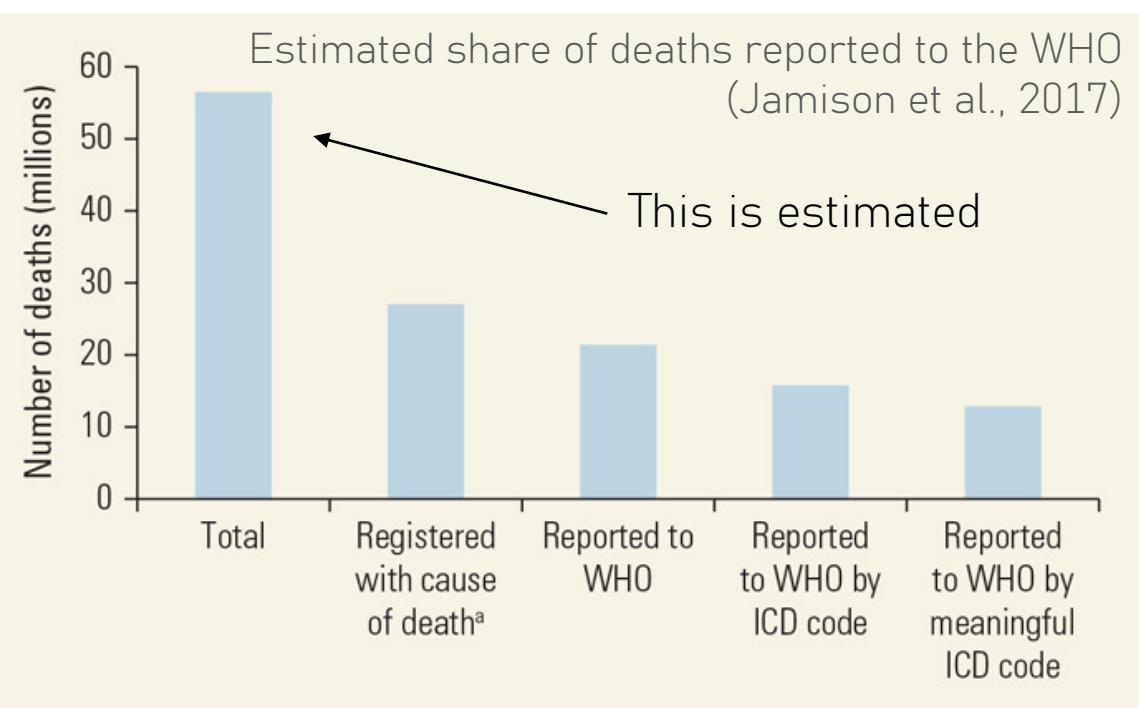


Source: WHO, Global Health Observatory (2022)

Note: Data points are taken as single-year observations between 2007 and 2016, depending on the country.

Estimated share of deaths reported to the WHO (Jamison et al., 2017)

This is estimated



Why?

Why?



- **Civil institutions:**
Some countries lack a functioning civil registry to keep records of births and deaths.



- **Legal factors:**
Death registration doesn't have legal consequences, like inheritance or insurance, or penalties.



- **Labor, health infrastructure and technology:**
Some countries lack doctors, hospitals, testing, medical records and coroners to determine each person's cause of death.



- **Cultural factors:**
Some causes of death are stigmatized, like HIV/AIDS and suicide, and are misreported.

Many people die outside hospitals.

- These are also rough estimates
 - Data only comes from 49 countries, between 2005 and 2019.
 - Without a doctor and medical records, they may not be given a cause of death on the certificate.
- Countries also lack coroners & autopsies

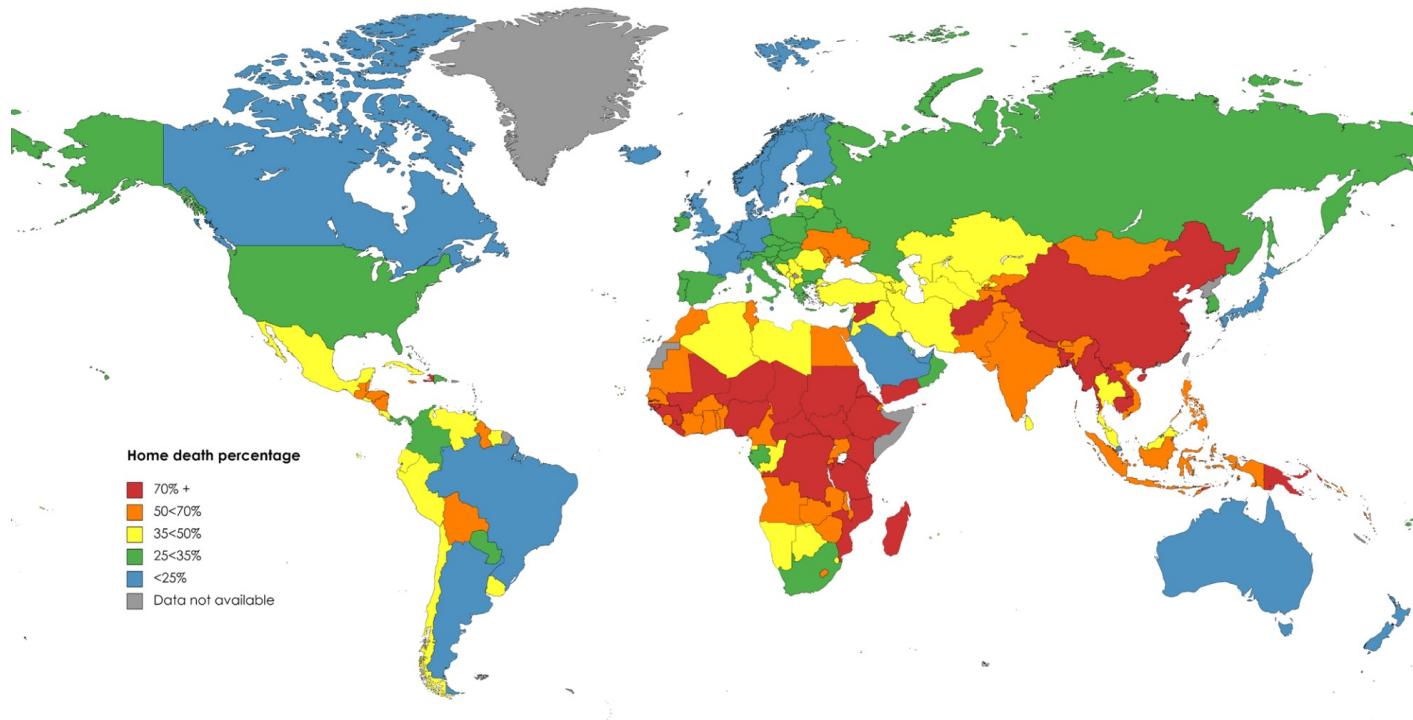
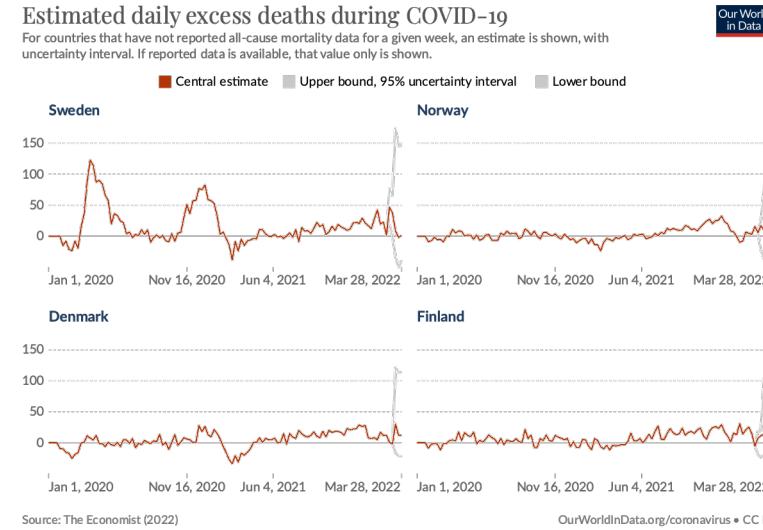


Figure 1 Map of home death percentage* by country, 2019. *Based on estimated and observed home death percentage in online supplemental table 2. In some locations home death percentage could not be estimated because data were not available for all covariates in the model. Map created using mapchart.net.

Estimated share of deaths that occur at home.
(Adair, 2021)



So how do we know what people
are dying from?



So how do we know what people are dying from?

- Sub-national studies & surveys
 - “Excess deaths” compared to a baseline
 - During Covid, this included household surveys and investigations into the number of burials
 - ❖ But this method is only appropriate when the excess can be attributed to a factor, like an epidemic or a disaster
 - ❖ We may lack data to know what the baseline is
 - Estimates from other countries
 - By comparing their other characteristics, e.g. poverty, climate, age-structure, healthcare spending, etc.
 - ❖ But we may lack good cause-of-death data from similar countries to extrapolate from
 - “Verbal autopsies” – interviews of relatives



The Million Death Study

A large study in India conducted between 1998–2014. "Nationally representative."

2.4 million households were interviewed, once every 6 months.

They were asked about deceased relatives and their symptoms before death.

Doctors used their answers to predict the causes of death.

Results were extrapolated to the rest of India (1.3 billion people in 2014).

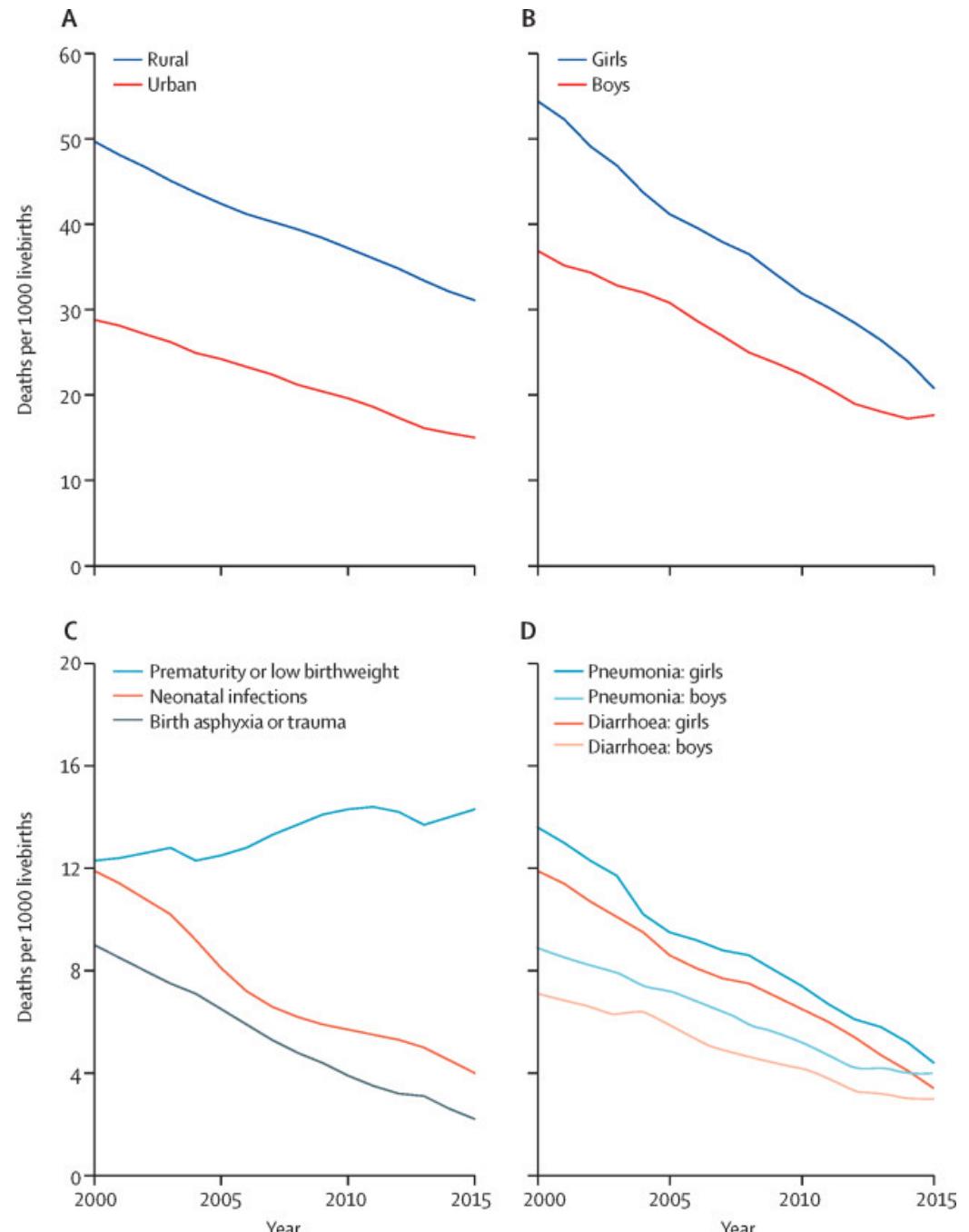
Budget of 2 million USD (\$1 per household)



The Million Death Study

A few results:

Infant and child mortality declined across causes, regions, and gender.

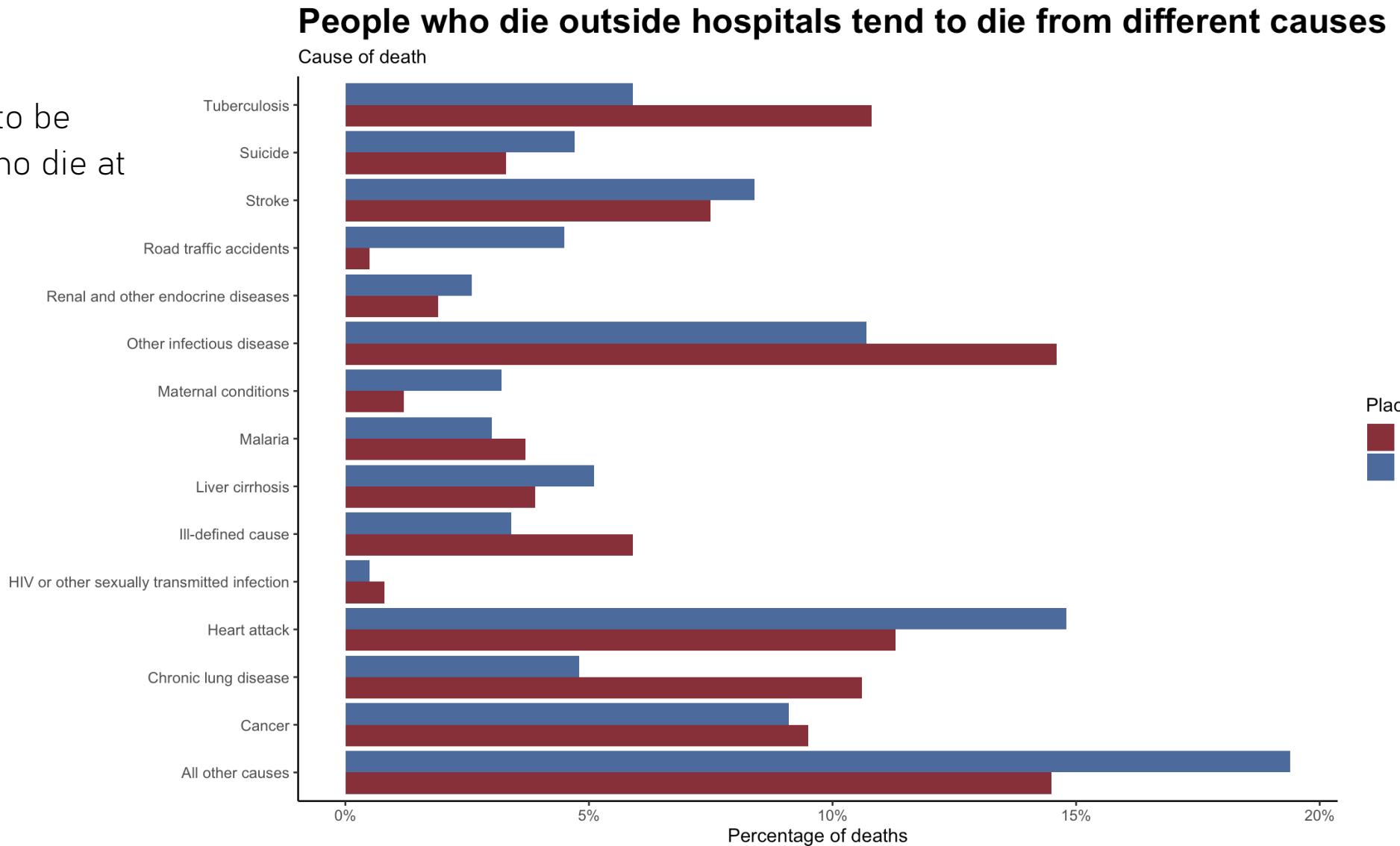


Trends in infant and child mortality in India.
(Million Death Study collaborators, 2017)

The Million Death Study

A few results:

Causes of death tend to be different for people who die at home.



The Million Death Study

A few results:

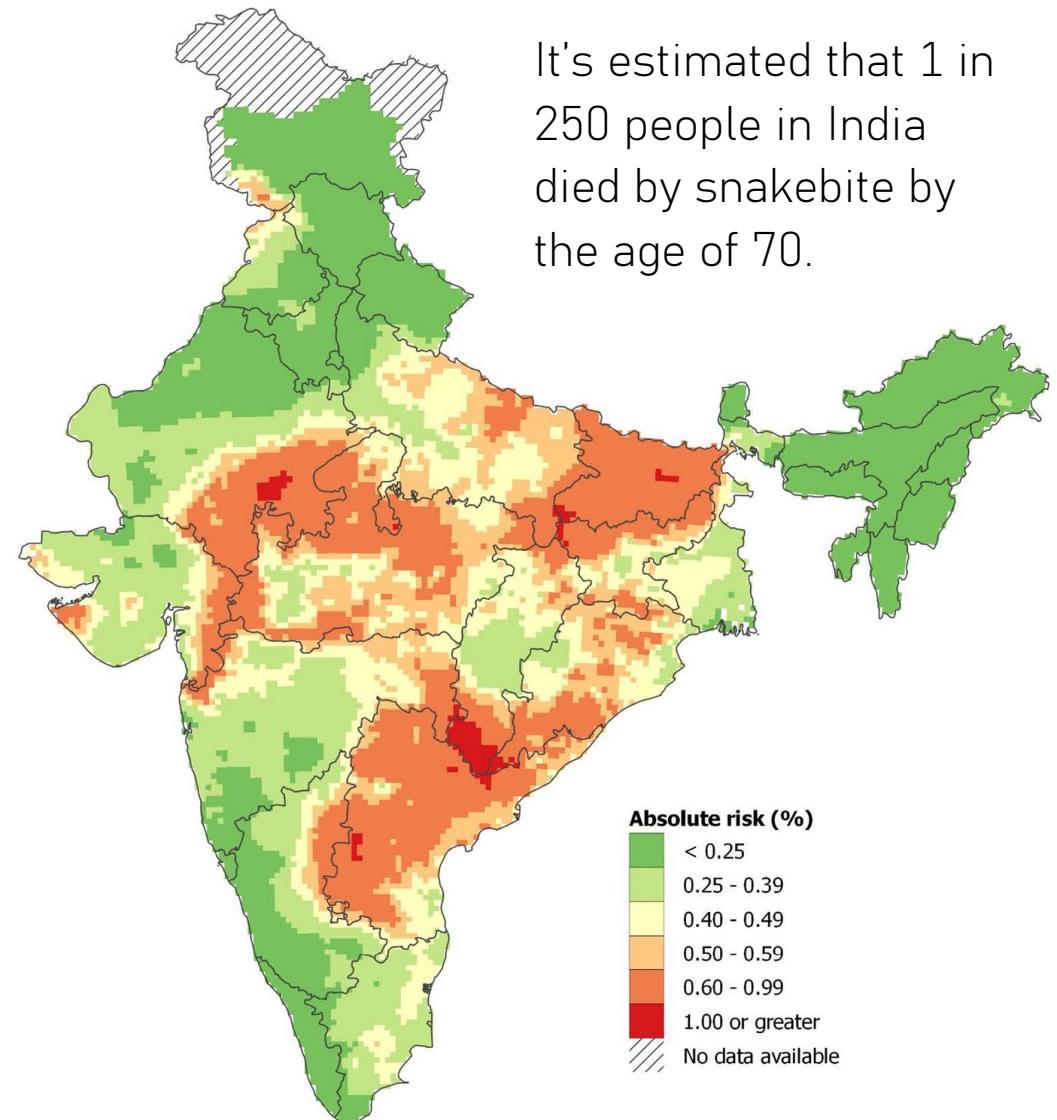
Around 45,900 die from snakebites in India per year.
(The WHO estimated between 11,000 – 15,000. The Indian govt recorded around 1,000)

About 94% of snakebite deaths occurred in rural areas.

77% occurred out of hospital.

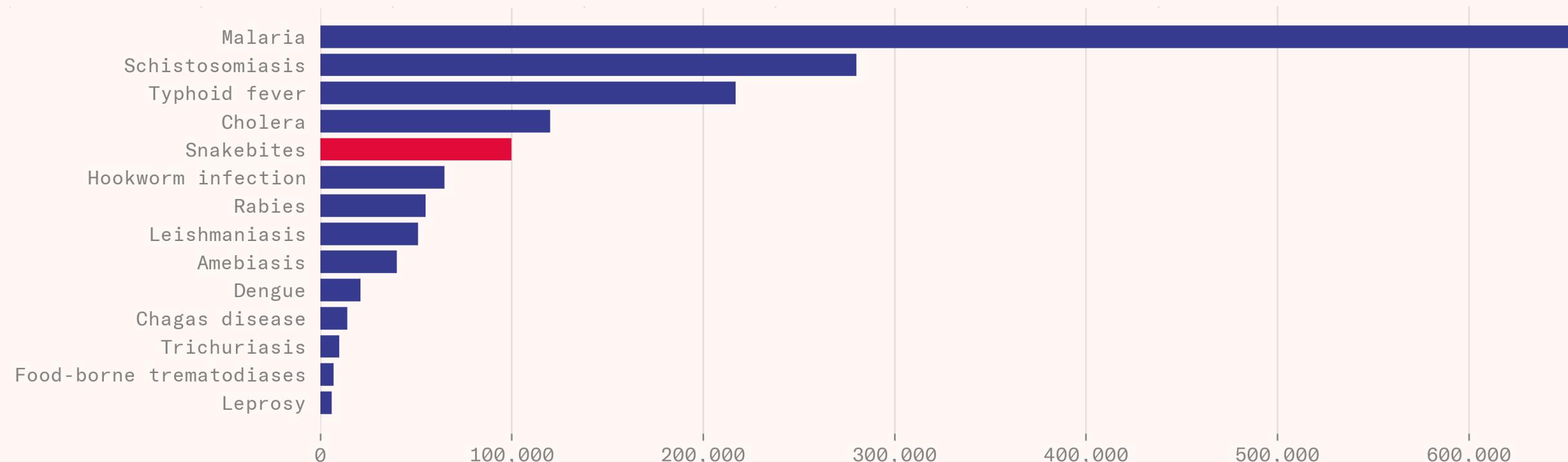
- This led to revisions to the WHO's estimates.

Lifetime risk of death from snakebites in India.
(Suraweera et al., 2020)



Annual global deaths from tropical diseases against snakebites

[source: Hotez (2012)]



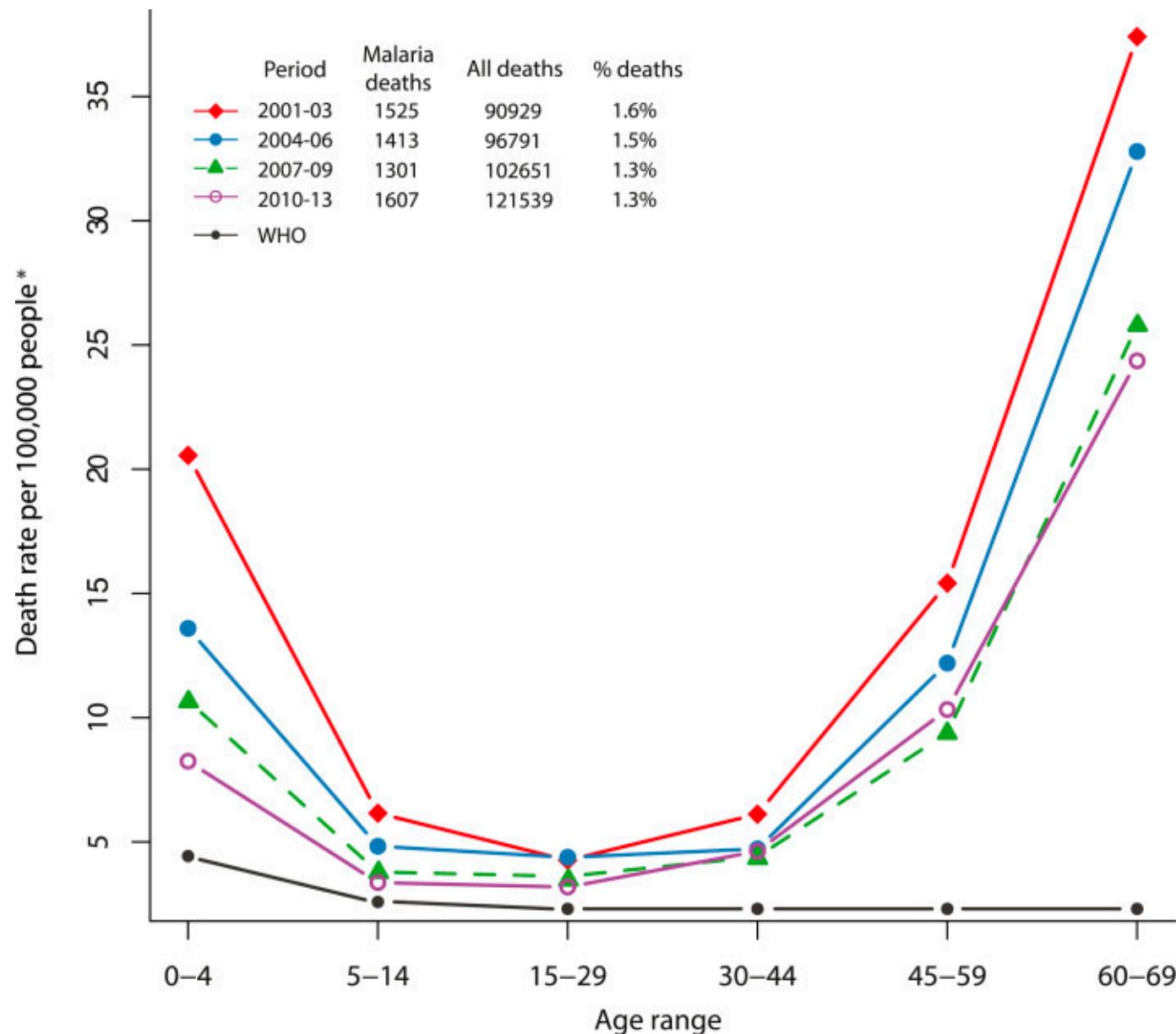
From: worksinprogress.co
Made by: joejamesronan.com

The Million Death Study

Mortality from malaria in India by age.
(Gelband et al., 2020)

A few results:

200,000 estimated deaths from
malaria in India per year (13x the
WHO's estimate then)



The Million Death Study

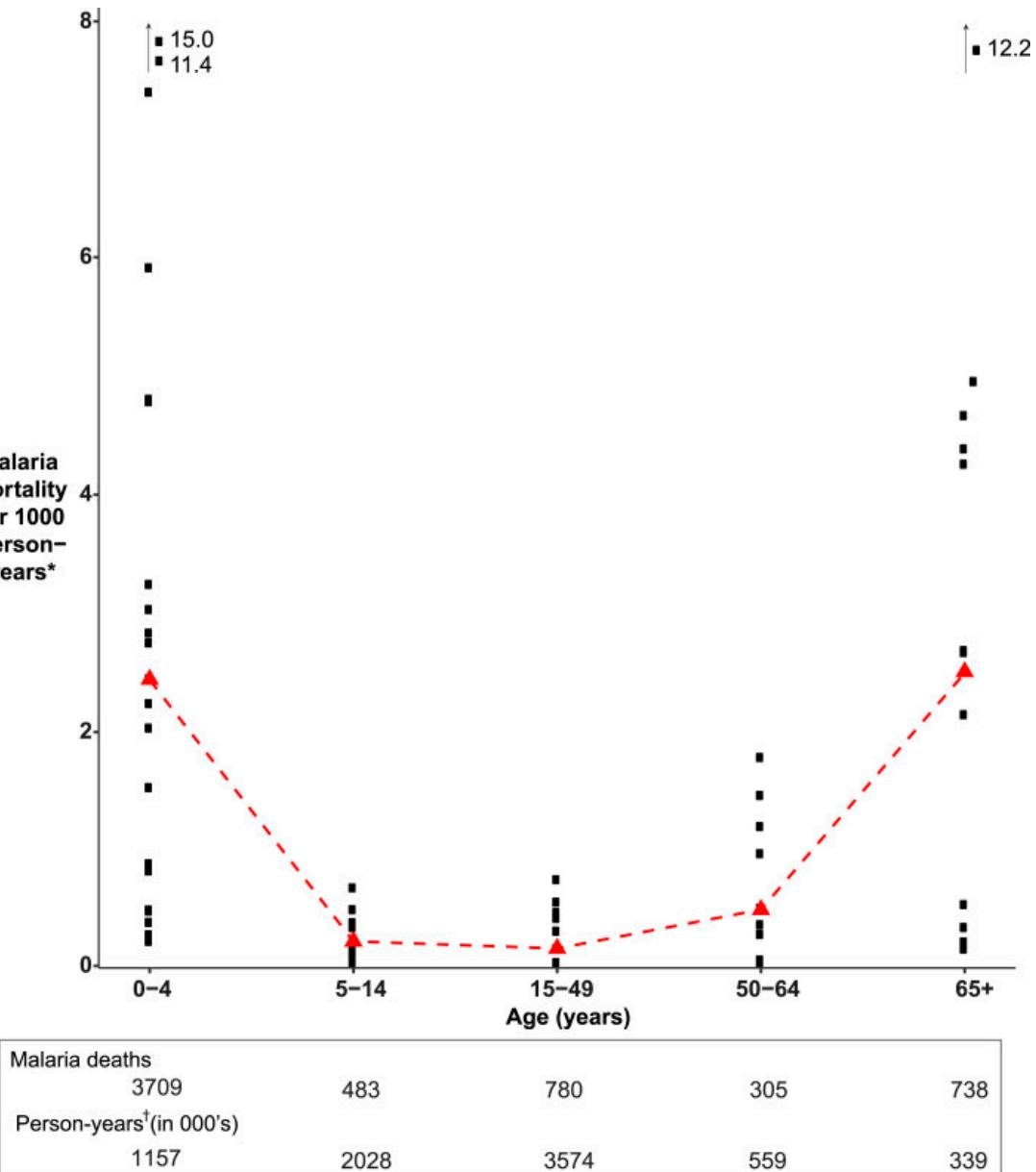
A few results:

200,000 estimated deaths from malaria in India per year ([13x the WHO's estimate then](#))

This age-mortality curve has also been found in the INDEPTH study (2000–2012),

- includes Burkina Faso, Cote d'Ivoire, Ghana, Kenya, Nairobi, Senegal, South Africa and Gambia.

Mortality from malaria across sites in Africa by age (Gelband et al., 2020)



The Million Death Study

A few results:

200,000 estimated deaths from malaria in India per year (13x the WHO's estimate then)

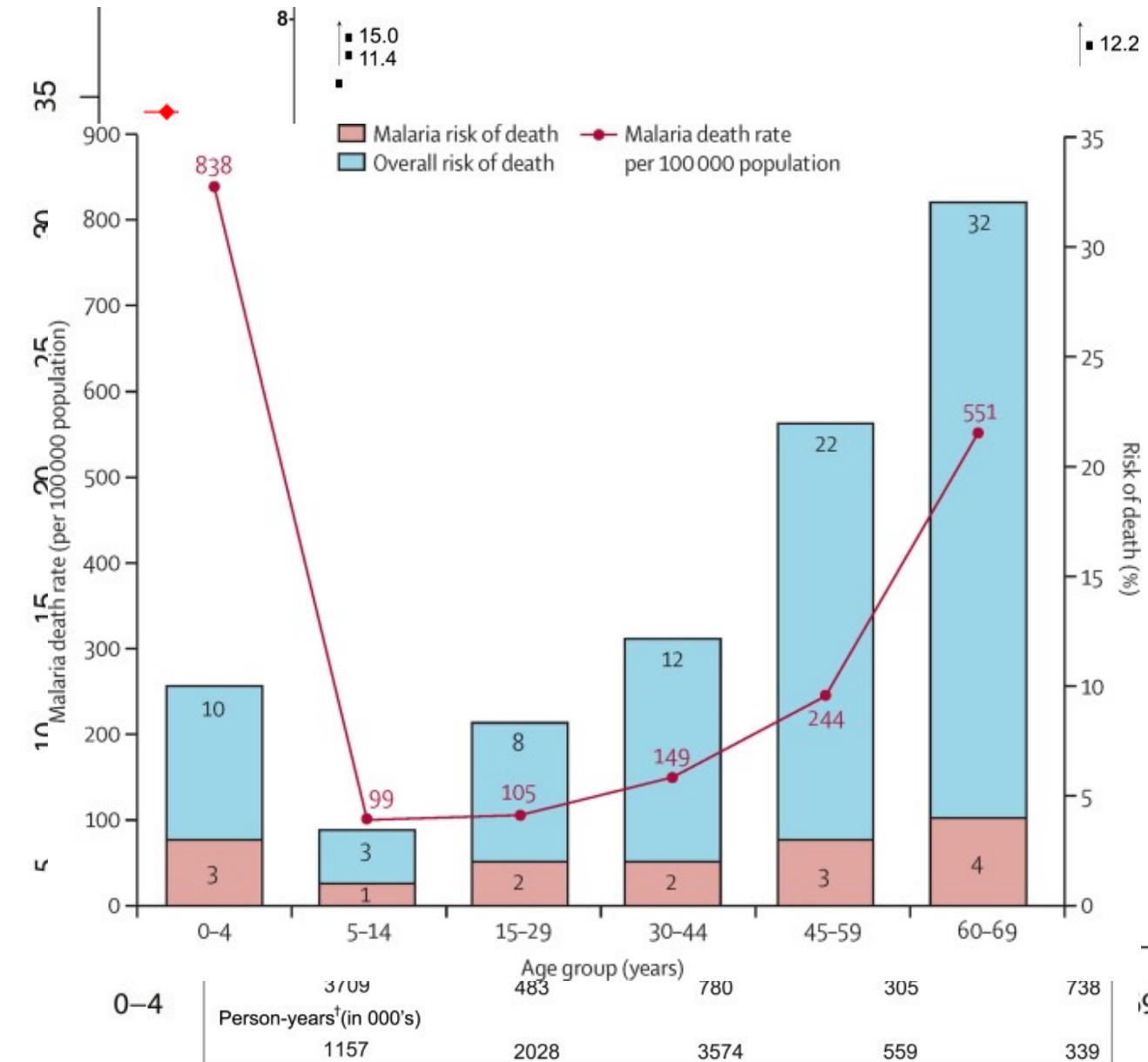
This age-mortality curve has also been found in the INDEPTH study (2000–2012),

- includes Burkina Faso, Cote d'Ivoire, Ghana, Kenya, Nairobi, Senegal, South Africa and Gambia.

And in Sierra Leone (2018–2020).

But there's still uncertainty!

Mortality from malaria across sub-Saharan Africa by age.
(Gelband et al., 2020)



Limitations:



In verbal autopsy studies...

- relatives may not recall or know the relevant symptoms
- lack of medical records & testing
- they may misreport evidence especially for stigmatized causes of death, like HIV/AIDS and suicide
- studies don't include the entire population – results are extrapolated

These limitations are larger for rare causes of death & causes that are difficult to distinguish – e.g. malaria vs other mosquito-borne diseases.



Even though they have limitations,
they are a step forward

- They give us benchmark figures
- They help understand causes of deaths outside hospitals
- *They improve estimates for other countries*
 - Better data from one region also means improved estimates for similar regions

Why were estimates so far off?

- Estimates tend to come from data from non-tropical countries & urban areas.
 - Where most people die in hospitals

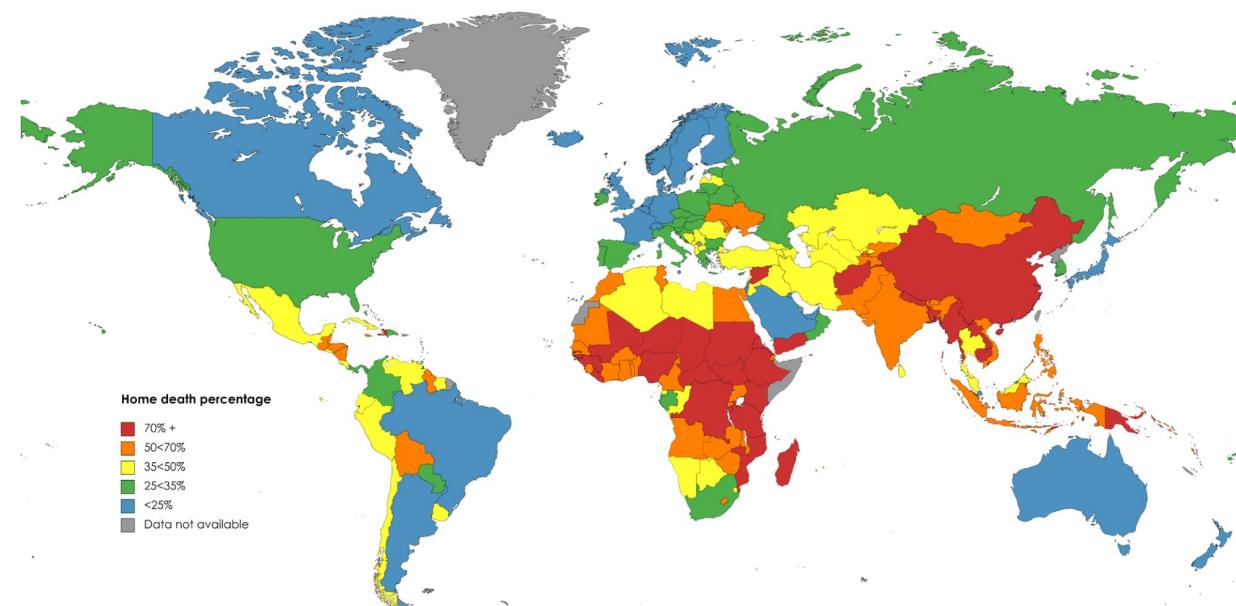
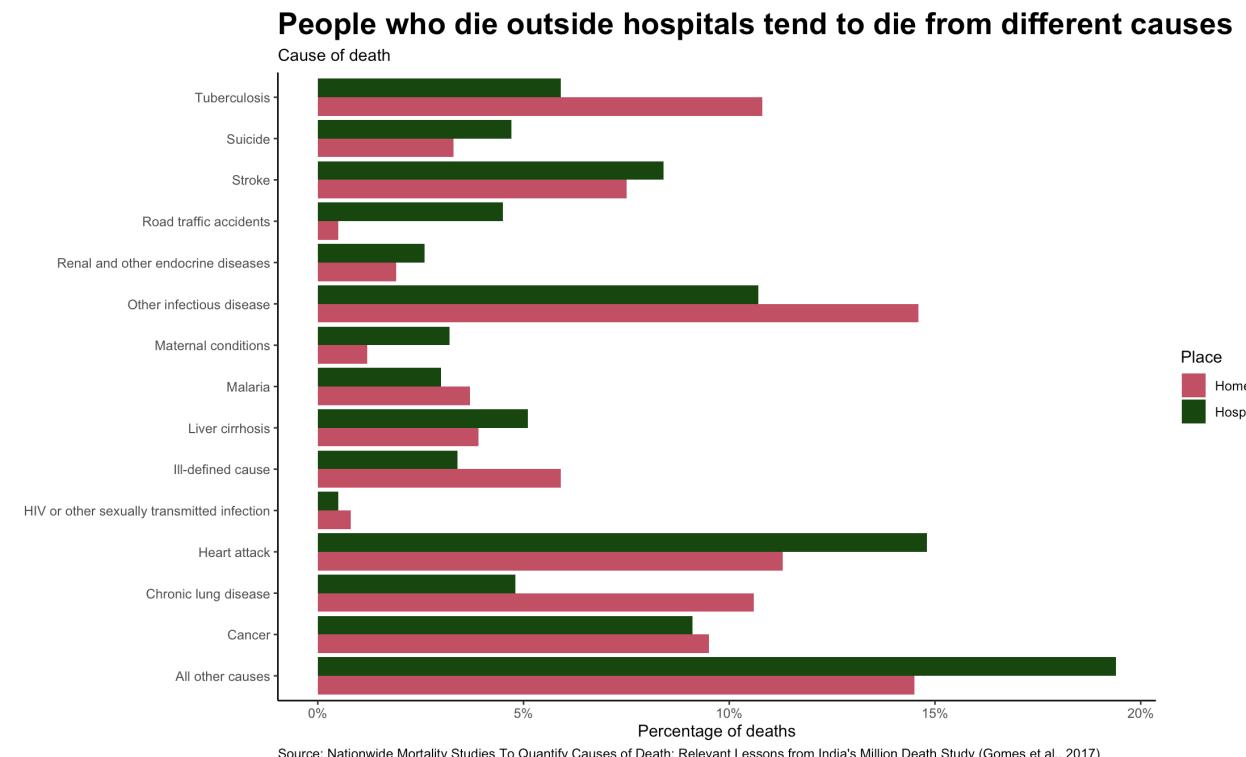


Figure 1 Map of home death percentage* by country, 2019. *Based on estimated and observed home death percentage in online supplemental table 2. In some locations home death percentage could not be estimated because data were not available for all covariates in the model. Map created using mapchart.net.

•

Why were estimates so far off?

- Estimates tend to come from data from non-tropical countries & urban areas.
 - Where most people die in hospitals
 - Where some causes of death (like snakebites) are less common



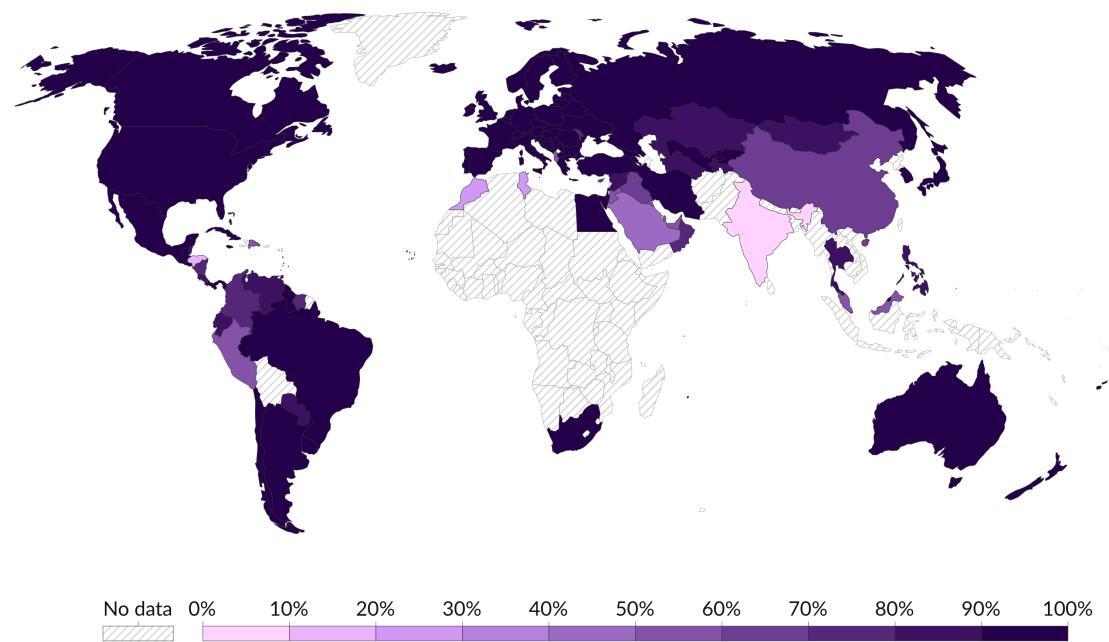
Why were estimates so far off?

- Estimates tend to come from data from non-tropical countries & urban areas.
 - Where most people die in hospitals
 - Where some causes of death (like snakebites) are less common
 - Where most deaths are counted
- Data is missing *systematically*.
- This makes it difficult to extrapolate.

Share of deaths for which the cause is registered

The number of deaths that have been registered with cause-of-death information in a country's vital registration system as a share of total estimated deaths.

Our World
in Data



Source: WHO, Global Health Observatory (2022)

Note: Data points are taken as single-year observations between 2007 and 2016, depending on the country.

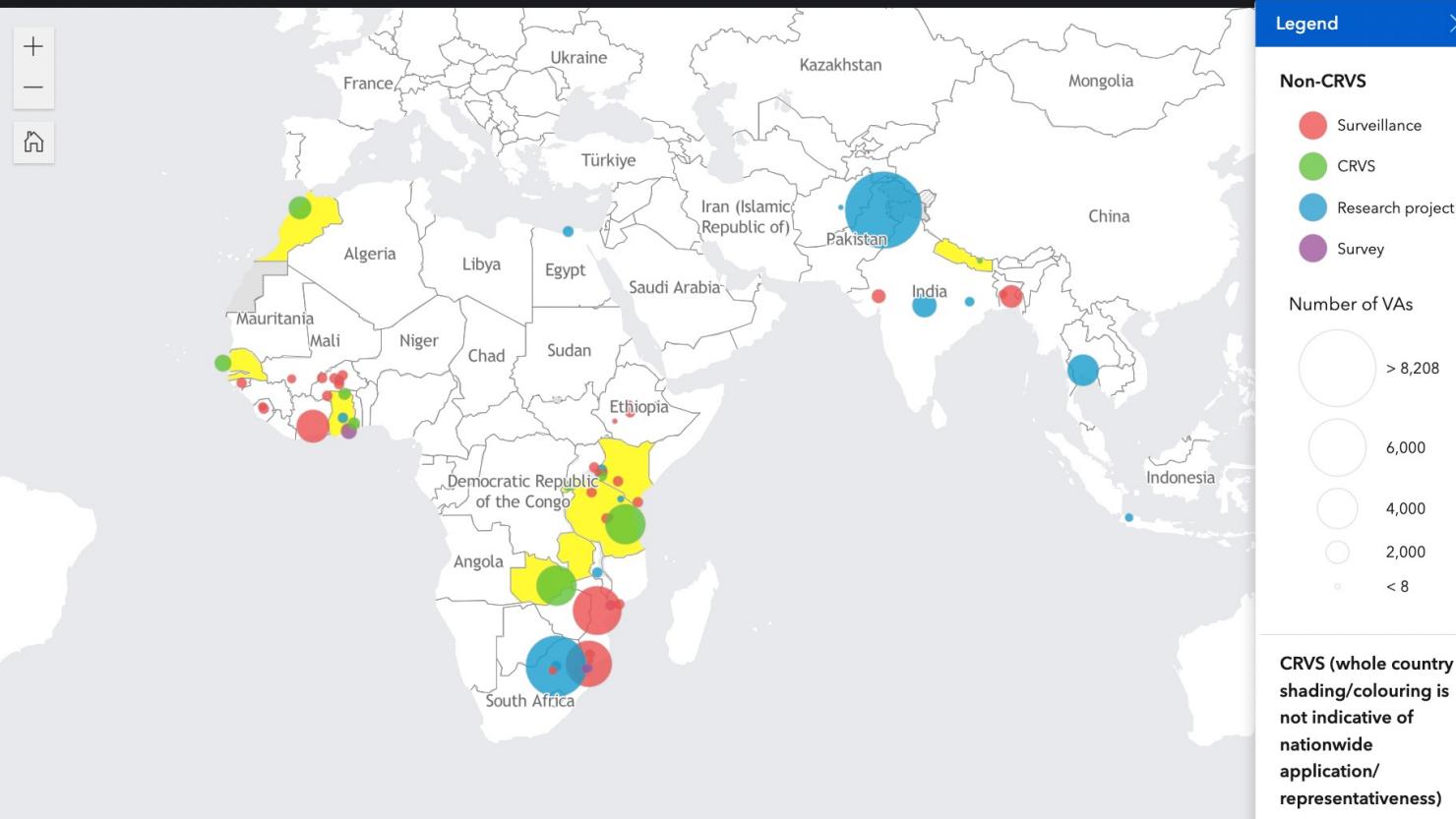
OurWorldInData.org/causes-of-death • CC BY



“Verbal autopsy studies” have also been conducted in other regions

They have also led to revisions of global estimates

What are we missing & underestimating in other countries?



Verbal autopsies performed by region.
(WHO, 2022)



Why does all this data matter?

Why does all this data matter?

- To put all other numbers in context.
 - For example, to calculate numbers per capita, rates per 100,000 people, % of people, etc.
- Planning & decision-making
 - By individuals, governments, NGOs and private companies
 - e.g. Which conditions are going untreated?
 - Where are doctors needed?
 - How many doses of medicines and vaccines are needed?
 - How much progress are we making?
 - Are new problems emerging?

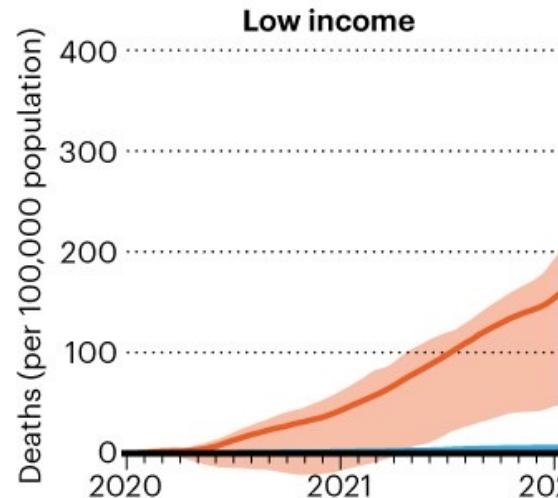
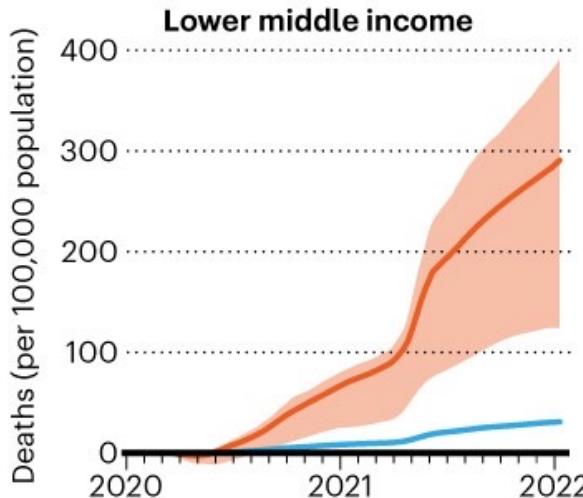
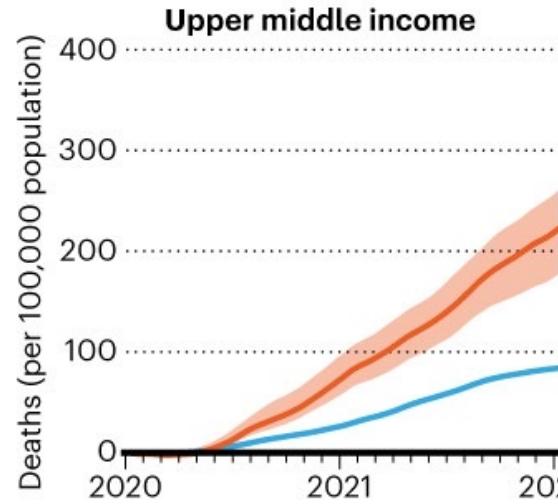
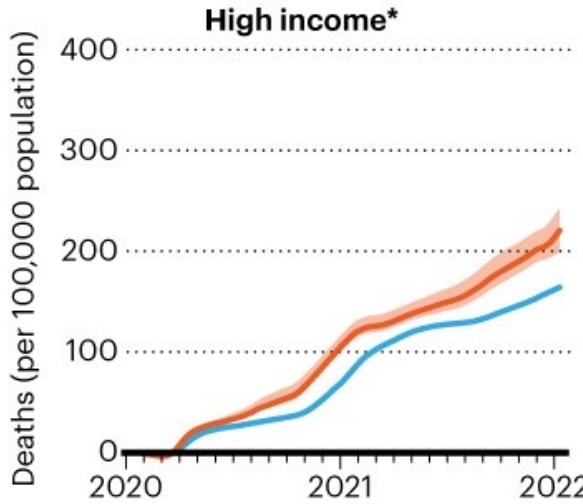
Even if different estimates agree that something is a large problem, mis-estimating numbers by a factor of 5-10x is a big deal.

RICH AND POOR

Official figures suggest that wealthy countries had the highest number of deaths per capita during the pandemic. But a model that estimates excess deaths suggests that is false: lower middle-income countries might have been hit hardest.

Confirmed COVID-19 deaths *The Economist* excess deaths estimate

95% confidence interval





Missing data in global health

- Is widespread
- Is highly unequal
- Can have large consequences

Doesn't mean we're completely in the dark

Isn't an excuse to not act

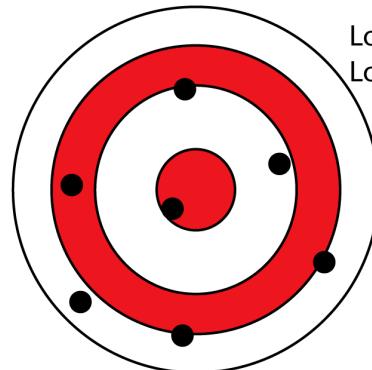
➤ Each step forward improves other estimates



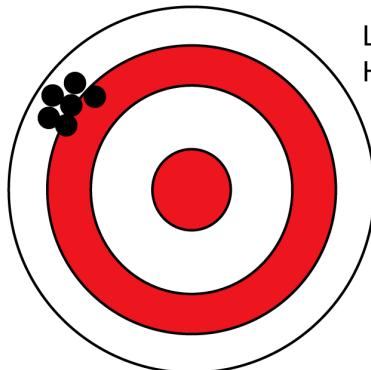
What we should aim for

What we should aim for

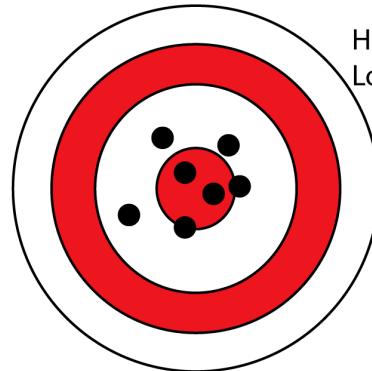
- More high-quality **underlying data**, less need for extrapolation
- Routine, long-term data
- More accurate & precise estimates
- Numbers we can be more confident in



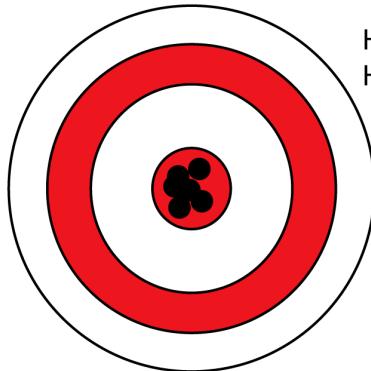
Low accuracy
Low precision



Low accuracy
High precision



High accuracy
Low precision



High accuracy
High precision

How to get there

How to get there

Readers & data communicators:

- Data quality should be emphasized alongside numbers – it affects how we interpret them, like having units on a chart

Researchers:

- Highlight missing data, uncertainty, and gaps in knowledge
- Deliberately study under-represented populations

Countries:

- Labour, technology, institutions & health infrastructure to collect & share data regularly, in the long-term
- Universal data coverage



Thank you!

Further reading

- Bonde, M. K. (2022). Advancing antivenom. *Works in Progress*, 09. <https://worksinprogress.co/issue/advancing-antivenom>
- Fortner, R. (2023). The tale of snakebite's fleeting spotlight—And why it encapsulates all that's wrong with global health. *BMJ*, p306. <https://doi.org/10.1136/bmj.p306>
- Gelband, H., Bogoch, I. I., Rodriguez, P. S., Ngai, M., Peer, N., Watson, L. K., & Jha, P. (2020). Is Malaria an Important Cause of Death among Adults? *The American Journal of Tropical Medicine and Hygiene*, 103(1), 41–47. <https://doi.org/10.4296/ajtmh.20-0036>
- Jamison, D. T., Gelband, H., Horton, S., Jha, P., Laxminarayan, R., Mock, C. N., & Nugent, R. (Eds.). (2017). *Disease Control Priorities, Third Edition (Volume 9): Improving Health and Reducing Poverty*. The World Bank. <https://doi.org/10.1596/978-1-4648-0527-1>
- Mathers, C. (2017). *Monitoring trends in causes of death – India and the Million Death Study*. <https://colinmathers.com/2017/09/24/monitoring-trends-in-causes-of-death-india-and-the-million-death-study/>

References

- Adam, D. (2022, January). The pandemic's true death toll: Millions more than official counts. *Nature News*. <https://www.nature.com/articles/d41586-022-00104-8>
- Brambor, T., Goenaga, A., Lindvall, J., & Teorell, J. (2020). The Lay of the Land: Information Capacity and the Modern State. *Comparative Political Studies*, 53(2), 175–213. <https://doi.org/10.1177/0010414019843432>
- Jamison, D. T., Gelband, H., Horton, S., Jha, P., Laxminarayan, R., Mock, C. N., & Nugent, R. (Eds.). (2017). *Disease Control Priorities, Third Edition (Volume 9): Improving Health and Reducing Poverty*. The World Bank. <https://doi.org/10.1596/978-1-4648-0527-1>
- Adair, T. (2021). Who dies where? Estimating the percentage of deaths that occur at home. *BMJ Global Health*, 6(9), e006766. <https://doi.org/10.1136/bmjgh-2021-006766>
- Gomes, M., Begum, R., Sati, P., Dikshit, R., Gupta, P. C., Kumar, R., Sheth, J., Habib, A., & Jha, P. (2017). Nationwide Mortality Studies To Quantify Causes Of Death: Relevant Lessons From India's Million Death Study. *Health Affairs*, 36(11), 1887–1895. <https://doi.org/10.1377/hlthaff.2017.0635>
- Causes of neonatal and child mortality in India: A nationally representative mortality survey. (2010). *The Lancet*, 376(9755), 1853–1860. [https://doi.org/10.1016/S0140-6736\(10\)61461-4](https://doi.org/10.1016/S0140-6736(10)61461-4)
- Kasturiratne, A., Wickremasinghe, A. R., de Silva, N., Gunawardena, N. K., Pathmeswaran, A., Premaratna, R., Savioli, L., Laloo, D. G., & de Silva, H. J. (2008). The Global Burden of Snakebite: A Literature Analysis and Modelling Based on Regional Estimates of Envenoming and Deaths. *PLoS Medicine*, 5(11), e218. <https://doi.org/10.1371/journal.pmed.0050218>

References (continued)

- Mohapatra, B., Warrell, D. A., Suraweera, W., Bhatia, P., Dhingra, N., Jotkar, R. M., Rodriguez, P. S., Mishra, K., Whitaker, R., Jha, P., & for the Million Death Study Collaborators. (2011). Snakebite Mortality in India: A Nationally Representative Mortality Survey. *PLoS Neglected Tropical Diseases*, 5(4), e1018. <https://doi.org/10.1371/journal.pntd.0001018>
- Suraweera, W., Warrell, D., Whitaker, R., Menon, G., Rodrigues, R., Fu, S. H., Begum, R., Sati, P., Piyasena, K., Bhatia, M., Brown, P., & Jha, P. (2020). Trends in snakebite deaths in India from 2000 to 2019 in a nationally representative mortality study. *ELife*, 9, e54076. <https://doi.org/10.7554/elife.54076>
- Bonde, M. K. (2022, December). Advancing antivenom. *Works in Progress*, 09. <https://worksinprogress.co/issue/advancing-antivenom>
- Hotez, P. J. (2012). The Four Horsemen of the Apocalypse: Tropical Medicine in the Fight against Plague, Death, Famine, and War † †Presidential address given at the 60th Annual Meeting of the American Society of Tropical Medicine and Hygiene, December 7, 2011, Philadelphia, Pennsylvania. *The American Journal of Tropical Medicine and Hygiene*, 87(1), 3–10. <https://doi.org/10.4269/ajtmh.2012.11-0814>
- Gelband, H., Bogoch, I. I., Rodriguez, P. S., Ngai, M., Peer, N., Watson, L. K., & Jha, P. (2020). Is Malaria an Important Cause of Death among Adults? *The American Journal of Tropical Medicine and Hygiene*, 103(1), 41–47. <https://doi.org/10.4269/ajtmh.20-0036>
- Carshon-Marsh, R., Aimone, A., Ansumana, R., Swaray, I. B., Assalif, A., Musa, A., Meh, C., Smart, F., Hang Fu, S., Newcombe, L., Kamadod, R., Saikia, N., Gelband, H., Jambai, A., & Jha, P. (2022). Child, maternal, and adult mortality in Sierra Leone: Nationally representative mortality survey 2018–20. *The Lancet Global Health*, 10(1), e114–e123. [https://doi.org/10.1016/S2214-109X\(21\)00459-9](https://doi.org/10.1016/S2214-109X(21)00459-9)
- World Health Organization. (2022). *Verbal autopsy standards: The 2022 WHO verbal autopsy instrument*. https://cdn.who.int/media/docs/default-source/classification/other-classifications/autopsy/2022-va-instrument/verbal-autopsy-standards_2022-who-verbal-autopsy-instrument_v1_final.pdf?sfvrsn=c8cf2dda_8