An Investigation into the Quality of HMIS Data

Nathan Franz

3/23/2020

The Question

Email 1

> The last NFHS was in 2015. In the decade before, hospital birth increased dramatically from 40% to 80%. Has it further increased since? The only data source that *might* help answer this is health monitoring system (HMIS) data from the government. But we don't know how reliable it is. If someone cleans and organizes the HMIS data, it could be matched to NFHS-4 data to see whether it reliably reported the proportion of hospital births in different places in 2015. (It probably did not reliably report deaths.) If it did, then let's see what it says about the proportion of births that took place in hospitals in 2019 and see if we trust it.

Email 2

> This is substantially going to be a task of assessing data quality – do these data make any sense? Do some indicators or states make sense but not others? Whatever you end up learning about it will be useful because I have other colleagues who work on health in India who are interested in it to.
>
> I'm also interested in a very specific question: do the numbers on place of delivery (proportion of babies born at home, in public hospitals, and private hospitals) in the HMIS match a high quality sample survey (the NFHS-4)?
>
> To get started, I shared a google drive folder that contains HMIS (health monitoring information system) data on maternal and child health. Here are some steps.
>
> Download 2014-2015 data.
>
> First figure out whether it seems to be cumulative throughout the year or quarterly data? One thing that is annoying is that there are a lot of repeat file names which might make working with these spreadsheets more difficult.
>
> Once you'd figured the data out a bit, compute the proportion of births it claims occurred in private, public, and at home in each state for 2014-15.
>
> Then download and open the NFHS-4. Look at place of birth (using weights of course) for 2014-15 births. You'll have to code the place of birth variable yourself based on a wider set of categories, but it should not be too hard.
>
> Compare the two data sources. Does the HMIS match the NFHS? (could you make a graph or a table to show it?)
>
> If so, could you please look at what fraction of births happen at home, private, and public in each state in 2018-2019 in the HMIS?
>
> If, along the way, you find other interesting things, please let me know. (One thing is that I think the HMIS basically does not report neonatal mortality.)

HMIS Data Characteristics

The HMIS data is provided in two explicit levels of aggregation: at the district level and at the sub-district level. For the purpose of investigating the quality of HMIS data compared to NFHS-4 data, we will use the district-level data.

For the year 2014-2015, the data is provided **cumulatively** each quarter. That is, data is available for the periods

- from April 2014 to June 2014 – the first quarter of Indian fiscal year (FY) 2014-2015 – in the directory "MonthUpToJune",
- from April 2014 to September 2014 – the first two quarters of FY 2014-2015 – in the directory "MonthUpToSeptember",
- from April 2014 to December 2014 – the first three quarters of FY 2014-2015 – in the directory "MonthUpToDecember", and
- from April 2014 to March 2015 – all of FY 2014-2015 – in the directory "MonthUpToMarch".

Within each of the quarterly directories, there are yet more directories for each of the Indian states containing district- or sub-district-level data. There is also a directory with data aggregated to the state level, which is what we will use for comparison to NFHS-4 data. The file from which we'll extract the data has the following address: "./nrhm_files/district/2014-2015/MonthUpToMarch/_ALL INDIA/All.xls"

The HMIS reports 166 indicators. The indicator columns are color-coded: white for counts and gray for percentages. Those we care about are called "Number of Home deliveries", "Deliveries Conducted at Public Institutions", "Institutional deliveries (Public Insts.+Pvt. Insts.)", and "Total reported deliveries".

State	% Home	% Pub. Inst.	% Pvt. Inst.
A & N Islands	3.72748	96.27252	0
Andhra Pradesh	5.211907	52.16508	42.62301
Arunachal Pradesh	6.357544	80.19682	13.44563
Assam	15.3392	73.79186	10.86894
Bihar	23.64617	76.34688	0.0069494
Chandigarh	1.447695	98.55231	0
Chhattisgarh	25.80345	57.88689	16.30967
Dadra & Nagar Haveli	1.251977	83.80338	14.94465
Daman & Diu	1.574173	53.6286	44.79723
Delhi	6.288519	77.64141	16.07007
Goa	.1157025	63.6584	36.22589
Gujarat	2.416069	40.11806	57.46587
Haryana	11.37654	53.37545	35.24801
Himachal Pradesh	17.75469	67.95181	14.2935
Jammu & Kashmir	9.073519	84.97659	5.949898
Jharkhand	21.44771	60.38848	18.16381
Karnataka	.9186705	67.01209	32.06924
Kerala	.1705298	30.97543	68.85404
Lakshadweep	0	100	0
Madhya Pradesh	11.51237	80.6953	7.792329
Maharashtra	1.697953	53.04405	45.258
Manipur	20.08797	64.88353	15.0285
Meghalaya	48.5994	39.85056	11.55005
Mizoram	9.637074	72.86702	17.49591
Nagaland	22.89588	64.5219	12.58222
Odisha	11.43811	79.14854	9.413345
Puducherry	.0136509	73.18727	26.79908
Punjab	9.951732	48.89772	41.15054
Rajasthan	5.091767	71.39291	23.51532

State	% Home	% Pub. Inst.	% Pvt. Inst.
Sikkim	2.047952	77.46004	20.49201
Tamil Nadu	.0736521	67.82365	32.10269
Telangana	3.084601	60.66509	36.25031
Tripura	13.15953	81.30605	5.534418
Uttar Pradesh	26.78585	65.24159	7.972555
Uttarakhand	21.66616	62.22997	16.10387
West Bengal	17.51872	63.65038	18.83089

NFHS-4 Data Characteristics

One interesting thing: I've never worked with survey data before, but I would've expected the per-state means of the state sample weights to equal 1 (or 10^6, since the state sample weight variable contains six decimal places without a decimal point). Instead, the means range from 92% to 107%.

State	Mean of state sample weight	Mean st. sample wt. for dlv. plc. observed
A & N Islands	1.019731	.9401651
Andhra Pradesh	1.000626	.9806303
Arunachal Pradesh	1.003405	1.014932
Assam	.9950057	.9870515
Bihar	1.002083	1.003952
Chandigarh	.9950939	.984368
Chhattisgarh	1.022396	1.009044
Dadra and Nagar Haveli	1.000107	1.005672
Daman and Diu	1.009258	1.044274
goa	1.009291	.9981192
gujarat	.979265	.9724191
haryana	1.002750	1.002528
himachal pradesh	.9854493	.9825162
jammu and kashmir	.9590977	.9334011
jharkhand	1.004140	1.006503
karnataka	.9728785	.9999459
kerala	1.001710	.9967074
lakshadweep	1.024807	.975109
madhya pradesh	.9944348	.9939861
maharashtra	.9635219	.9775645
manipur	.9604686	.9582052
meghalaya	.988156	.9870083
mizoram	.9231265	.9071155
nagaland	.966884	.985118
delhi	1.066457	1.068031
odisha	.9944992	.9954692
puducherry	.9929863	.9707636
punjab	.9953682	1.007983
rajasthan	.9995916	.9988474
sikkim	.9858245	.9894643
tamil nadu	.9901204	.9917371
tripura	.9770556	.9687295
uttar pradesh	.9939065	.9918465
uttarakhand	.9851471	1.005475
west bengal	.9782621	.9866215

State	Mean of state sample weight	Mean st. sample wt. for dlv. plc. observed
telangana	.9780245	.9960077

The NFHS-4 reports 1,315,617 observations of 1,340 variables. Those we care about are called “sv005” for state sample weight, “m15” for place of delivery, and “v024” for state.

I had to create

state	% home	% other	% private	% public
andaman and nicobar i	3.09	0.32	4.33	92.27
andhra pradesh	8.34	0.11	53.27	38.28
arunachal pradesh	47.34	0.33	9.61	42.72
assam	29.22	0.15	10.65	59.99
bihar	35.93	0.23	16.18	47.66
chandigarh	8.36	0.00	19.27	72.38
chhattisgarh	29.70	0.07	14.36	55.87
dadra and nagar havel	12.05	0.00	21.55	66.40
daman and diu	9.90	0.00	47.09	43.01
goa	3.13	0.00	38.66	58.21
gujarat	11.11	0.21	56.05	32.63
haryana	19.48	0.08	28.44	52.00
himachal pradesh	23.05	0.51	14.81	61.62
jammu and kashmir	13.73	0.61	7.54	78.12
jharkhand	37.76	0.30	20.13	41.81
karnataka	5.58	0.15	32.90	61.37
kerala	0.11	0.00	61.54	38.35
lakshadweep	0.72	0.00	35.00	64.28
madhya pradesh	18.97	0.21	11.37	69.46
maharashtra	9.60	0.11	41.38	48.91
manipur	30.54	0.35	23.43	45.67
meghalaya	48.20	0.40	11.91	39.49
mizoram	19.86	0.34	16.01	63.78
nagaland	67.05	0.13	7.67	25.14
delhi	15.47	0.05	28.88	55.60
odisha	14.14	0.44	9.52	75.91
puducherry	0.11	0.00	17.94	81.95
punjab	9.50	0.01	38.84	51.65
rajasthan	15.83	0.20	20.48	63.49
sikkim	5.29	0.00	11.98	82.72
tamil nadu	0.97	0.09	32.27	66.67
tripura	20.00	0.11	10.82	69.07
uttar pradesh	31.82	0.34	23.34	44.49
uttarakhand	31.10	0.26	24.87	43.77
west bengal	24.22	0.58	18.62	56.57
telangana	7.84	0.45	61.09	30.62