Health Economics

Dr Pratap C Mohanty

Department of Humanities and Social Sciences,

Indian Institute of Technology Roorkee

Week – 11

Lecture 53- Data in Health: Sources and Indicators- I

Welcome friends to our NPTEL MOOC module on Health Economics. This is our eleventh week, and we have targeted discussing health issues in terms of their data, what different sources of data are available, what the indicators are, and how one should search and include them in their work as a researcher. So, we will discuss. In the first two lectures, we will emphasise the micro and macro sources and indicators of data. Then, in the third, we will attempt to discuss the basic analysis using basic models. Then, the latter two will emphasise the software and the packages, where we can answer how it is available, etc.

So, here we will emphasise the sources of data. Before that, let me recapitulate what we did in our previous week. We discussed the concept of efficiency, and we also discussed different aspects of data envelopment analysis. This lecture briefly introduces you to data in healthcare, its sources, indicators, and challenges in data collection. Starting with the background, health economists usually face questions that require the correlations between variables under examination for analysis.

So, the obvious aspect is how to delve into the economic relationship, specifying the models/ econometrics model, which requires data. Sequence is like understanding the investigation of economic relationships that requires specification of the economic model, and then further, it requires understanding of data. All these helps in the specification of the appropriate model and choice of estimation method. So, much of the work of health economists relies on the type of data available, which is usually divided into observational study data and administrative data. So, in this lecture we will be emphasising on observational studies data, that largely explains the micro-foundations or the micro information at a very desegregated settings, from the perspective of individuals. However, these can be used for macro analysis when employed at an aggregate level. I mean, collating the micro will generate the macro aggregates. Observational data are often designed to provide data on a representative sample of the underlying population. These may include both quantitative and qualitative data. The administrative data includes micro and macro aspects as well. However, these are not designed primarily for research purposes and may include only quantitative data.

Understanding the observational studies data: There are broadly two types of observational studies. One is at the primary level, and the other is secondary. Again, primary is of two types: qualitative and quantitative and secondary is just quantitative.

So far as qualitative primary data is concerned since we said these are observational, these can be collected through interviews, FGDs (Focus group discussions), observation case studies etc. Quantitative data is covered through surveys (collecting objective and subjective measures) or experiments. Whereas, in case of secondary (that is quantitative), we usually base it on the large scale unit level surveys, which research institutions usually carry out with the help of the government. So, let us elaborate on each of these a little more.

We will further emphasise what is called qualitative. As I already discussed, these include interviews within the qualitative framework, and these are mostly unstructured, which also allows flexibility. However, they can also be semi-structured or structured based on research needs. Subjective health assessment, exploring patient experience, etc., are some examples of qualitative information from the interviews.

The second one is 'focus group discussion', which consists of the focused persons, especially the stakeholders who are bearing the context upon which the study is based. This is usually considered a limited set of groups and used to be catering to their locality or their surroundings. They must be in a very closed, compact group of interchanging ideas and discussions. It is a group discussion where a facilitator is always needed, and the facilitator leads a discussion on a specific topic followed by participants sharing their experiences. For example, doctor's perception of a newly introduced health service should be discussed through a focus group discussion approach.

Another is called 'ethnographic observation'. This involves observing and recording an individual or group's behaviour or interactions in a particular setting. For example, exploring how different tribes use traditional medicines to improve their health issues. Another type of qualitative micro or observational data is 'case studies'. This involves an in-depth analysis of a particular individual, group or organisation's phenomena, usually over an extended period. For example, analysing COVID-19's impact on a particular group can be considered to be a case study.

So, next is the quantitative but primary one/ primary data. First, through surveys. In quantitative research, surveys are used to collect data from individuals' samples using a standardised questionnaire method. Roberts et al. (2019), used surveys to collect data on quality-of-care delivery and the uptake of evidence-based practices. Note that many health surveys now include objective measurement of health as part of clinical research assessment along with the subjective health assessment, which alone might give inaccurate results.

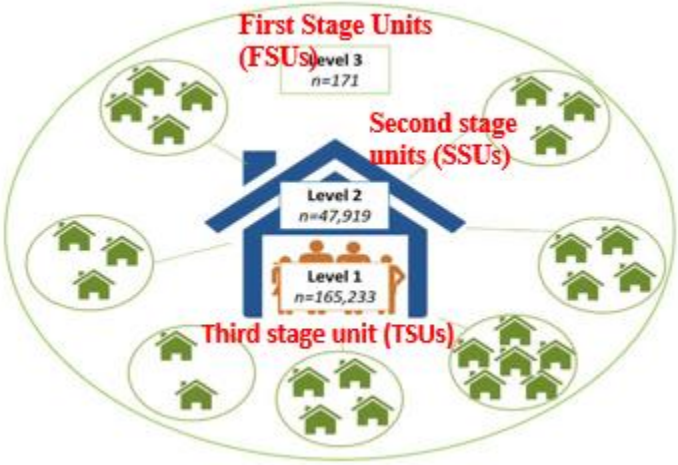
Another approach of quantitative information is called 'experiments'. In health economics, experimental research is used to evaluate the effectiveness of an intervention. A clinical trial is a type of experiment that tests the safety and efficacy of a new drug or medical device. Another example can be the randomised controlled trial, which is very popular recently, which we have also referred to in our second week on RAND experiment.

One thing to note here is that as RCTs (randomised controlled trials) are very expensive, time-consuming, and are not always practical or ethical. Economists sought quasi-or natural

experiments that exploit variation in any variable (across time and space) to analyse the relations. For example, in 2015, Grossman analysed the causal impact of change in education on health.

Secondary data, as I already mentioned, is usually the large-scale unit level data sets, collected by academicians, research institutions, or the government of India, with all possible aspects including its ethical clearance etc. So, unit level data refers to detailed data concerning sampled units at the ultimate stage, along with the sampling weight for that stage.

We also discussed this aspect in my lecture on handling large-scale data. If you search in YouTube, you will also find these things on handling large-scale survey data, especially the emphasis on weight. Weight will represent the state or district, so that the findings from the data you can claim represent the entire area. It is also called microdata, or disaggregated data or even raw data. It depends on where you are actually working, and I will discuss it. Usually this involves first stage, second stage and third stage:



So, usually the final stage is considered to be the individuals or the households. How households are actually considered as the surveying unit in NSS data. So, TSU, SSU, FSU, etc., are the words used in the survey data, which we discussed in this module.

There are advantages of unit level data. As I just said, these are nationally representative, but it also depends upon the survey. But nowadays, all the survey conducted by government of India or the bigger institutions usually represent the data at national, state, and district levels. So, accordingly, different weights are given. Sometimes you might require I-weight, sometimes P-weight, sometimes A-weight (analytical weight), etc. We discussed this in another module on 'survey data on healthcare'. So, I am not discussing much here, because our focus is on understanding health economics concepts.

Let us understand the advantages of unit-level data in more detail. These provide information about the characteristics of individuals or entities such as households, business enterprises, facilities, farms, or even geographical areas such as villages or towns. Studying relationships

and interactions among phenomena allows an in-depth understanding of socio-economic issues.

This is key to design projects and formulating policies. Targeting interventions and monitoring and measuring the impact is becoming very helpful from these datasets. Now, I am citing and referring you directly the sources of the unit-level datasets in India. It shows how you should work on it. One by one we are mentioning. This is a national sample survey

https://microdata.gov.in/nada43/index.php/catalog/152

This is national family health survey (NFHS)-

https://www.dhsprogram.com/data/dataset/India_Standard-DHS_2020.cfm?flag=0

So, NSS is here, then this is consumer pyramid:

https://consumerpyramidsdx.cmie.com/

This is LASI (Longitudinal Aging Survey of India):

https://www.iipsindia.ac.in/content/LASI-data

This is India's Human Development Survey:

This is SAGE database on healthcare:

And this is Young Life Survey (especially useful for cohort etc.):

We will emphasise now. So, the first aspect is understanding the richness of the data where you get the in-depth. So far as the coverage is concerned, I think the NSS (National Sample Survey) data and NFHS are considered the best, even globally. It is one of the advantages of the huge population in this country, and that too, the government of India had a long plan to cover a large scale survey. These two, especially NSS, are considered to be among the best in the world, especially in terms of design.

So far as NFHS is concerned, it has already covered 5 rounds. So, 5 rounds have already been covered. Starting in 1992, the latest round was the NFHS-5. If you click on this website, you will be redirected to the site. We have given all the links. I will just tell you, let me explain a bit. Otherwise, let me open this first. National Family Health Survey, from the word itself, you can understand that the focus is on family and their health, and the major focus is on women issues, women and child. The questionnaire, the latest round that was covered, is 2019 till 21. This is the latest round. This is called NFHS-5.

If you click on this link, all the databases are available. You need to register with your college identity card. So, it will give you the data. The portal is DHS datasets. If you click on this, if I just click on this, it will direct me to that site. I am just going to open it, and let you know how it works.

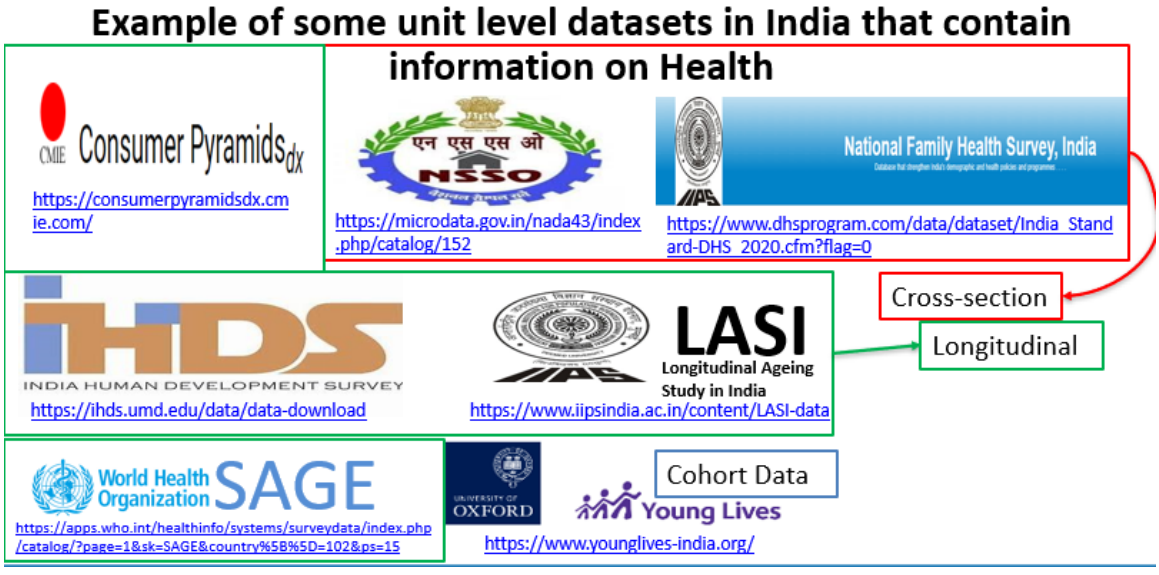
Here is the datasets, and datasets are in fact available in different categories (just see once). Actually, in NFHS, if you read that correctly, it is your individual record data. Individual record is the word they have written only to explain you or help you in the dataset in terms of its categories. These are systematically presented, but that does not mean they have actually diluted the raw data.

You can get the raw data directly and cross-check whether it is right or not. You can get the data directly in Stata format, SAS format, or SPSS format. So, individual record refers to the women profile, then children record etc., all datasets are available here. This refers to the latest 2019-21 dataset. Other dataset you can just click here. You need to first register, and then filter of country etc., it will be visible on your page, on your desk. So, open registration, and all the links will be activated, and you can download them. So, this is one.

The second aspect I need to emphasise is the NSS. NSS, as I already mentioned, we have dedicated modules of my own. There are two modules which are successfully running in the NPTEL platform. One is on 'handling large scale data', and another one is on 'survey data on healthcare'.

So, anyway, my focus was on health. So, I did not record any other, but however you can get all the sources over here. So, you just click over here, it is directed as, the latest round of dataset on health in the 75th round of national sample survey, conducted in the year 2017 to 2018, for over one year and all those things are important. You completely need to download the report, then, it is data layout, then, it is readme file, then, other schedules. Schedule 25.0 is important. The schedule means the questionnaire. Then, you need to get the raw data from here (in different formats). You have to click here to get micro data. So, it asks for a registration and will then be directed to the page to download the data.

However, there are other datasets also, which NSS provides: on agriculture, on debt and investment, then, on labor issues (employment unemployment round), or latest one is PLFS (Periodic labor force survey), then it also provides information about time-use survey (TUS) (So, time use survey). Besides this, you will also get consumer expenditure rounds as well. These rounds are very interesting and relevant for the researchers to publish their papers in journals and newspapers (wherever you want). I think further guidance are available in our module. In YouTube, if you search with large-scale data, NSS etc., it will be directed to our page. So, I am not focusing on this. These are purely cross-sectional datasets.



Others are important, I am just highlighting the green border that emphasises the longitudinal nature of the data. Longitudinal means, the data is provided or surveyed over different time periods with the repetition of the households, and these are also called panel data. Another interesting aspect of the survey, somewhere you can also track, like the young life survey, is panel data, specifically called cohort data. The cohort group is actually emphasises other rounds, like age group of a set of indivuduals is surveys again after five years. The other five years of the same individuals are covered in another round. The same cohort, like today right now suppose I am around 40. So, if the survey is conducted after two years, I would have been 42. So, the person with me (the age group) is covered as 42. So, I am also included there, but the cohort is emphasised. So, there have been five cohort series so far in the young lives. Hence, it is called a cohort. It is also called a panel but specifically called cohort data.

And coming to the interesting data through IHDS. It emphasises on human developmental issues such as education, health, income etc. So, also this track gives information about crime records, education, and the children's cognitive level, but one of the demerits is that the latest round was in 2011-12. So, this is relatively older. However, since this represents the area, the district, the state, and India in particular, these datasets are also useful.

Beside this, the consumer pyramid is also well referred, to as through the CMI. All the links are given, I am not going to open at this moment. IHDS has two rounds (I probably I missed this). But consumer pyramids is claiming to be the biggest longitudinal data. It has covered so many rounds. You will get interesting aspects on consumer expenditure, even on health issues as well. There are 27 rounds of these consumer issues, which you can follow from the CMI site. Another one is called LASI (Longitudinal Aging Survey of India). The focus here is on ageing/ aged population (45 years onwards).

Some more should also be referred. Another observational data set is the WHO SAGE dataset. It has a focus on the health systems, so will be useful. Some work you can also refer to from World Bank, especially at a micro level, it represents six states, selected districts and the impact of COVID etcetera can also be derived from this World Bank micro data. If you search or click on this link, it will be directed to that page. Given the time limit, I am not explaining, I am happy to guide you in our live sessions.

Besides that, the indicator derived from observation studies data includes mortality, morbidity rates, healthcare utilisation and their related expenditures, disease prevalence. Like, within mortality, we will have a complete coverage on non-communicable diseases, biomarkers, self-assessed health. It focuses on different parameters of morbidities and public health surveillance. How far do these also help in calculating morbidity, mortality rates, etc., can be tracked. In terms of ‘healthcare utilisation and their related expenditure, ' you will get enough information about out-of-pocket expenditure. These include hospital visits, outpatient care, preventive service, and cost, etc. Disease prevalence can also be tracked from those observational studies. I think you can follow from the slides.

However, there are challenges involved in observational data collection, such as data privacy and confidentiality. Observational studies data collection and usage must adhere to strict privacy regulations to safeguard individual information and maintain confidentiality. Data quality and completeness-related questions are often raised. Challenges arise, ensuring in accuracy, completeness and consistency. Hence, how far it is creating some challenges in terms of validity and reliability is up to the researcher to give or add tools and techniques to validate. Further, it has resource constraints which is a bigger challenge to have a better coverage of the survey.

So, what we are heading for the next lecture? It is on data in healthcare, especially focus is on administrtive data sources. In order to study it further, we have given you the link here, like from our module on Exploding Survey Data on Healthcare. You can follow this link. Even the YouTube platforms and these links are given for you. Rest, if you want to clarify how the survey data are used and codes etcetera are used, you can read to other articles, which we have cited here.

So, these are all, at this moment. I hope it will be helpful. Thank you. I should stop here. Thank you.