India Policy Insights
Podcast Transcript

Host: Taamra Segal, Communications and Outreach Manager, The Lakshmi Mittal and Family South Asia Institute. Harvard University

Guest: Professor S.V. Subramanian, Professor of Population Health and Geography, Harvard University

Professor Subramanian discusses his ongoing research project India Policy Insights for which he has been awarded a 2.2million-dollar grant by the Bill and Melinda Gates Foundation. His current research interests include developing and applying data science approaches for precision public policy in the context of health, nutrition, and development with a focus on India; and understanding individual and population inequalities in health from a multilevel and cross-comparative perspective.

BEGIN TRANSCRIPT

Taamra Segal: Welcome to this episode of the India in Focus podcast. My name is Taamra Segal, and I am the Communications and Outreach Manager at Harvard's Lakshmi Mittal and Family South Asia Institute. This podcast is brought to you by the Mittal Institute at Harvard University and Times of India. For this episode, we are joined by Professor Subramanian, who is a professor of Population, Health and Geography at Harvard University. His current research interests include developing and applying data science approaches for precision public policy in the context of health, nutrition, and development with a focus on India; and understanding individual and population inequalities in health from a multilevel and cross-comparative perspective. Professor Subramanian, welcome to the podcast.

Professor Subramanian: Thank you very much Taamra it's my pleasure to be here.

Taamra Segal: I'll start off with our first question: We know that any progress on key issues related to health, nutrition and population requires policies and actions that are geographically precise and effective. This is where the India Policy Initiative, a collaborative initiative comes in. Can you give us an overview of the India Policy Initiative?

Professor Subramanian: Thank you Taamra it's a pleasure to be here and let me start by first of all thanking the Lakshmi Mittal and family Harvard institute for hosting me here during this time. And thank you for taking the time to learn about the India Policy Initiative that we currently have. To give you a broad overview of what the India Policy Insights initiative is about, it is a cross cutting data science platform that cuts across various domains including health, nutrition, and population. Within these domains we are creating policy indicators that are aligned with the government of India's many programmes which touch on these domains. As part of this initiative, we will be building a comprehensive interactive dashboard that would allow users to perform various functionalities and understand how India's critical policy indicators are performing.

One of the unique aspects of this platform is its comprehensive nature with regards to geographical coverage. The geographical units or levels that India Policy Insights is focusing on includes districts. As you know districts are a very key component of administering a lot of these programmes, and a lot

of the social and health programmes in India and a primary target /partner are administrators at various levels, such as the district collectors. Another two levels of geographies that are, probably for the first time, being given salience and prominence are the parliamentary and assembly constituencies in India.

As you would know, both these geographical units have a member representative who is in charge of ensuring that the vision /needs of the constituents are reflected. Bringing this critical leadership into the discussion on policy through filling in this data gap is a key innovation of India Policy Insights. Currently there is nothing out there where you can find how a particular constituency is doing, whether it's parliamentary or assembly on any of the major policy indicators.

Last but not the least are the more micro geographic units, such as sub districts, which are a level below the districts where a lot of the implementation operation occurs in the administration of these programmes, as well as gram panchayats and villages which are the ultimate decentralized unit. And as you may know, gram panchayats and villages are both a political governance unit, i.e., there are elections there, so that's the most decentralized unit with regards to both our people's aspirations as well as the accountability at that level. Not only are these governance units, but that's also where rural India lives. People don't live in districts or states, people live in *mohallahs* and villages. That's the broadest coverage that we have, and this would be a platform that will provide a comprehensive dashboard for all our user groups to engage on how India is performing.

As a last point, our crucial partners here, who we are working very closely with, are India's leading think tank organization, NITI Aayog, where I also happen to have the privilege of being the honorary senior fellow to the government of India. They are the nodal partner in this effort. We are also trying to build partnerships with other groups, especially the *Lal Bahadur Shastri* national academy of administration that has the mandate and the mission to train our administrators right from the foundation level up till phase five when they become joint secretaries and later, secretaries. So, these are our two government of India partners and lastly, our third partner is the International Institute for Population Sciences. As we move forward, we will incorporate more high frequency data that is coming from the govt, so their administrative data.

Taamra Segal: Thank you for that wonderful overview. My next question is that as part of your research, you have been comparing data across districts. Is there anything that you have found particularly interesting/surprising as part of this research?

Professor Subramanian: We've been doing research not just on districts but also on these other geographic units as I mentioned. A few common themes have emerged in this context. We find that in India if you ask, what are the big drivers of variation, where does life differ? The one geographical unit, and one which we are aware of, are the states. The states in India are very different. On different outcomes, they tend to vary a lot. We are very well versed with this idea of state differences. However, when we look at these other units, including districts and parliamentary constituencies, yes there are variabilities within a state, between districts or between parliamentary constituencies. However, a much larger fraction of variability lies *within* a district and not *between* districts. So, a lot of the research that we've been doing, on a variety of health, nutrition and population indicators seems to consistently point out this notion of 'states matter a lot' and micro villages, assembly constituencies and most likely sub districts are going to matter a lot. So, the micro units are very important as are the

very macro units, but it's the mezzo level units that appear to be somewhat less important in relative terms. Based on this observation driven by data, for the first time we are able to truly understand this multi-level layer of variation. The states differences one can anticipate, why that might happen because different states can potentially do different things. A particular state may choose to do more in some domains such as health and nutrition whereas other states may be constrained and not be able to do so.

So, we can expect some big differences since a lot of these issues related to health, nutrition and population tend to be state subjects. The other aspect that I mentioned within districts, that's where social and economic life happens. The variation that takes place within districts is very important. These are some of the ways in which the data has revealed something different that we did not know before in such concrete ways.

Taamra Segal: How does the work done through the India Policy Initiative interplay with sustainable development goals?

Professor Subramanian: As you know, the sustainable development goals have a much broader coverage and aren't only related to health and nutrition. It includes energy and infrastructure, for example. In the initial stages we are focusing on the strength of the data. We are focusing on health and nutrition where we have had a lot of data. We are still in the process of collecting data through secondary sources, that directly or indirectly relate to SDGs. We will then develop those SDGs at each of these geographical scales. In this regard, we are also trying to work with NITI Aayog which is nodal agency in India right now. We are trying to work collaboratively within that domain.

To start with we are focusing on health and nutrition, we demonstrate what these platforms can do, and then expand that to a much wider array. When you make these platforms cross cutting, it allows you to look at indicators across domains. This offers unique advantages, and at the same time we are looking at how to make our research policy relevant and additionally we are building collaborations with specific programs that are running in the country. That helps us ground this data science platform in a way that suits that program.

Taamra Segal: How much of a discrepancy is there in data that is reported at the state and district level when compared with data that is reported within these macro zones, e.g., Gram Panchayat?

Professor Subramanian: This relates to one of the points that I made earlier. When you start looking at the village level, the variability increases. What does that tell us? That tells us that it's fine as a starting point to identify a district but what we really want is to decide where to focus our resources. Not only are resources scarce but in order to bring in efficiency and equity, you need to develop some degree of prioritization, and prioritization at a very micro scale with a degree of precision so that you get the resources to people who need them. That is the only way you can bring efficiency and equity because its first going to the village that requires it.

So, in what sense does looking at these within district geographical units, e.g., gram panchayat, villages, sub districts or Assembly Constituencies, bring us that precision in our public policy. It is in that sense that we are adding value over and above what we have come to learn from looking at state and district patterns. Having said that, what we also hope that this project will demonstrate, is we can't wait for the perfect data to arrive. At the district and state level there is plenty of data, but as we zoom in, our ability to get reliable data that everyone can agree upon, is not quite where we are yet. However,

we need to start somewhere. My vision is that we will be able to pick 50 districts even or 10,000 villages out of the half a million villages that exist across India, where we feel that the data provided by the government is reliable. How do we also build, through this exercise, whereby data quality is improved? It has to be done in a more institutionalized way.

Taamra Segal: Could you elaborate a bit on the behind-the-scenes scientific operations that are required to produce India Policy Insights?

Professor Subramanian: I really thank you for this question because when we were trying to develop this grant, a very large part of the motivation was to produce something for dissemination, for translation. There has been a recognition that there needs to be some science behind the scenes before we can produce these in a robust way, that was very important. One of the scientific efforts that we made was for example, if you have data available at the district level and you want to compute something at the parliamentary constituency level. Because the two boundaries are not *co-terminus*, how do you create an approximation for a parliamentary constituency based on district level data?

For this we used geographical information systems approaches, which is called the *crosswalk* method. It is based on population weighting with a combination of some statistical aspects as well. Secondly, when you do have sub district data, e.g., in the national family health survey we have cluster level data (villages) however, different clusters may have different sample size. These are survey data, so not everybody is measured. For example, if I have to estimate the prevalence of underweight children at the cluster level, its coming from a *sample* of children, not *all* the children. Similarly, if I aggregate all the clusters in that district, it gives me a robust estimate of the mean for that district as well as the variation but it's coming from a sample of clusters, not all the clusters. There might be a thousand clusters but you're getting information from 100 of them, because it's a sample survey.

When you have a sample survey and especially when you go to small area estimation, a lot of the scientific work here has to draw and contribute to this methodology of small area estimation. Some clusters may vary in the number of children, for example, and you need to take that into account in your estimation process. You don't want a sampling variability to be driving your estimates of the mean prevalence etc.

This was an example of where we had to tweak our precision weighted methodology in order to take account of this variability. We used machine learning models to come up with robust estimates. Science is an iterative process, so you can always improve upon it. We also want to reach out to district collectors to see if we can get some ground validation for our estimates. We are also trying to come up with more triangulated estimates. This is not an academic exercise, its ultimately about the child who will be the beneficiary. We aim to do more policy prescriptive research, as opposed to police descriptive research.

Taamra Segal: What other initiatives are you working on?

Professor Subramanian: All the other initiatives that I'm working on are driven by a very talented group of people that I'm privileged to have, which includes students, PhD students, high school students and undergraduates. But broadly speaking I have two other, bigger initiatives. One that I am engaged in is a very similar effort in the United States on creating policy indicators at the level of congressional districts. The administrative geographies do not speak very easily, so to speak, with the

electoral geographies. When we created a COVID dashboard for the congressional districts, we got a lot of responses from congressional representatives from various states or committees requesting us to do more so that they can serve their constituents well. So, I am very interested in that project that is happening, which is mainly on congressional districts.

Lastly, as a big project, I'm also intrigued and interested in this idea of heterogeneity. We have focused too much on averages and I feel it's time to move beyond averages and look at the variability, and whether it's at the individual level, or the population level. I have found that more often than not, at least in the exercises that I have done, the heterogeneity of the variation is the norm, while average is the exception. This has conceptual, methodological, and empirical implications.

These are the two big parallel initiatives that I have been working on recently and I'm also very passionate about one course that I teach at Harvard. It is on multilevel statistical models and these are regression techniques that allow you to undertake analysis simultaneously at multiple scales. So both individual and population, and you can define population in various ways.

This is something that I've been very fortunate to be able to teach at the Harvard school of public health and I get amazing students who continue to push me. I'm also hoping that as India Policy Insights is developed, that will lend itself to training and education programs. Training programs are already part of the scope for the administrators, but I'm also thinking these could be seen as educational platforms, right from the high schoolers to graduates.

Taamra Segal: That was wonderful, thank you so much for joining us for this podcast, it's been a pleasure to host you and to learn more about the work that you've been doing

Professor Subramanian: Thank you so much Taamra this has been a great pleasure.