**Title: Population Heterogeneity Under SSPs at the Sub-National Level in India**

**Authors**

**Samir KC**[1,2,3,4] and Moradhvaj Dhakad[1,2]

1. International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria;
2. Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/OAW, WU), Vienna, Austria
3. Asian Demographic Research Institute (ADRI) at Shanghai University, Shanghai, China
4. Population and Statistics Research Hub (PSRHUB), Lalitpur, Nepal

**Abstract:**

India is a subcontinent of significant heterogeneity in terms of demographic and socioeconomic indicators. Population and urbanisation projections are developed nationally through shared socioeconomic pathways (SSPs). Both spatial and socioeconomic population heterogeneity are ignored, resulting in biased future population changes. Applying a multidimensional demographic model, this study tested different alternative assumptions on fertility, mortality, migration, and education transition corresponding to SSP1 (Rapid Development scenario) and SSP4 (Inequality scenario) at the subnational level and compared them with national-level assumptions. Preliminary results show that India's total population size and the trajectories significantly differ between the national and subnational narratives corresponding to SSP1 and SSP4.

## Introduction

In 2023, India overtook China as the world's most populous country, accounting for nearly 18% of the world's population. India has unique importance due to its sheer population size and heterogeneity. India's population changed marginally from 235 million in 1901 to 358 million in 1951, rapidly increasing to 1.43 billion by 2021[1]. Within India, according to the last census in 2011, Uttar Pradesh (UP) was India's most populous state (200 million), larger than the world's fifth-most populous countries, Brazil (195 million) and Pakistan (174 million)[2]. India's least populous territory, Lakshadweep, had about 60,000 inhabitants, equal to Greenland and Bermuda.

A population projection is important information used to determine future population trends and their impact on socioeconomic and environmental factors. They are core inputs for understanding future needs related to health care, education institutions, employment, social security, transportation, housing, jobs and energy facilities, and other goods and services (Ref). Population studies commonly demand population projections for planning and understanding the future climate change-related processes in which the population is either a contributor via consumption or is vulnerable to adverse impacts. Changes in population affect the demand for energy and natural resources, which has a detrimental effect on pollution, biodiversity, and climate change, and also affects the resilience and vulnerability of society to environmental changes (Muttarak and Jiang 2016)

As such, on request from the climate change research community, demographers have developed the population and urbanisation projections at the national level in the shared socioeconomic pathways (SSPs) framework.[3]. However, spatial and socioeconomic population (mainly in SSPs' urbanisation projection) heterogeneity is ignored, resulting in biased future population changes at the national and global levels. Further, in India, KC et al. (2018) have shown that considering 35 states (S) and rural/urban (R) place of residence in the population projection (in addition to age, gender, and education - AGE) leads to a much higher total population compared to AGE only projection. The difference is mainly due to larger populations in areas with lower socioeconomic levels (e.g., UP and Bihar), where higher fertility rates are increasing faster.

India is a subcontinent of significant heterogeneity regarding demographic factors such as fertility and mortality levels and socioeconomic development indicators like education and per capita income across the regions. The total fertility rate (TFR) was reported to be the highest at 3.0 (urban=2.4, rural=3.1) in Bihar and the lowest in Punjab at 1.6 (urban=1.6, rural=1.7) during 2019-21[4]. Life expectancy at birth varied from 63.2 and 66.5 years in Madhya Pradesh to 72 and 78 years in Kerala for males and females, respectively, during 2011-15[5]. The child mortality rate (4q1) among illiterate mothers was five times the child mortality among mothers with 12 years of schooling (IIPS and Macro, 2021). Migration patterns differ mainly across the states; Maharashtra and Delhi have the largest net in-migration, while Uttar Pradesh and Bihar have large numbers of net out-migration[6]. Education attainment varies considerably across the region in India. Andhra

[1] United Nations, "World Population Prospects 2022: Summary of Results."
[2] ORGI, "Data 2011 Census of India."
[3] KC and Lutz, "Demographic Scenarios by Age, Sex and Education Corresponding to the SSP Narratives"; Kc and Lutz, "The Human Core of the Shared Socioeconomic Pathways."
[4] IIPS and ICF, "National Family Health Survey (NFHS-5), 2019-21."
[5] ORGI, "Compendium of Sample Registration System Year Books (1981–2016)."
[6] ORGI, "Data 2011 Census of India."

Pradesh has the lowest literacy at 66 percent, and Kerala has the highest literacy at 96 percent. Similarly, the per capita income differs largely within the country, from US$7,032 for Goa to US$ 681 for Bihar.

**Demographic transitions:**

There was a wide variation in the speed and initiation of the demographic transition among these groups. Differences will influence variations in regional population growth in fertility and mortality levels. Urban areas like Tamil Nadu, Kerala, Himachal Pradesh, and West Bengal reached a fertility level below replacement level in 1991, and their population will peak in some years. On the other hand, some states in northern India, such as Bihar, Uttar Pradesh, Chhattisgarh, and Rajasthan, continue to have higher TFR and grow at a high rate for many years. The demographic transition also leads to significant changes in the population's age structure, which leads to the differential impact of population momentum on future growth. The regions with higher fertility rates will continue to experience population growth despite declining fertility rates due to population momentum. Overall, fertility and mortality rates are higher in rural and urban areas and differ by state. It is well known that socioeconomic developments significantly influence the demographic indicators of fertility, mortality, and migration. The variations in socio-economic development across the regions continue to cause demographic heterogeneity.

**Translation of SSPs storylines at the subnational level**

Considering the regional heterogeneity in India, this paper analysed the projected population from national and subnational alternative narratives on future fertility, mortality, migration, and educational transitions for SSP1 (rapid development) and SSP4 (Inequality). Using KC et al.'s[7](2018) model, we first define the SSP narratives for India. Then, we deliberate whether national-level SSP assumptions (e.g., the low fertility variant of SSP1) can be applied to all subnational regions. In doing so, India's states are divided into four groups based on current fertility levels by rural and urban areas defined by a TFR of more than 2.1 in 2019: "High fertility region (High-Fert)" and TFR of less than and equal to 2.1 are considered low fertility regions (Low-Fert).

Therefore, considering the general rationale of SSPs and the storylines underlying the individual SSPs, this paper defines new storylines for two SSPs (SSP1 and SSP4) at the subnational level for India. These storylines are translated into alternative fertility, mortality, migration, and education assumption scenarios for the subnational groups.

The details of the assumption under each SSP at the subnational level are given in Table 1. The SSP1 assumes the future is moving toward a more sustainable path. In particular, the story narrates that education and health investments accelerate the demographic transition, leading to a relatively low population. We assumed a low fertility trajectory in both Low-Fert and High-Fert regions in rural areas and High-Fert in urban areas. Under this scenario, improvement in health care will reduce the unmet need for family planning. Investment in education will raise the marriage age, and women's employment opportunities will result in lower family size. In Low-Fert urban regions, we assumed a medium-fertility trajectory because there would be more favourable

[7] "Future Population and Human Capital in Heterogeneous India.", 2018, PNAS

conditions for women to care for their children and their personal growth. Women from the Low-Fert in urban areas can maintain their personal growth with more children's pronatalist conditions.

The SSP2 scenario continues with uneven progress toward achieving the development goal of recent decades. Under this scenario, we assume all areas will follow the medium path for fertility, mortality, education, and migration.

The SSP4 is the scenario of high inequality within countries. The limited access to high-quality education, health services, and family planning leads to high population growth. Women from this group have a higher number of children due to lower education and economic participation, low mean age at marriage, and limited access to modern contraception. The overall poor status of women in this region will continue to keep higher fertility. We assume the High-Fert groups will experience higher fertility and mortality in rural areas; this group is left behind in education, economics, and other indicators. In addition, low levels of maternal nutrition and public health infrastructure limitations will continue to cause higher mortality for this group. While the urban population has a relative advantage over the rural population, we assume that the high-fert urban regions will follow medium fertility and mortality trajectories. In the Low-Fert group, which already has lower fertility, rural areas could experience medium fertility and mortality trajectories. In contrast, the urban areas will have better conditions and follow low fertility and mortality trajectories.

Table 1 Matrix under the SSPs definitions for fertility, mortality, migration, and education at the subnational level (number of areas in parenthesis)

| | | Fertility | | Mortality | | Migration | | Education | | Reclassification |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Rural | Urban | Rural | Urban | R-to-U | U-to-R | Rural | Urban | |
| SSP1 | LowFert. | Low (24) | ***Med*** (32) | Low | Low | Med | Med | High | High | High |
| | HighFert. | Low (11) | Low (3) | Low | Low | Med | Med | High | High | |
| SSP2 | LowFert. | Med | Med | Med | Med | Med | Med | Med | Med | Med |
| | HighFert. | Med | Med | Med | Med | Med | Med | Med | Med | |
| SSP4 | LowFert. | ***Med*** | Low | Med | ***Low*** | Med | ***Low*** | Low-Med | Low-Med | High |
| | HighFert. | ***High*** | ***Med*** | ***High*** | Med | ***High*** | ***Low*** | Low-Med | Low-Med | |

**Data and Methods:**

This study used population distributions by age, sex, and educational attainment from the Indian Census 2001, 2011, Age-specific fertility rate (ASFR) by educational attainment and rural and urban regions, and life expectancy for 20 larger states taken from Sample Registration Surveys (SRS) between 1999-2020[8]. The overall mortality levels by educational attainment were also taken from the SRS. The under-five mortality by mother's education was estimated from the National

[8] ORGI, "Sample Registration System Year Book 2020."

Family Health surveys between 1992-93 to 209-21[9]. The life expectancy by education attainment in the adult age group was estimated from the India Human Development Surveys between 2004-05 and 2011-12. The migration data from Census 2011 is extracted from Tables D03 and D04[10]. For this study, we used migrants by place of last residence (by state and rural-urban), place of destination (state and rural-urban), age, sex, education, and duration of residence (<5 years). The education attainment data was retrieved from the census 2011 and the National Sample Survey Office for 2017-18[11].

*Multi-State Demographic Projection Model (MS-Dem):*

This study used the methods of multidimensional population dynamics to deal with the sources of population heterogeneity in addition to the conventional age and sex structure, which is explicitly differentiated by educational level, urban-rural place of residence, and residence in high fertility and low fertility regions.

We used statistical software R in our calculation and have developed an R-package named **M**ulti-**S**tate **D**emography. This package can model population projections by age and sex and any combination of three more dimensions, namely, education and four sub-national dimensions - rural/urban and/or High-Low-Fert group of the rural-urban region of India. The package was released in July 2017 in R-forge (https://r-forge.r-project.org/R/?group_id=2281), and the first update was released in Jan 2018.

**Results and conclusion**

Figure 1 shows age-sex and education pyramids for the 2011 population composition for India and 35 states by rural-urban. In India, a larger proportion of women have never been to school, which differs greatly across the regions. They are important sources of population heterogeneity, influencing future population dynamics. Some regions have a larger share of the no-education group, while some areas have a large percentage of the secondary-educated group.

Figure 1 Population structure by age-age and education, 2011, India, and in states by rural-urban regions (names not shown).

[9] IIPS and ICF, "National Family Health Survey (NFHS-5), 2019-21"; IIPS, "National Family Health Survey (MCH and Family Planning, India 1992-93"; IIPS and ICF, "National Family Health Survey (NFHS-4), 2015-16: India"; IIPS and Macro International, "National Family Health Survey (NFHS-3), India, 2005-06"; IIPS and ORC Macro, "National Family Health Survey (NFHS-II), 1998-99: India."

[10] ORGI, "Data 2011 Census of India."

[11] National Sample Survey Office, "Key Indicators of Household Social Consumption on Education in India, July 2017-June 2018, NSS 75th Round."

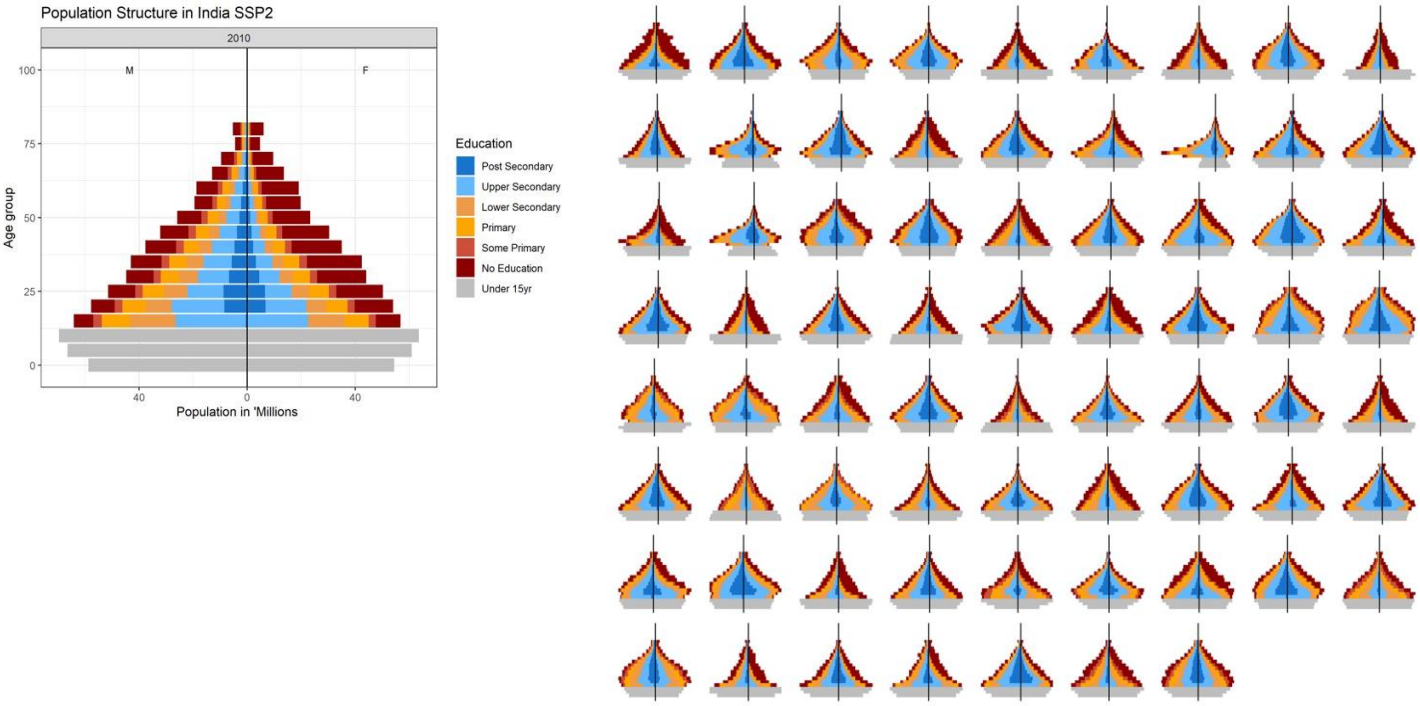



Figure 2 compares India's population projection based on the national and subnational assumptions under the SSP1 and SSP4 scenarios and the medium variant under SSP2. Under the SSP1 scenario, population sizes based on the sub-national assumption are relatively higher than national-level assumptions. The differences between the projections from national and subnational assumptions under the SSP4 are much more prominent. Similarly, the urbanisation process is higher under the subnational assumption under SSP1.

Figure 2 Population size and urbanisation of India projected from the national level (SSPnat) and subnational level (SSPnew) assumptions in SSP1, SSP2, and SSP4.

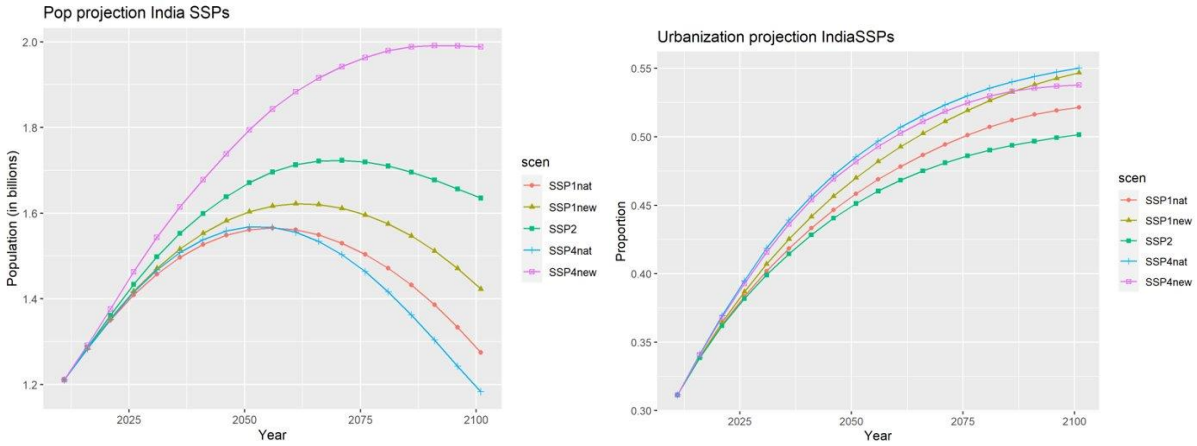


Figure 3 shows that under SSP1, the proportion of the no-education population has reduced over the years, and the upper secondary and post-secondary population share has risen to the peak of the population in 2060. Meanwhile, under the SSP4, higher population growth peaked in 2090, and the larger population remained uneducated.

Figure 3 India Population in 2011–2100 by broad age group and educational attainment under SSP1, SSP2, and SSP4 scenarios.

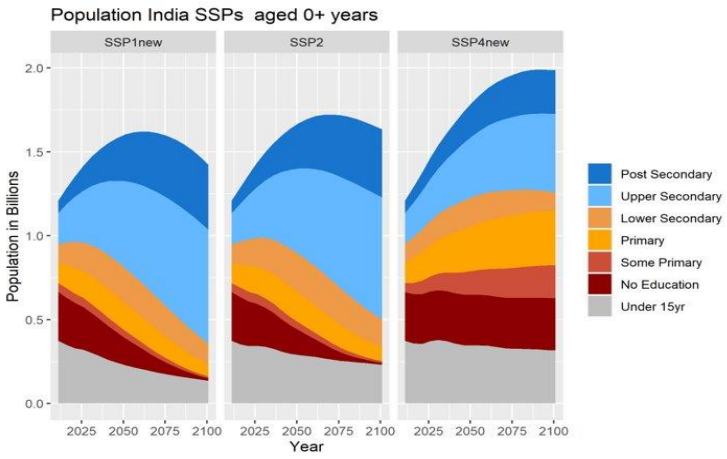
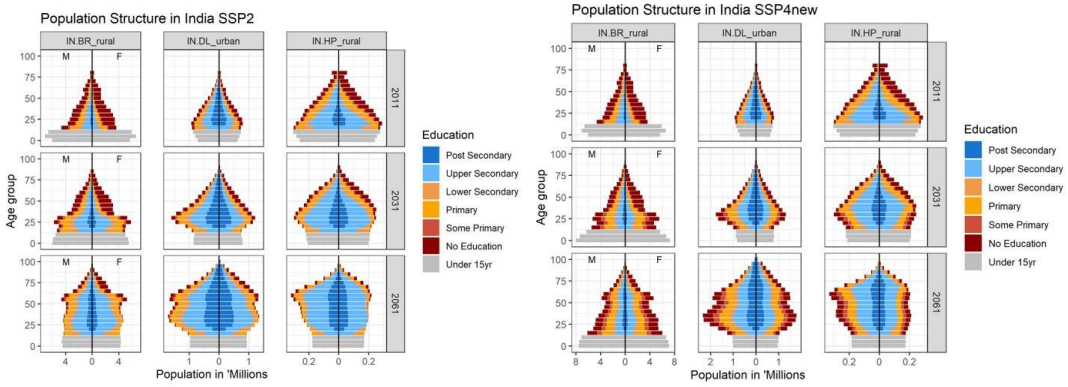



Figure 4 compares the age, sex, and educational attainment pyramids for selected regions of India projected under SSP2 and SSP4 scenarios. It indicates significant heterogeneity in the population by age and education attainment for these regions.

Figure 4 The projected population for selected regions: Bihar rural, Himachal Pradesh rural, and Delhi Urban under SSP2 and SSP4 scenarios.




This study concludes that considering the demographics and education, regional heterogeneity is important for individual SSPs narratives in India. Therefore, we will focus on refining our narratives and the model with the latest available data, particularly migration.