Births, deaths, and other uncertainties in India

CASI Data Seminar series 2023

Aashish Gupta

24 May 2023

Overview

- Records of births, deaths, and people alive are some of the most basic data that states and societies collect

- There are a quite a few sources, with different strengths and limitations

    - Censuses

    - Administrative data (Civil registration, MCCD …)

    - Survey data (Sample Registration, NFHS, IHDS, AHS …)

    - Newer, "innovative", and crisis sources (Insurance, media, disease surveillance)

- Despite data deficiencies, it is still possible to learn quite a lot

- Always helpful to think about signal v noise, biases, and selection into data

Diversion: A long(er) history

- The *Arthashastra* (Rangarajan 1987)

(b) the preparation of the budget and maintenance of detailed accounts of revenue and expenditure as prescribed.[5] {from 2.6.10-27}
A wise Chancellor is one who collects revenue so as to increase income and reduce expenditure. He shall take remedial measures if income diminishes and expenditure increases. {2.6.28}

(ii) Administration of the Countryside:

(a) The Chancellor shall divide the countryside into four provinces and [clearly] demarcate the boundaries of each village. He shall appoint a Governor (*sthanika*) for each province and a Record Keeper (*gopa*) for every group of five to ten villages.

(b) [The Record Keepers (for the villages under their charge) and the Governors (for their provinces) shall maintain records as follows.] The villages shall be classified as best, average or lowest. They shall also be classified according to whether they are [tax paying or] tax-exempt, whether they supply soldiers [in lieu of tax], and whether they supply [fixed amounts of] grain, cattle, gold, forest produce, labour or other commodities. [Within each village,] every plot of land shall be numbered and its use recorded according to the classification: cultivated or fallow, dry or wet cultivation, park, vegetable garden [and orchard], enclosed area, forest, sanctuary, temple, water works, cremation ground, rest-house, public drinking-water facility, holy place, pasture and road. These records shall be used for determining the location of fields, forests and roads [in case of boundary disputes] and to record transactions such as gifts, sales, charitable endowments and tax exemptions. [Likewise,] each house shall be numbered and classified as whether tax-paying or tax-exempt. Records of the inhabitants shall also be kept under the following headings: (i) the *varna*; (ii) occupation (such as farmer, cowherd, trader, craftsman, labourer or slave); (iii) the number of males and females as well as the number of children and old people, their [family] history, occupation, income and expenditure; (iv) livestock and poultry owned; (v) the amount of

3

Diversion: A long(er) history

- The *Arthashastra* (Rangarajan 1987)

220   *The well-organized state*

(b) the preparation of the budget and maintenance of detailed accounts of revenue and expenditure as prescribed.[5]   {from 2.6.10-27}
A wise Chancellor is one who collects revenue so as to increase income and reduce expenditure. He shall take remedial measures if income diminishes and expenditure increases.   {2.6.28}

(ii) Administration of the Countryside:

(a) The Chancellor shall divide the countryside into four provinces and [clearly] demarcate the boundaries of each village. He shall appoint a Governor (*sthanika*) for each province and a Record Keeper (*gopa*) for every group of five to ten villages.

(b) [The Record Keepers (for the villages under their charge) and the Governors (for their provinces) shall maintain records as follows.] The villages shall be classified as best, average or lowest. They shall also be classified according to whether they are [tax paying or] tax-exempt, whether they supply soldiers [in lieu of tax], and whether they supply [fixed amounts of] grain, cattle, gold, forest produce, labour or other commodities. [Within each village,] every plot of land shall be numbered and its use recorded according to the classification: cultivated or fallow, dry or wet cultivation, park, vegetable garden [and orchard], enclosed area, forest, sanctuary, temple, water works, cremation ground, rest-house, public drinking-water facility, holy place, pasture and road. These records shall be used for determining the location of fields, forests and roads [in case of boundary disputes] and to record transactions such as gifts, sales, charitable endowments and tax exemptions. [Likewise,] each house shall be numbered and classified as whether tax-paying or tax-exempt. Records of the inhabitants shall also be kept under the following headings: (i) the *varna*; (ii) occupation (such as farmer, cowherd, trader, craftsman, labourer or slave); (iii) the number of males and females as well as the number of children and old people, their [family] history, occupation, income and expenditure; (iv) livestock and poultry owned; (v) the amount of

- Nainsi's Census (Peabody 2001)

TABLE 1
Munhata Nainsi's Household List for Merta, A.D. 1673

2512 Mahajan jatis					
2158	Banyas				
1371	Oswal	161	Agarwal	75	Kandelwal
551	Maheshri				
354	Various other jatis included [with Mahajans]				
282	Bhojag	28	Bhat	4	Niratkari
40	Khatri				
669 Brahmans, all kinds (*nyat*)					
82	Pokarana	54	Dahma	14	Gujarati
13	Rajpur	44	Sarasut	50	Gaud
42	Gujar Gaud	7	Sankwal Upadhiya	7	Sanavas
140	Parik	105	Sirimali		
53 There are fifty-three houses of Kayasthas					
42	Bisa	7	Dasa	4	Bhatnagar
7 Todarmals					
190 Rajputs					
292 Sipais					
31	Pathan	126	Deswali	11	Kazir
124	Turk Sinbandhi Topchi				
8 Red Ochre Islamic physicians (*geru tabab*), descendants of Isar					
2124 Purifying (*pavan*) jatis					
114	Darzi	293	Julaha	33	Ghosi
91	Sonar	44	Kasabgar	167	Mochi
97	Nai	17	Kunjda	30	Jatiya Vangar
10	Nilgar	2	Chitara	60	Bhedbhunja
12	Sikalgar	1	Bajigar	6	Tirgar
10	Kholwar	4	Babar Nai	11	Lakhara
27	Kasara	45	Khatik	57	Pinjara
	Thanthara				
54	Khati	26	Jatiya Andhodi Range	10	Ghanchi
23	Tanboli	21	Nagar Nayak	21	Sogagar
20	Sabangar	7	Acharaj Khapan Khosa	6	Bhardhari
34	Kumbhar	182	Mali	105	Multani
18	Gancha	91	Sonar	2	Tabab
35	Bajdar	62	Teli	50	Dakot
11	Bharava	31	Kalal	8	Hamal
38	Silavat	51	Chhinpa	4	Bhadihar
111	Dhobi	81	Kahar	1	Sarago

3

Some basics

- The balancing equation (For India, 2022 - 2023)

    - Population (2023) = Population (2022) + Births (2022) - Deaths (2022) + Net Migration (2022)

    - Change = Births - Deaths + Net Migration

- The United Nations World Population Prospects does provide these estimates for India (and for the rest of the world)

    - Gerland, P. (2014). UN Population Division's methodology in preparing base population for projections: case study for India. *Asian Population Studies*, *10*(3), 274-303.

    - Alternative estimates: Wittgenstein Centre Human Capital, Institute for Health Metrics and Evaluation

Mortality 101 (with an example from the SRS)

Crude death rates, 2018:

- UP: 6.6 deaths per 1,000
- Kerala: 6.9 deaths per 1,000

## Mortality 101 (with an example from the SRS)

Crude death rates, 2018:

- UP: 6.6 deaths per 1,000
- Kerala: 6.9 deaths per 1,000
- It turns out that Uttar Pradesh
  is much younger than Kerala

# Mortality 101 (with an example from the SRS)

Crude death rates, 2018:

- UP: 6.6 deaths per 1,000
- Kerala: 6.9 deaths per 1,000
- It turns out that Uttar Pradesh is much younger than Kerala
- And looking age-specific mortality helps avoid confusion
- We can age-standardize crude death rates. Or we can look at life expectancy.

Higher mortality in Uttar Pradesh relative to Kerala across the life course
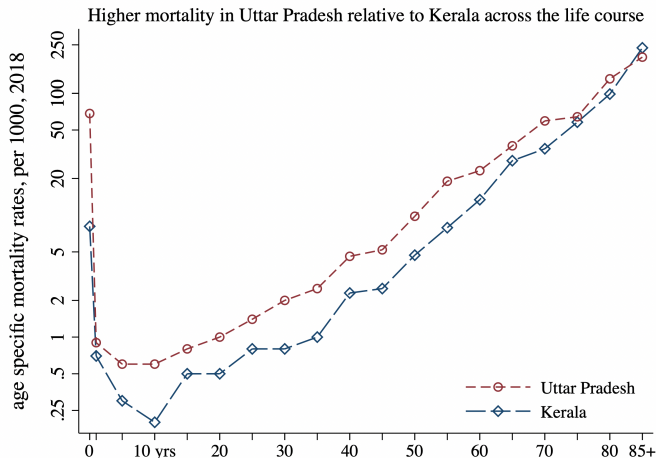
(Period) Life Expectancy at Birth

- The average number of years a (hypothetical) cohort would live
  - if they experienced the set of prevailing age-specific mortality rates
  - in a population (in a given period)
- Life expectancy at birth, 2014–2018, SRS
  - Uttar Pradesh: 65.3 years
  - Kerala: 75.3 years

"Something's just not right – our air is clean, our water is pure, we get plenty of exercise, everything we eat is organic and free range, and yet nobody lives past 30"

© ALEX GREGORY/THE NEW YORKER/THE CARTOON BANK
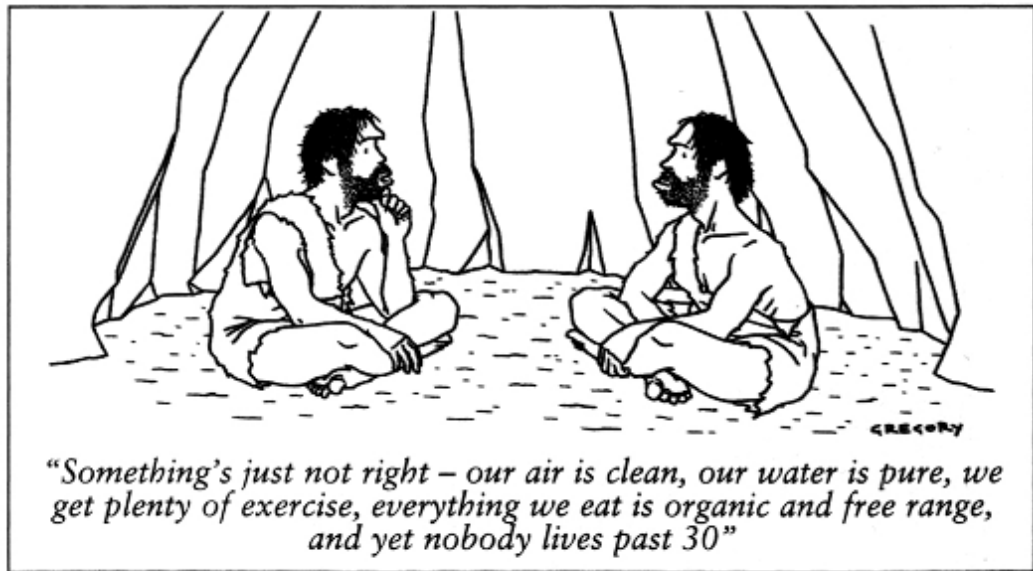
And similarly for fertility

- The total fertility rate
  - "The number of children a woman would have if she lived from age 15 to age 50 and experienced the age-specific fertility rates of the period in question"
 - Uttar Pradesh (2018): 2.9
  - Kerala (2018): 1.7

The SRS

- The Sample Registration Survey monitors vital events in a nationally representative panel of villages and urban blocks ($> 8{,}800$)

  - Set up in late 1970s, reports available from then on
  - Panel adjusted after censuses (1977-78, 1983-85, 1993-05, 2004, 2014)
  - Exceptionally large sample (> 8.3 million individuals)
  - A "dual record" system, with continuous monitoring
  - Annual reports available with some delay, but not underlying microdata
  - Some longstanding issues (infant mortality in Bihar, for instance) and some new ones (before the pandemic)

- Available data: Crude birth rates, crude death rates, age-specific mortality rates, maternal mortality, age-specific fertility rates, sex ratio of birth

But what if we just registered all births and deaths?

Birth registration completion (2014-21): 89%
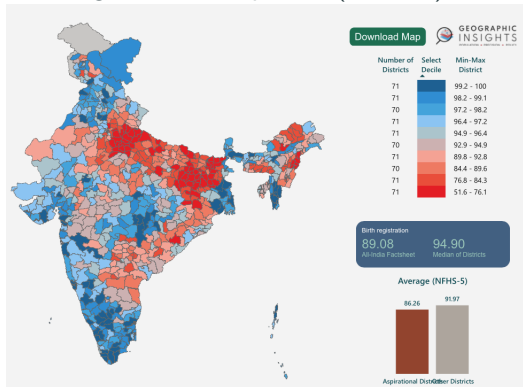
But what if we just registered all births and deaths?

Birth registration completion (2014-21): 89%
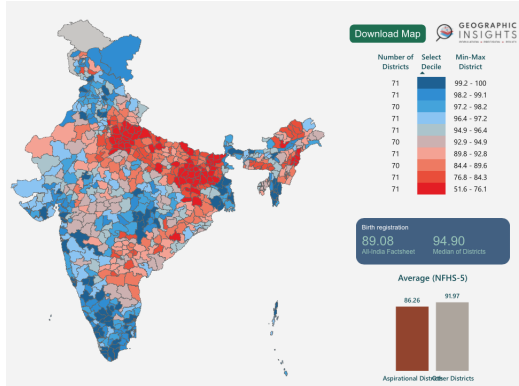
Death registration completion (2017-21): 71%
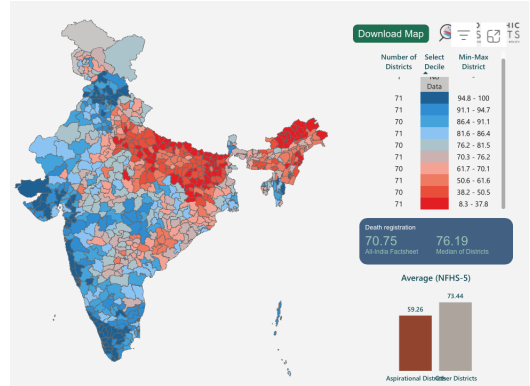
Civil Registration System

- Incomplete, but improving

- Good quality in several states

- Microdata not available

- Lacks denominators

- But annual summary reports published with some delay

- Worth trying to ask for data politely (and beg, borrow, or …)

Age-specific mortality rates in Kerala, 2006-2018 (Gupta & Mani 2021)

Reported deaths in MP shot up in April and May 2021

■ 2018 ■ 2019 ■ 2020 ■ 2021
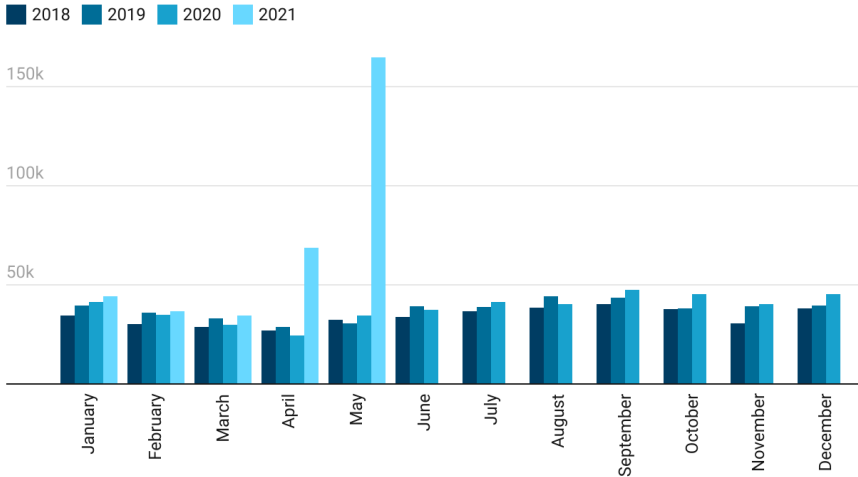
Source: Civil Registration System • Created with Datawrapper

Deaths registered in Madhya Pradesh (Rukmini S 2021)

But beware of pitfalls



(b) Excess mortality P-scores, WHO estimates for India
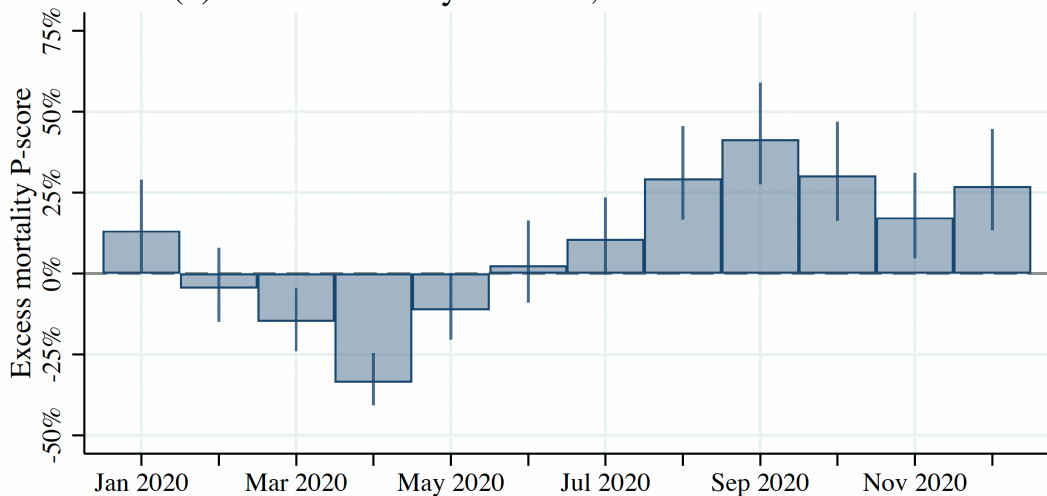
Msemburi et al 2023

So is there any high-quality data that is also publicly available?

- The National Family Health Surveys (NFHS 1-5)
  - India's version of Demographic and Health Surveys (> 400 surveys, > 90 countries)
  - Have exceptionally large sample sizes: $> 600{,}000$ households in NFHS-4/5
  - Ask "birth history" from women in reproductive ages (15-49)
    - All births they ever had
    - Month and year of birth
    - Survival status of child
    - Age at death if child died
    - Month and year of death if child died
  - Child mortality and fertility rates for last couple of years more widely used
  - Also includes a wealth of other information, measured well for the most part

Figure 1: India has low fertility despite a high proportion of married women
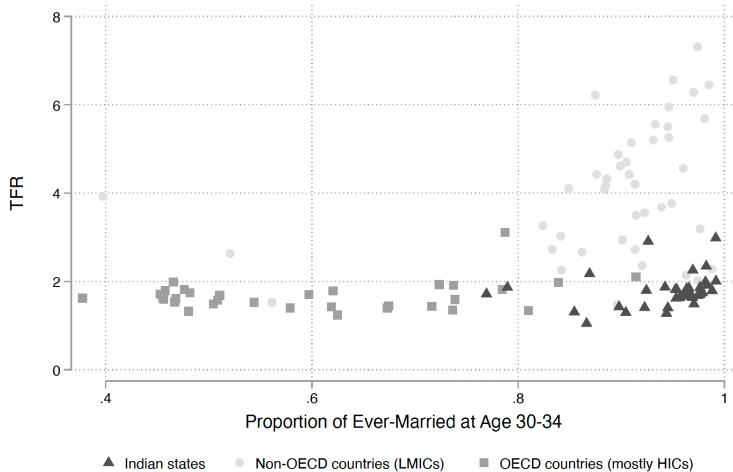
A closer look at TFR in Indian states

Information on all-cause mortality in the NFHS surveys

- Question on recent deaths of usual household members
  - e.g. since January 2017 in NFHS-5 (2019-21)
  - Follow-up questions include
    - age at death,
    - sex of deceased individual,
    - month and year of death
    - whether death was during pregnancy or within two months of pregnancy
    - whether death was due to external causes?
- Can be used to generate age-specific mortality rates
  - When combined with information on alive individuals
  - Which are observed in the household roster of the survey
  - Code: https://github.com/lekhajokha/e0_social_disparities_nfhs

Reliable mortality rates

a) Female, 1997-2001

b) Male, 1997-2001

c) Female, 2012-2016

d) Male, 2012-2016

e_0 NFHS (1997-2000) -- 61.3 years [95% CI 60.7-62.0]
e_0 SRS (1997-2001) -- 63.3 years

e_0 NFHS (1997-2000) -- 60.6 years [95% CI: 60.0-61.1]
e_0 SRS (1997-2001) -- 61.4 years

e_0 NFHS (2013-2016) -- 69.6 years [95% CI: 69.3-69.9]
e_0 SRS (2012-2016) -- 70.2 years

e_0 NFHS (2013-2016) -- 66.0 years [95% CI: 65.7-66.3]
e_0 SRS (2012-2016) -- 67.4 years

Mortality rates per 1,000 ($_n m_x$)

SRS — NFHS
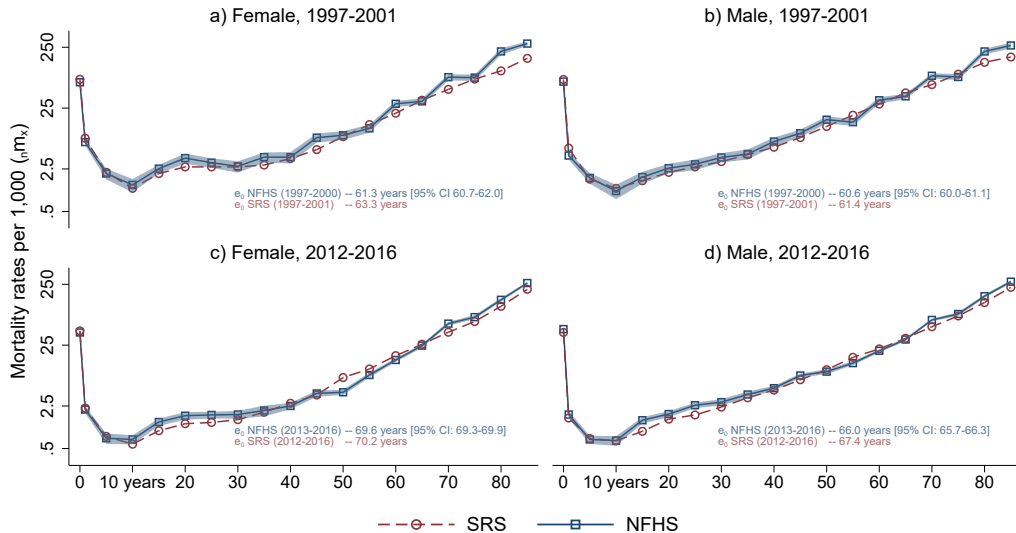
Reveals uncertainty about maternal mortality levels

Figure 1. *Comparison of maternal mortality ratios by region and data source*

Note: SRS estimates are reported rates from SRS Maternal Mortality Bulletins. NFHS estimates and 95% CIs account for NFHS survey weights and clustering of observations within primary sampling units. UN/WHO 2014 ratio and uncertainty interval are calculated using the BMat model by the UN MMEIG.

Franz et al 2022

Can we learn something about Indian society from these data?

Large and persistent disparities in life expectancy at birth in India
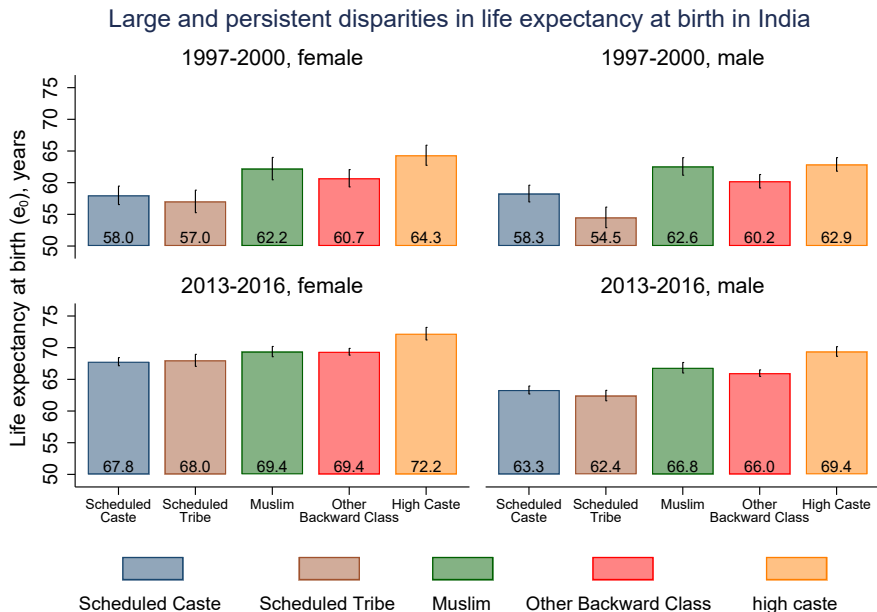
And what about the pandemic?

- NFHS-5 conducted between 2019 and 2021

- First phase in 2019 (22 states and union territories out of 36)

- Second Phase disrupted by the 2020 Indian lock-down

  - Completed in early 2021, before the delta variant surge

  - 14 States/UTs

  - Not a random-sample, but similar in characteristics

  - Representative of a fourth of India's population

  - 23% of >600,000 NHFS-5 interviews conducted in early 2021
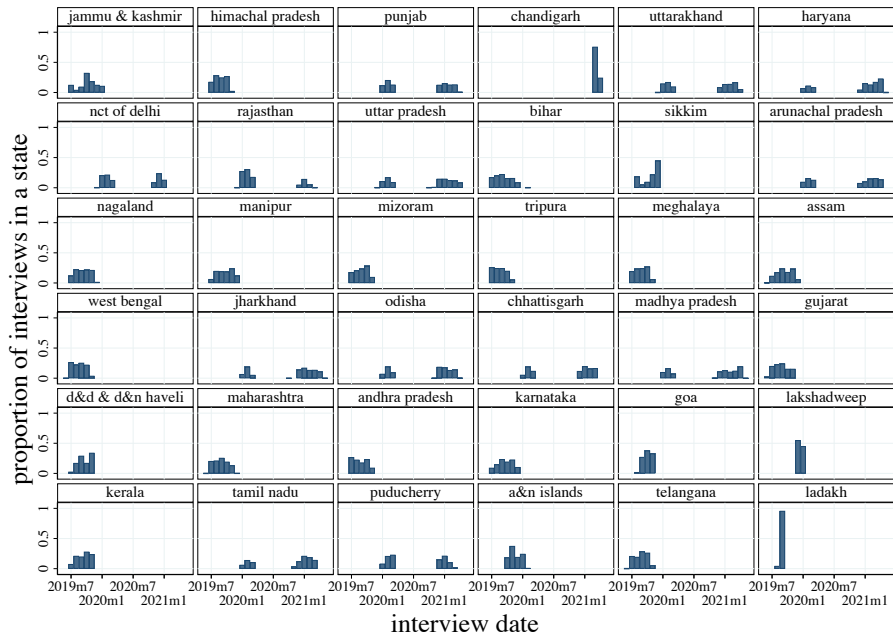
# Excess mortality in 2020 in NFHS-5 2021 subsample

(a) Percentage increase in 2020 crude death rate relative to previous years
(excess mortality P-scores)

- NFHS-5 2021 Subsample, relative to 2019 baseline
- NFHS-5 2021 Subsample, relative to 2018-19 baseline
- WHO Estimates (Msemburi et al 2023) for India

(b) Recorded COVID-19 cases in NFHS-5 2021 sub-sample states

(c) Recorded COVID-19 deaths in NFHS-5 2021 sub-sample states

Gupta et al 2023
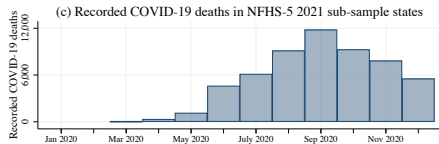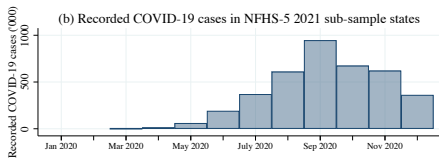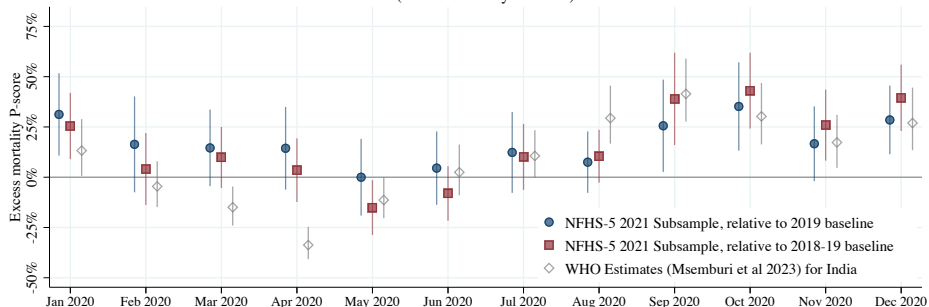
Always helpful to compare data sources
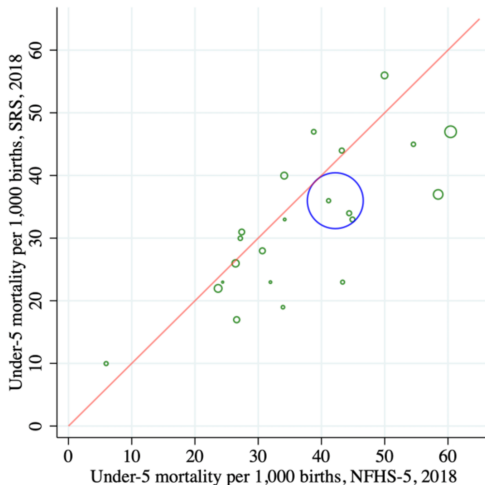
Banaji et al 2022

Other useful surveys

- The "Annual" Health Survey 2007-2012
  - For 9 poor Indian states
  - Publicly available at DSDR (Coffey et al 2022)
  - Helped show that social disparities in India cannot be explained by economic disparities (Vyas et al 2023)
- The India Human Development Survey
  - A panel survey with two waves (2004-05 and 2011-12) already available
  - Includes similar birth history data as NFHS-5
  - Mortality information in "tracking" file of the panel
  - Helped show that socioeconomic status is associated with adult mortality (Barik et al 2017)
  - Third wave currently in field
- Health and Demographic Surveillance Sites

Other surveys

- The Longitudinal Aging Study of India
    - Includes questions on recent deaths of household members
    - And a large number of measured bio-markers (2018-19)
    - Also a panel of individuals older than 45, second wave currently in field
- The health rounds of the National Sample Surveys
    - Also include questions on recent deaths of household members
- Preliminary investigations reveal that mortality rates from module on recent deaths of household members in LASI and NSS are not reliable
- Consumer Pyramids Health Survey
    - Biased sample that does not measure mortality well

Other administrative sources

- Health Management Information System
  - Primarily based on records of public health facilities
  - Also used to understand pandemic mortality
  - Captures births better than it captures deaths

Other administrative sources

- Health Management Information System
  - Primarily based on records of public health facilities
  - Also used to understand pandemic mortality
  - Captures births better than it captures deaths

Table 1
Comparison of birth and mortality indicators in 2014 and 2018 at the national level, based on the Health Management Information System (HMIS), the National Family Health Survey (NFHS) and the Sample Registration Survey (SRS).

Indicator	2014			2018		
	HMIS	NFHS	SRS	HMIS	NFHS	SRS
Birth Rate	16.8	NA	21.0	15.9	NA	20.0
Neo-Natal Mortality Rate	4.4	31.5	26.0	8.6	NA	23.0
Post Neo-Natal Mortality Rate	1.8	8.6	13.0	3.3	NA	9.0
Infant Mortality Rate	6.1	40.1	39.0	11.9	NA	32.0
Child Mortality Rate ages 0-4	1.2	9.6	11.0	2.4	NA	9.0
Child Mortality Rate ages 1-4	0.2	2.0	1.7	0.4	NA	1.1
Proportion of Institutional Births (%)	68*	81	79	94	NA	83

Chatterjee et al 2022

Other administrative sources

- Integrated disease surveillance program
- COVID-19 cases and deaths (ICMR + states)
  - Parallel to IDSP
  - "It was the best data. It was the worst data."
- Medically certified causes of death
  - Parallel to CRS
  - Only about one in five deaths have a cause of death certificate

Crisis sources

- Crematoria

- Newspaper reports

  - Lockdown related deaths (Aman et al 2020)
  - Deaths due to lack of oxygen (Priya et al 2021)
  - Comparison of published obituaries in Kerala with confirmed COVID-19 deaths (Madhavan 2020)

- Insurance records

The effects of India's COVID-19 lockdown on critical non-COVID health care and outcomes: Evidence from dialysis patients

Radhika Jain [a,*], Pascaline Dupas [b]

Some other helpful resources

- District level population projections, 2000-2020:
  https://spatialdata.dhsprogram.com/population-estimates/

- DHS STATcompiler and GitHub repository

- Spatial population maps (population, births, age): worldpop.org

Neglected themes (apart from data quality)

- Causes of death
  - Collapse of nationally representative research on "verbal autopsy" methods
- Still-births
  - little reliable
    measurement
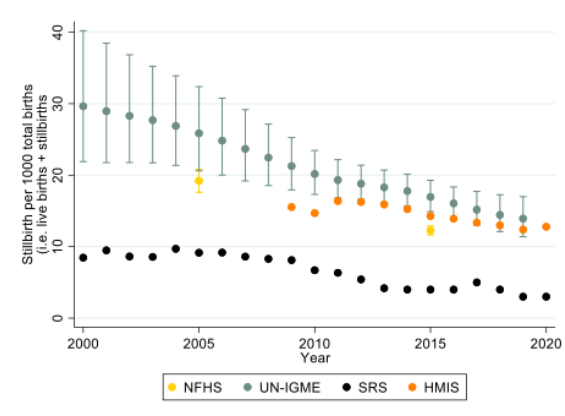
Hathi et al (in progress)

Some tips

- Important in itself; but also as an outcome and a determinant
  - Neglected theme in social sciences
  - And in public discussions
- Both discrepancies and consistencies between sources can be revealing
  - e.g. official COVID-19 deaths v excess death estimates
- Possible to learn even from biased sources
  - Specially if we are careful and acknowledge biases
- Worth paying attention to selection into data and measurement errors
  - Especially in an unequal society like India