International Institute for
Applied Systems Analysis
Schlossplatz 1
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342
Fax: +43 2236 71313
E-mail: repository@iiasa.ac.at
Web: www.iiasa.ac.at

Working Paper **WP-17-004**

Population projection by age, sex, and educational attainment in rural and urban regions of 35 provinces of India, 2011-2101: Technical report on projecting the regionally explicit socioeconomic heterogeneity in India

Samir K.C. (kc@iiasa.ac.at)
Markus Speringer (speringe@iiasa.ac.at)
Marcus Wurzer (wurzer@iiasa.ac.at)

Approved by

Wolfgang Lutz
Program Director, World Population Program
April 2017

Contents

Abstract

This working paper documents and explains our methodological approaches and technical details about how we conducted subnational population projections for India. This research is motivated by two research questions: **(1)** *How does the accounting of socioeconomic heterogeneity, measured by educational attainment, improve population projections for India?*, and **(2)** *How will changing patterns in urbanization affect the population projections, depending on the spatial scale (national vs. subnational) considered in the projections?*

Projections at national and subnational level can provide essential information for planning and implementing government policies, including the allocation of budget and resources. In a country like India national projections ignoring spatial and socioeconomic heterogeneity would be too short-sighted considering its sheer population size of 1.2 billion in 2011.

It was surprising to see that our population projections for India with baseline scenario were consistent with the UN medium variant and Wittgenstein Centre SSP2 until 2070. We found that while our fertility assumptions are lower, our mortality assumptions were also lower and compensated for the lower number of births (and no international migration) with higher number of survivors. The results show that the overall fertility for India is lower than estimated/assumed by UN and Wittgenstein Centre due to lower starting values in our projection as well as due to explicit consideration of education in the model. This results in a rapid TFR decline to about 1.85 children per woman in the next two decades and stabilization for the rest of the century. The projection resulted in slower rate of urbanization in India from 31% in 2011 to 40% in 2051, compared to the UN urbanization projection and we presented several explanations for that.

Acknowledgments

This working paper provides a detailed technical documentation about the methodological approach developed and implemented by Samir K.C., Marcus Wurzer and Markus Speringer to project subnational population (the first administrative level [**FAL**] below national level) by education and residence for India. The methodology is applicable to other country cases in the world, whereby we are currently focusing on Asia to produce comprehensive datasets for multiple countries including China, Nepal and others.

This work is part of the IIASA cross-cutting project entitled "Socioeconomic Heterogeneity in Model Applications" (SCHEMA), that focuses on the integration of different IIASA models GAINS, GLOBIOM, MESSAGE[1] and our projection model in an interdisciplinary research process between IIASA's Air Quality and Greenhouse Gases Program, Energy Program, Ecosystems Services and Management Program and World Population Program.

We would like to thank our colleagues at IIASA, Gregor Kiesewetter, Narashima Rao and Hugo Valin, who are working with us at the SCHEMA Project, for the constructive work and vital discussions that made this paper and all the correlated efforts possible.

Participants of the course on multistate population projection in Bangalore India, especially Akansha Singh who helped preparing fertility data from the Sample Registration Survey.

[1] See model descriptions in appendix.

About the Authors

Samir KC is Project Leader of the Modelling Human Capital Formation project at the World Population Program at IIASA, Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU). He is also Professor and Leader of the Human Capital and Development group in Asian Demographic Research Institute at Shanghai University.

Markus Speringer is Research Assistant at the World Population Program at IIASA, and Research Scientist at the Vienna Institute of Demography, of the Austrian Academy of Sciences, Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU).

Markus Wurzer is Research Scholar at the World Population Program at IIASA, Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), and Lecturer at the Institute for Statistics and Mathematics, Vienna University of Economics & Business.

Population projection by age, sex, and educational attainment in rural and urban regions of 35 provinces of India, 2011-2101: Technical report on projecting the regionally explicit socioeconomic heterogeneity in India

Samir K.C.
Markus Speringer
Markus Wurzer

1 Introduction

Scholars and policymakers have increasingly realized that the causes of changes in environment and life support systems and its consequences on human societies are largely determined by not only total population growth but also changes in population compositions. This includes the consideration of heterogeneous characteristics of the populations and their spatial distribution, answering questions who the populations are (age, sex, educational attainment, etc.), and where they are (rural or urban residence, and geographic locations e.g. coastal or hinterland). Although almost all integrated assessments of socioeconomic and environmental changes consider population dynamics as one of the key driving forces, most of the modeling and analysis use population size as a scalar and the only demographic variable, and ignore the important impacts of socio-demographic heterogeneity.

To bridge the gap, IIASA initiated an internal cross-cutting project "Socioeconomic Heterogeneity in Model Applications" (SCHEMA). The main goal of the project is to address the following question:

> *How does better accounting of SCHEMA in systems analysis improve our prediction of global environmental change and human well-being and the design of related policies?*

This question is being addressed through a cross-cutting activity involving four IIASA programs and the following large-scale IIASA models: GAINS, GLOBIOM, and MESSAGE[2].

Specifically, the research focuses on the following overarching questions:

- *How will changing patterns in urbanization and income distribution influence the patterns of human consumption (e.g., food, energy), and what are the associated pressures on the environment and human well-being (e.g., clean air)?*
- *How do environmental policies affect different socioeconomic groups, and overall inequalities and social justice?*

[2] See model descriptions in appendix.

These refinements in model specifications will enable new policy-relevant research questions to be informed by IIASA models using a broad set of well-being metrics. This project will also develop a common knowledge pool on the representation of socioeconomic heterogeneity, and strengthen the information flows between IIASA models.

Specifically the work of our group is motivated by two research questions with focus on social and spatial heterogeneity in population dynamics: **(1)** *How does the accounting of socioeconomic heterogeneity, measured by educational attainment, improve population projections for India?*, and **(2)** *How will changing patterns in urbanization affect the population projections, depending on the spatial scale (national vs. subnational) considered in the projections?*

In this working paper we aim to illustrate the technical and methodological details about how we developed demographic projections of Indian population, disaggregated by age, sex, educational attainment, rural/urban residence, and by 35 states, as the first exploration study. The projected demographic dynamics, and income distributions, will serve as fundamental indicators of socioeconomic heterogeneity and input for IIASA's energy model MESSAGE, food and land use model GLOBIOM, and air pollution model GAINS in studying future changes in energy consumption, food demand, transportation demand, and air pollution.

2 Data and Methods

As a first step, we acquired data from the website of the Office of Registrar General of India (ORGI). The following data from two latest Censuses (2001 and 2011) and various reports from the Sample Registration Survey (SRS) for the period since 1999 were used in this exercise:

- Population distribution by age, sex, and educational attainment (Census 2001, 2011).
- Age-specific fertility rate (ASFR) by educational attainment and rural and urban regions, and for 20 larger states 1999-2013 (SRS)[3]. Recently, the ASFR by rural and urban regions for the remaining 15 smaller states and union territories (UT) were published.
- Life expectancy by rural and urban regions and for 17 larger states[4] – 2010-2013 (SRS). The Crude Death Rate (CDR) is used for proxy states.
- Internal migration between rural and urban regions of states/UT (Census 2001, data not yet released from 2011).

These data were first processed to produce distribution and estimates consistent with our age and education categories. We defined six levels of educational attainment, namely: "no education", "some primary", "completed primary", "completed lower secondary", "completed upper secondary" and "completed post-secondary". The population was disaggregated in five-yearly age groups with '100+' as the last age group.

[3] Missing states/UT in SRS's fertility data: Andaman & Nicobar Islands, Arunachal Pradesh, Chandigarh, Dadra & Nagar Haveli, Daman & Diu, Goa, Lakshadweep, Manipur, Meghalaya, Mizoram, Nagaland, Puducherry, Sikkim, Tripura, and Uttarakhand.

[4] Missing states/UT in SRS's life tables: in addition to 15 missing in SRS's fertility data (see Footnote 3) Delhi, Chhattisgarh, and Jharkhand.

The definition urban type of residence is according to the Census 2011 and was based on population size (>5000 inhabitants), population density (400 inhabitants per km^2), and proportion of males working in non-agricultural sector as main occupation (>75%). Local administrative units that fulfill these criteria should be classified as "Census Town" (CT). Deviations from this rule are possible, but in general this would mean that villages exceeding this thresholds would be automatically reclassified from rural to urban areas. Additionally the urban definition identifies administrative urban regions that are known as "Statutory Town" (ST). (See details in Section 2.4, Census India, 2011)

Details of estimation methods and their results are discussed in the following sections: fertility (Section 2.1), mortality (Section 2.2), internal migration (Section 2.3), urbanization (Section 2.4), and educational attainment (Section 2.5).

Future scenarios are defined in the last sub-section (2.6). So far, we defined a business-as-usual-scenario (medium variant scenario) or baseline scenario by mostly continuing the trend. For the baseline scenario, we defined future pathways for the three demographic components fertility (2.1.4), mortality (2.2.2), and (internal) migration (2.3.1) for each of the 70 sub-national units (35 states/UT and urban/rural residences). We could get only education-specific data for fertility. We also defined urbanization process through reclassification of administratively rural areas to urban areas (see Section 2.4.2). For the education component we defined five pathways for education transitions for the baseline scenario (Section 2.5).

2.1 Fertility

2.1.1 A fertility pathway for India

The demographic transition theory explains how fertility declines from a very high level (~7 children per women) to a low value (~2.1 children per woman, below replacement level). India is currently moving towards the end of the demographic transition, with a TFR of 2.32 children per woman in 2013 (ORGI, 2014).

Assumed decline is too slow in global level projections…

The trend extrapolation indicates further decline in the fertility level of India. The United Nations Population Division (United Nations, 2015) expects that the fertility will decline below 2.0 by 2035-2040 (with 2.48 estimated for 2010-2015). The speed of decline seems to slow down with less than half a child in the next 25 years, however, the result is based on an extrapolation of India's rate of decline in the past along with the experience of other countries that have gone through similar levels (i.e., TFR of 2.5 or so) in the past. The question is whether this speed assumed by UN is reasonable or is it too slow?

Projection done by the Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU) (WIC) has similar expectation for the future that starts with a bit higher TFR value (2.54) in 2010-2015 and goes to less than 2.0 by 2035-2040. (Lutz et al., 2014; WIC, 2015) India's earlier projection of TFR, conducted based on Census 2001 data (ORGI, 2006), predicts the fertility to decline to 2.52 by 2011-2016, starting with 3.13 in 1996-2001, which is quite close to the WIC starting point for 2010-2015 (derived from the UNPD 2010 revision).

…but local projection has caught the faster decline

However, the SRS indicates that the Indian fertility might have declined faster than anticipated in 2001 (ORGI, 2006). What could have triggered the decline? Obvious guesses are increasing educational attainment of women, success of family planning policies (reached the mass-proliferation of 2 children ideal), urbanization, modernization (e.g., use of contraception), economic growth, etc. The expectation by ORGI predicts quite well the reported value in 2013. ORGI predicted that the fertility level of 2.0 will be reached during the period of 2021-2025, almost 15 years earlier than predicted by UN and WIC. Does this mean the projections done globally (***UN and WIC***) are wrong about the future expectations in India? What are they missing? It is possible that both have not considered spatial and population heterogeneity within India and relied on the extension of the national trend. Or maybe, Indian demographers (at ***ORGI***) were overly optimistic about the spread of the low fertility ideals, especially in rural areas of Bihar and Uttar Pradesh with highest level of fertility. However, the level of fertility seen in Kerala, Tamil Nadu, West Bengal and many other states clearly shows the plausibility of fertility declining beyond 2 children per woman.

So, the local projection experts were closer to the truth in predicting, why?

Due to high variability in the TFR level between the Indian states (spatial heterogeneity), instead of assuming overall Indian fertility and then deriving the state's fertility, a bottom up approach could have led to the expectation of faster decline. However, our preliminary results (not shown here) reveal that considering spatial heterogeneity leads to slower decline in TFR than when projected at the national level due to the weighting effect of large population size in under-developed regions in India (e.g., Uttar Pradesh and Bihar). Most of the spatial heterogeneity, however, can be explained by different composition of the population such as education level, place of residence, overall development level, ethnicity, cultural practices, religion, etc.

Rising fertility in the future…

Another observation is that the decline in fertility could reach a floor and rise again; the lowest level of fertility for different populations could vary and indicate context-specific factors (such as openness – woman actively participating in the labor force – vs traditional value regarding woman being a housewife).

We recognize that the pathways could vary for each state, however, the uncertainty about the state-specific minima remains. One way to solve this is to assume the average Indian pathway and use the minima reached by India, separately for rural and urban region (Author's opinion).

Education differential in minima

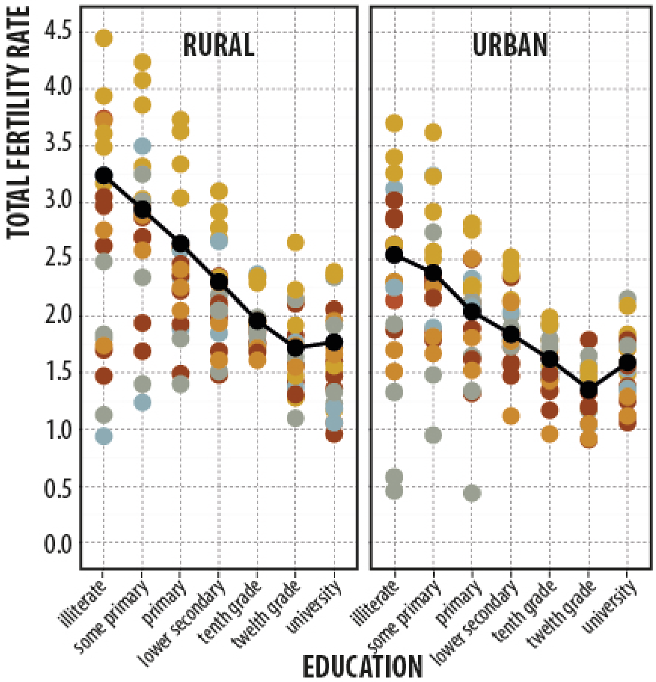
Another question is whether women with different levels of education might have different minima. We argue that for women with up to primary completed there will be higher minima than for those with secondary education. The primary completion is by the age of 12 (i.e., before women enter child-bearing age) and, essentially, the child-bearing is not disrupted by enrolment in the school/college. However, for those with secondary and above the school years spill beyond ages 15 disrupting the early child-bearing ages and therefore result in different minima.

As seen in the SRS data, the fertility level among upper-secondary educated women is the lowest in both rural and urban regions. In India, university educated women were the first group to complete the demographic transition, followed by higher secondary educated women, and then by lower secondary educated woman. Also, the transition has occurred earlier in urban regions, mostly explained by the higher proportion of educated women in these regions compared to rural areas. However, there is clearly an independent rural/urban effect that is explained by factors other than education, such as cost of living, larger residential and recreational space for children to play, family and social support in raising children, etc.

2.1.2 Education differential in fertility

In the states/UT of India exists a consistent linear education differential (Figure 1), except for the highest education group, where the TFR is higher than among the women with upper secondary education. This is based on SRS data from 1999-2013. (ORGI 2014) However, in some states/UT the differential is diminishing, meaning that the less educated women are following the path of fertility experienced earlier by women with higher education. This is true for women with no education or some primary education, and it is happening at greater speed.

Figure 1. Total Fertility Rate in India by level of education and states/UT separately for Rural and Urban place of residence (ORGI 2017)



The data also reveals that for women with education up to lower secondary, normally achieved by the age of 15, a convergence could occur quickly. For high school graduates, the school age extends to 18 and beyond. Here, one can imagine a direct impact of education on fertility of women due to the fact of simply attending school. The fertility is lowest among this group in most states/UT.

It is likely that in the future women with completed lower secondary education might follow the fertility ideal or path of the upper secondary. Hence, we could imagine one path for all. Kerala is an interesting case, with a TFR of 1.69 for women with completed lower secondary education. The TFR for women with upper secondary completion is higher (2.13) and much higher for women with postsecondary completion

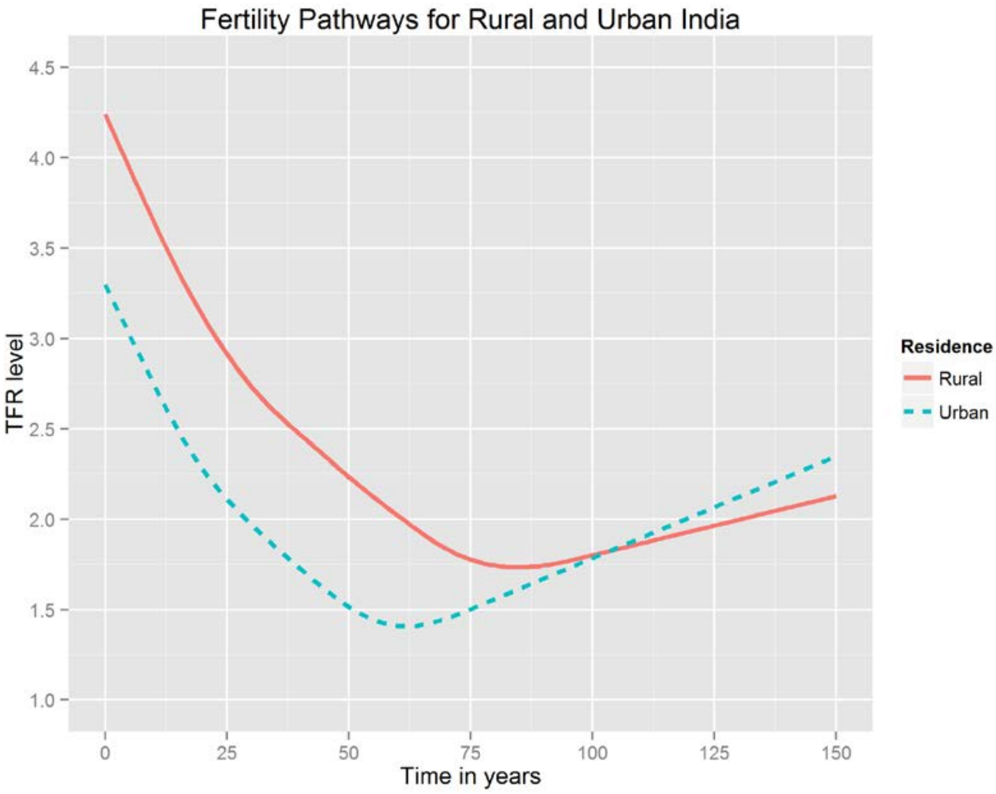
(2.35). For tertiary educated women, the fertility level after reaching a bottom, perhaps due to the tempo effect, has resurged to a higher level. This raises questions about the mechanisms of such developments, urban/rural differential, and if national fertility could reach levels as high as in Kerala.

2.1.3 Fitting fertility pathways

We defined two fertility pathways for rural and urban type of residence in India using education-specific TFR from the period 1999-2013. Trends show that fertility is declining rapidly among women with no education or some primary education. The fertility rate among primary educated and lower secondary educated is declining slowly and seems to be levelling off. The fertility rate among upper secondary educated women is the lowest in many cases, often because their education happens during the age of 15-19 years and therefore the births during this period are missed. And finally, the fertility rate among tertiary educated women is often slightly higher than among upper secondary educated women.

As explained in Section 2.1.1, after various iterations, final national pathways separately for rural and urban region were chosen by first aligning the fertility trend for each education category and then fitting a smooth spline separately.

Figure 2. Fertility Pathways for Rural and Urban India modelled based on SRS* (2007, 2010-2013)



Note: *Ultimate TFR are set 1.75 and 2.08 for rural and urban regions respectively

The smooth spline in Figure 2 shows that fertility declines to a level of 1.73 for rural and 1.40 for urban, and then increases (this is a phenomenon observed in many

Western countries). However, it is also likely that fertility will remain at the lower level, as observed in Southeast Asia (Basten et al., 2014). Very low fertility is observed in many Northern states/UT of India mostly among urban dwellers and could follow the Southeast Asian pattern. However, in many Southern states/UT as well as among the most educated women, the TFR is not so low, around 1.8 (e.g., in Tamil Nadu). Further, both UN (mostly in the range of 1.6-2.0) and WIC (1.75) assume higher levels for ultimate fertility. Based on these arguments for the baseline scenario, we assumed a TFR of 1.75 as the ultimate value at which fertility among all groups in urban regions will converge. For rural regions, we expect this ultimate level to be higher than 1.75. The gap between the ultimate fertility levels in rural and urban regions equals the gap between the minima of the two fertility pathways (0.33, see Figure 2).

The two national fertility pathways were then used to project the education-specific fertility in 70 sub-regions of India. In cases where the fertility level is already below respective pathways, the gap was allowed to remain to carry forward the low fertility behavior of women in the region.

2.1.4 Our rules for fertility projections

In the following we summarize the rules for the fertility pathway projections that we applied for our baseline scenario:

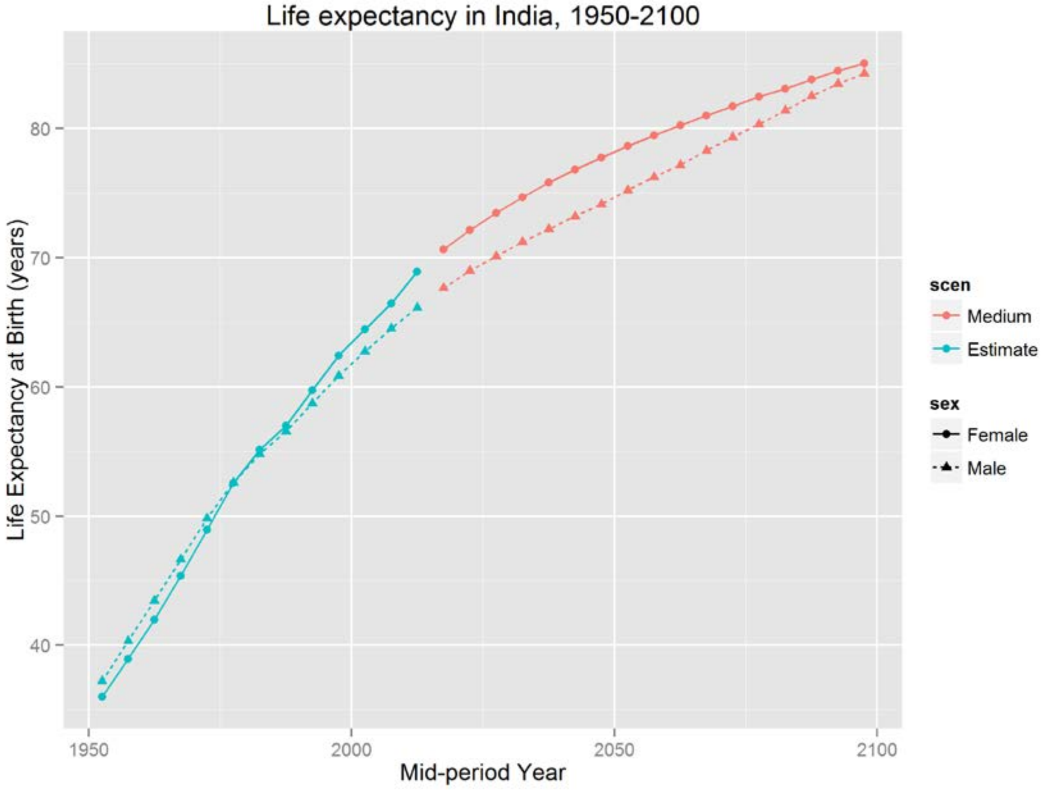
1. Fertility for women with up to completed primary education will level off at the ultimate values assumed for rural (2.08) and urban (1.75) regions.
2. Fertility for women with at least lower secondary education will follow the same path with some lag.
3. If the current value is already less than the minima, we let the difference be maintained.
4. Fertility for women with at least lower secondary education will level at 2.08 for rural and 1.75 for urban regions.

2.2 Mortality

Sex-specific life tables for each state/UT were downloaded from the SRS website (ORGI, 2014), separately for rural and urban regions for 17 states/UT[5]. These life tables were estimated based on registered deaths during 2009-2013. Unfortunately, the education-specific life tables were not available at the national and the states/UT level. So far, we could not find the education-specific mortality differential through other sources, except for infant and child mortality by mother‘s educational level in the DHS. Therefore, we did not apply the education differential in mortality and left it for future updates.

[5] Missing life tables, states/UT in SRS: Andaman & Nicobar Islands, Arunachal Pradesh, Chandigarh, Chhattisgarh, Daman & Diu, Dadra & Nagar Haveli, NCT of Delhi, Goa, Jharkhand, Lakshadweep, Manipur, Meghalaya, Mizoram, Nagaland, Pondicherry, Sikkim, Tripura, and Uttarakhand.

Figure 3. Life expectancy at birth among males and females in India, UN estimates and medium variant projection

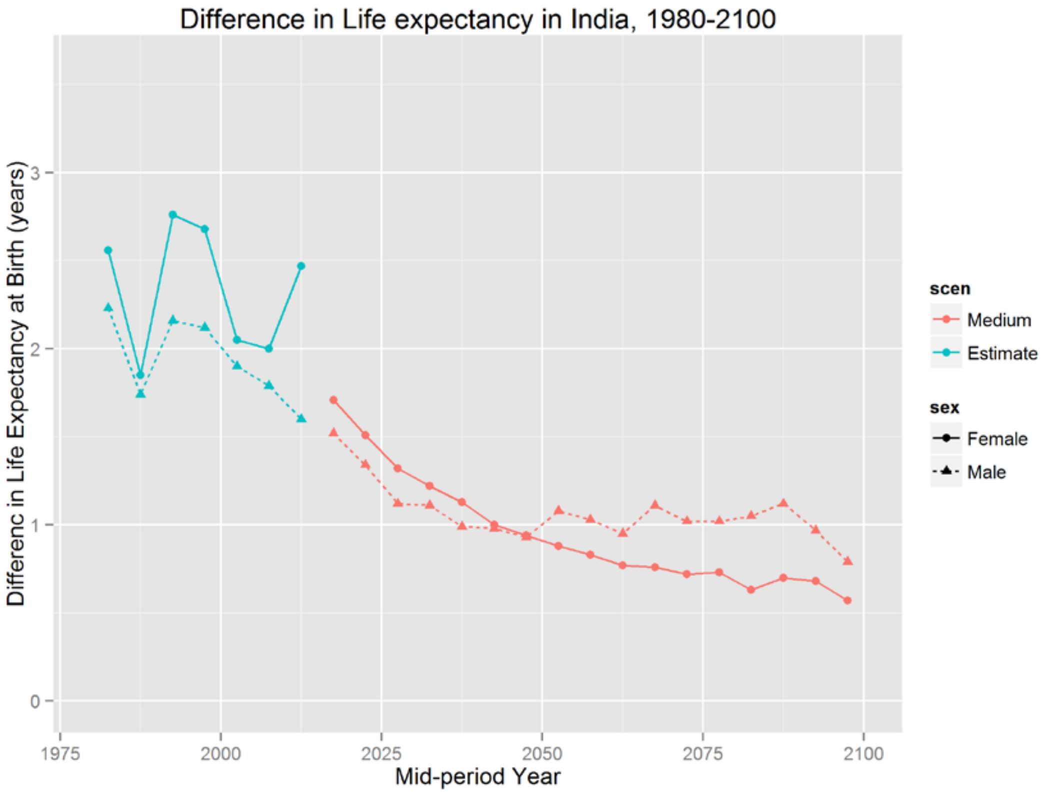


Source: The World Population Prospects, 2015 revision (United Nations, 2015)

In India, SRS estimates for life expectancy at birth for females and males were 69.3 years and 65.8 years respectively for the period 2009-2013 (midyear as 2011). (ORGI, 2014) The SRS values were slightly higher than the UN estimates for the period 2010-2015 (midyear 2012.5), see Figure 3, with 68.9 years and 66.1 years for females and males respectively. (United Nations, 2015)

In the past, the mortality situation was worse for females. For the first time, in 1980-85 life expectancy at birth among females (55.1 years) became higher than that for males (54.8 years), see Figure 3. The sex difference widened as the increase in life expectancy at birth for females increased faster than that for males, 2.05 vs 1.9 years (between 2000-2005 and 2005-2010), see Figure 4, and further widened with a gain of 2.47 vs 1.6 years for males and females respectively between the periods 2005-2010 and 2010-2015.

Figure 4. Gain in life expectancy at birth among males and females in India, UN estimates and medium variant projection



Source: The World Population Prospects, 2015 revision (United Nations, 2015)

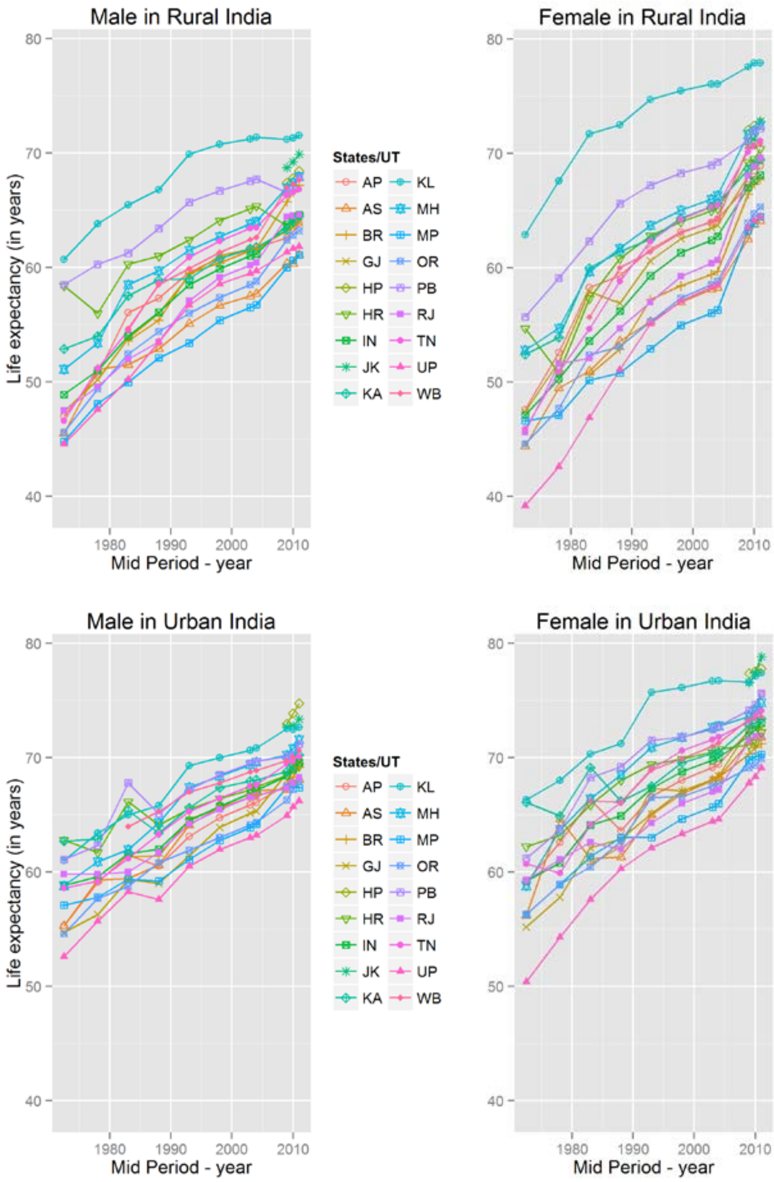
In UN medium variant, the gain in life expectancy at birth for males and females is assumed to decline in the future (see red color in Figure 4). For males, it is a continuation of the trend in the gain that stabilizes after 2040 at around one year per five years. For females, the gain for the first projection period 2015-2010 seems to be smaller than it would have been in the case of trend extrapolation. Also, in the future the gain among females will decline further, which is a result of implicit assumption in the UN projection that at the higher level of life expectancy the gain will be slower.

2.2.1 Mortality at state level

At the state level, the life expectancy at birth varies between states/UT levels. In the most recent data from SRS life expectancy at birth is always higher in urban areas compared to rural regions with the exception of Kerala. Figure 5 shows the evolution of life expectancy at birth in 17 states/UT and for whole India, separately by sex and by place of residence.

We could observe that the spatial diversity is very high in India. Among the states/UT with available data, Kerala (KL) has always been a front-runner. More recent data, that includes Himanchal Pradesh (HP) and Jammu & Kashmir (J&K), shows both States with high life expectancy at birth with highest levels in urban regions. Within urban area, Uttar Pradesh (UP) has the lowest life expectancy at birth both for males and females. Whereas within rural areas, Madhya Pradesh (MP), Assam (AS), and UP have lowest level of life expectancy at birth. Over time, the life expectancy seems to be converging rapidly in rural areas. The convergence is happening faster among females than males.

Figure 5. Life expectancy at birth in India and its 17 states, SRS estimates



Note: (AP) Andhra Pradesh, (AS) Assam, (BR) Bihar, (GJ) Gujarat, (HP) Himachal Pradesh, (HR) Haryana, (IN) India, (JK) Jammu and Kashmir, (KA) Karnataka, (KL) Kerala, (MH) Maharashtra, (MP) Madhya Pradesh, (OR) Odisha, (PB) Punjab, (RJ) Rajasthan, (TN) Tamil Nadu, (UP) Uttar Pradesh, (WB) West Bengal

2.2.2 Baseline assumption for mortality

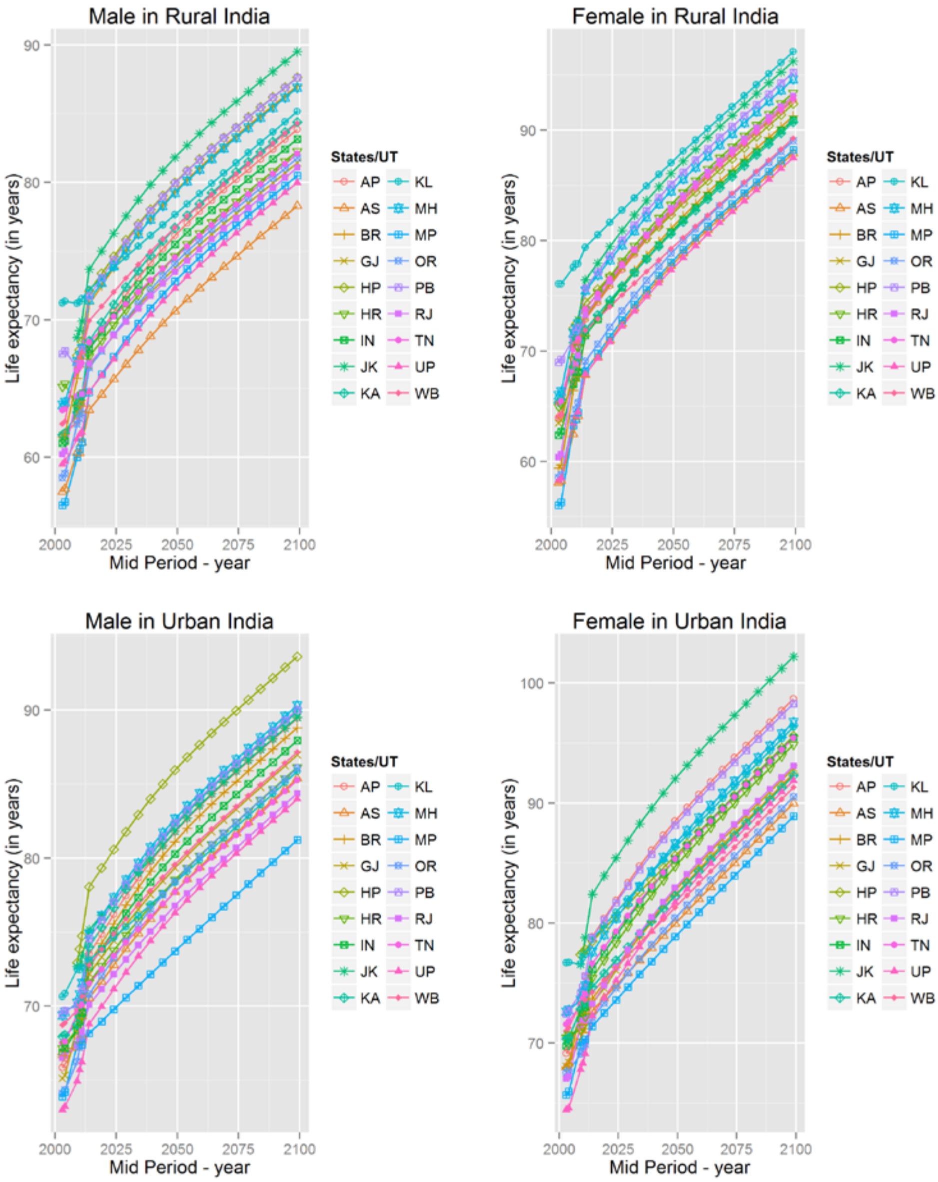
In order to project life expectancy into the future, we generated an average pathway for the future gain by regressing gain in life expectancy between two periods on the life expectancy of the initial period separately for males and females. We fitted simple linear regression and extrapolated the life expectancy into the future using the regression results and called it general predicted average gain. For each states/sex, we started with recently observed average rate of change and force it to converge to the general predicted average gain by 2030. Our narrative is that the convergence will carry on up until sometime in the future (we assumed it to be 2030, corresponding to the Sustainable Development Goals (SDG) target year) and then the regions will keep a similar rate of change in the future.

We have set a minimum value for the general predicted average rate. When it reached a certain value, we held it constant for the rest of the future, the values are 0.75 year per five years for males and 1 year for females. This leads to a widening of the gap in life expectancy between males and females, which we think will happen in the future – following the arguments by Oeppen and Vaupel (2017) that the limit to life is not yet reached. Few rules and limitations were imposed (Oeppen & Vaupel, 2017):

1. The five-yearly change in life expectancy at birth was limited to a maximum of 3 years.
2. The gain in life expectancy at birth will converge to the general predicted average gain by 2030.
3. Within each state, life expectancy in rural areas was restricted to remain lower or equal to that in urban regions.
4. The gap between rural and urban regions was limited to the most recent observed values.
5. (Not implemented yet) The gender gap in the life expectancy is not considered yet and we will further investigate to see if it is necessary.

Once the life expectancies were ready (as shown in Figure 6), we applied the Gompertz transformation method as implemented by KC et al. (2010) to produce life tables for the calculated life expectancy at birth. We used the life tables for India from the UN medium variant in the World Population Prospect 2015, as standard life tables.

Figure 6. Life expectancy at birth in India and its 17 states, SRS estimate up till 2011 and projections thereafter – convergence to national average rate of gain by 2030

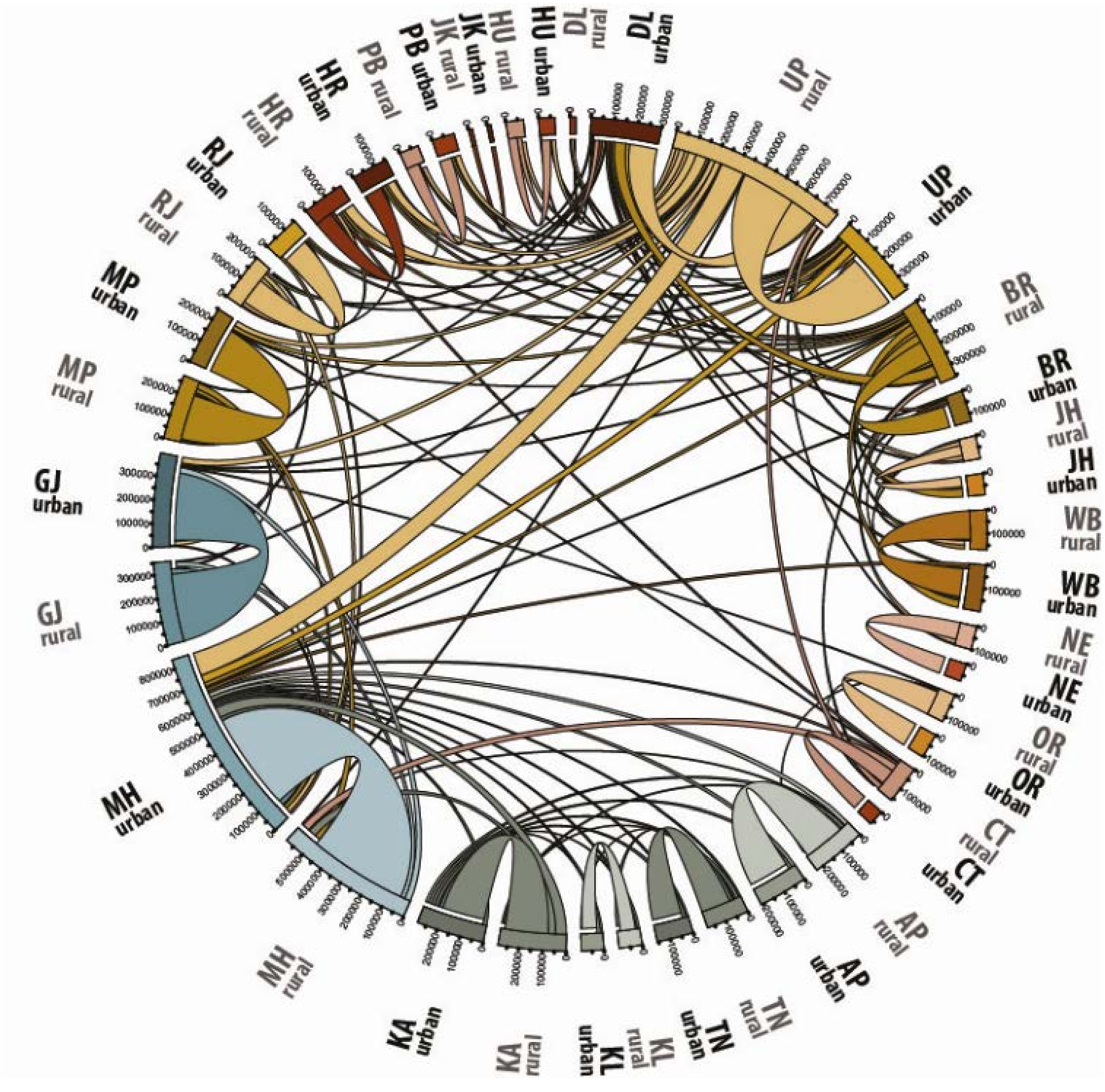


Note: (AN) Andaman and Nicobar Islands, (AP) Andhra Pradesh, (AR) Arunachal Pradesh, (AS) Assam, (BR) Bihar, (CH) Chandigarh, (CT) Chhattisgarh, (DN) Dadra and Nagar Haveli, (DD) Daman and Diu, (DL) Delhi, (GA) Goa, (GJ) Gujarat, (HR) Haryana, (HP) Himachal Pradesh, (HR) Haryana, (IN) India, (JK) Jammu and Kashmir, (JH) Jharkhand, (KA) Karnataka, (KL) Kerala, (LD) Lakshadweep, (MH) Maharashtra, (MP) Madhya Pradesh, (MH) Maharashtra, (MN) Manipur, (ML) Meghalaya, (MZ) Mizoram, (NL) Nagaland, (OR) Odisha, (PY) Puducherry, (PB) Punjab, (RJ) Rajasthan, (SK) Sikkim, (TN) Tamil Nadu, (TR) Tripura, (UT) Uttarakhand, (UP) Uttar Pradesh, (WB) West Bengal

2.3 Migration (internal)

The internal migration between rural and urban regions within and between the states/UT, altogether 70 spatial units, is one of the main determinants of the population dynamics in India. The data for the flows estimates between rural and urban regions by states/UT was not readily available and had to be estimated from different available tables from the Census 2001 (see Figure 7) as 2011 data is not yet published.

Figure 7. Internal Migration in India by states and residence from the Census 2001



Note: (AN) Andaman and Nicobar Islands, (AP) Andhra Pradesh, (AR) Arunachal Pradesh, (AS) Assam, (BR) Bihar, (CH) Chandigarh, (CT) Chhattisgarh, (DN) Dadra and Nagar Haveli, (DD) Daman and Diu, (DL) Delhi, (GA) Goa, (GJ) Gujarat, (HR) Haryana, (HP) Himachal Pradesh, (JK) Jammu and Kashmir, (JH) Jharkhand, (KA) Karnataka, (KL) Kerala, (LD) Lakshadweep, (MP) Madhya Pradesh, (MH) Maharashtra, (MN) Manipur, (ML) Meghalaya, (MZ) Mizoram, (NL) Nagaland, (OR) Odisha, (PY) Puducherry, (PB) Punjab, (RJ) Rajasthan, (SK) Sikkim, (TN) Tamil Nadu, (TR) Tripura, (UT) Uttarakhand, (UP) Uttar Pradesh, (WB) West Bengal

Following are the steps and list of data used:

i. In the first step, we extracted the data from the Census 2001[6] for the total number of migrants during the last five years (five-yearly duration) by sex in the current place of residence (by states/UT and by rural/urban, destination), and by last place of residence (origin), which gives us the volume of migration flows by sex. As a next step we estimated the age distribution of migrants at the origin and the destination.

ii. Five-yearly age and sex distribution of migrants, who have been living in the current region (destination) since less than 10 years (10-yearly duration), is available by origin state/UT and by rural/urban[7]. Age distribution of those who moved during the last five years is ideal for our projections. Since the only available data is for those who moved during the last ten years, we used the 10-yearly duration data to estimate the age and sex specific out-migration rates by dividing the age-sex-origin-destination specific number of migrants by the total pre-migration population, adjusted for the flow by taking out in-migrants and adding out-migrants from the total population, at the origin[8].

iii. A closer look at the five-yearly age pattern of 10-yearly duration migration rates revealed some anomalies that called for splitting into two five-yearly duration migration rates as our projections will be done for five-yearly age groups in five-yearly time-steps. The main problem comes from the fact that the five-yearly age distribution of 10-yearly duration migration numbers is a sum of those who moved during the last five-years and last 5-10 years with a five-year lag in age. We used all the information that was available to fill the number of migrations by five-yearly age and duration. The missing values were then filled using the iterative proportion fitting by employing the R package "*mipfp*" (Barthelemy & Suesse, 2016).

iv. In the next step, the five-yearly duration of migrations were divided by the total pre-migration population (see Footnote 8) to obtain the five-yearly age and duration migration rates. It created a total of 9660 (70 origins x 69 destinations x 2 sexes) age patterns. Each age-pattern were inspected visually to identify oddities and were corrected. The problems in the age-pattern stems mostly from very small (even no) number of migration flows between two regions. We employed the rule that if the total number of migration between two regions is less than 1000 persons, we apply appropriate overall pattern of migration rates. Corrections were also done for the migration rates in last age-groups that were exceptionally high, mostly by smoothing or by forcing a ceiling.

---

[6] Census 2001, Results Table *D-3: Migrants by place of last residence, duration of residence and reason for migration*

[7] Census 2001, Results Table *D-12: Migrants by place of last residence with duration of residence as 0-9 years and age*

[8] Migration rate = (the number of migrants) / (the number of population at origin = current population + those who left the region – those who came to the region)

2.3.1 Future assumptions

We assumed that the age- and sex-specific migration rates will remain constant between the 70 regions of India. With a single set of data, it is difficult to know the trend. Once the data on migration from the Census 2011 will be released, we will conduct further analyses and update our projections. One side-effect of setting the flow rates constant, especially between urban and rural residence, is that in the future the number of people moving from urban to rural will increase due to increase in the number of people living in urban area and vice-versa. This could result in a reverse net-migration rates between urban and rural areas. We acknowledge this effect and will try to find a solution once the data from the Census 2011 is released.

2.4 Urbanization

The change in population size and structure in urban regions could occur due to: **i)** natural increase (births minus deaths); **ii)** migration (in- minus out-), and; **iii)** reclassification of a rural region to urban and vice-versa. The first two are the inherent parts of the projection model. However, urbanization through reclassification needs a separate analysis to understand what is happening and to determine how to make future assumptions.

2.4.1 Urbanization through reclassification

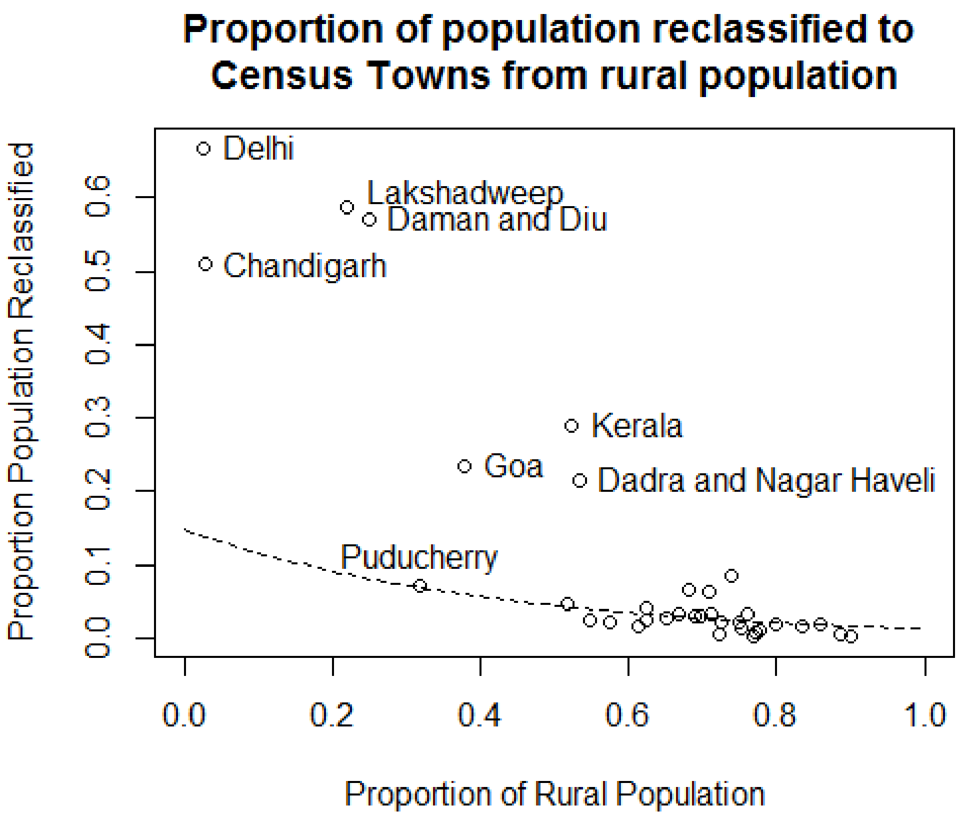
The occurrence rate of such a reclassification in terms of population is difficult to predict in the future. A recent paper by Pradhan (2013) has estimated the number of villages that were classified as Census Town (CT) in the Census 2011. The reclassification of villages into CT was based on three criteria, namely, population size, population density, and proportion of males working in non-agriculture as main occupation. The paper estimates that almost 29.5% of the growth in urban population (91m) is due to the new CTs (Table 2 in Pradhan, 2013). No urban area in 2001 was found to be declassified as village in 2011. Overall, in India about 2553 new CT were reclassified from villages within states ranging from 0 (in Mizoram) to 1 (in Sikkim and Arunachal Pradesh) to 526 in West Bengal. In Kerala, 93% of the urban growth was due to reclassification (346 new CT), which also describes the migration situation to and from urban Kerala. On the contrary, in Tamil Nadu only 25% of the urban growth was due to the reclassification (227 new CT), possibly due to the attractiveness of big cities, among others Chennai, to the migrants from the rest of India. (Pradhan, 2013) We used the data presented by Pradhan (2013) to estimate the proportion of population reclassified to CT.

2.4.2 Future assumptions

In total, Bhagat (2011) estimated that in the period from 2001 to 2011 about 44% of the urban population gains are due to natural growth, while 56% are due to net reclassification, expansion of boundaries, merge of settlements and migration. Pradhan (2013) showed that 29.5% of urban growth is due to the reclassification of rural settlements into CTs and he further implies that the remaining 26.5% are attributable to net reclassification of rural settlements into Statutory Towns (ST), the incorporation of such settlements into existing STs by expansion of their boundaries and migration. (Bhagat, 2011; Pradhan, 2013) Migration shall make up 22.2% points of this growth.

In order to make assumptions on the transition ratio from village to CT/ST for each state, we explored the relationship between various factors (rural population size, proportion of rural residence). Figure 8 shows a negative relationship between transition ratios and proportion urban population in each state. The seven outliers belong to two groups, the first group has a very high proportion of urban and smaller states/UT where more than 50% of the rural population make transitions to CT population, the second group consists of states like Kerala and Goa, also with higher proportion of urban population, but with relatively higher socioeconomic status among the states of India.

Figure 8. Proportion of population reclassified to Census Towns from rural population between 2001 and 2011



We excluded the seven outliers and fit a curve (general linear model – normally distributed error with log link function) ***log(y) = A+Bx***, (where, y is transition ratio and x is proportion of rural population). We then let the proportion of rural population at the end of each projection period predict the transition ratio from villages to CT. We let the seven outliers to be constant in the future. The predicted proportion was then used to reclassify rural population to urban population. We assumed that the age-sex-education distribution of the reclassified population will be the same as that of the rural population. In reality, the distribution of the reclassified population could be some kind of weighted average between the rural and urban distribution. We will consider this in future updates.

2.5 Educational attainment

We defined six levels of educational attainment, namely: “no education”, “some primary”, “completed primary”, “completed lower secondary”, “completed upper secondary” and “completed post-secondary”. The education distribution was available by more than six categories in the Census 2011. We aggregated for the six categories to match the International Standard Classification of Education definition (UNESCO, 2006) and studied the education transition between these six levels of education.

For a given educational attainment level, we defined the education attainment progression ratio (EAPR) to the next educational level as the proportion who completed the next level of educational attainment among those in the current level. For e.g., if in a cohort 90% have completed at least primary education and 45% have completed at least lower secondary (see Appendix, Table 2), then the EAPR to lower secondary completion is the ratio of the proportion of those who completed lower secondary education to those with at least primary education completed (i.e., 45%/90% = 0.5).

The education distribution in older cohorts provides information about cohort-specific education transitions in the past, which is necessary to study the trend. Distribution and transitions from consecutive cohorts can be used to analyze the trends in different education categories (see Figure 9).

Figure 9. Educational Attainment Progression Ratios in India for five educational levels among males and females by place of residence (rural and urban) in 2011 (Source: Census 2011, and own calculation)

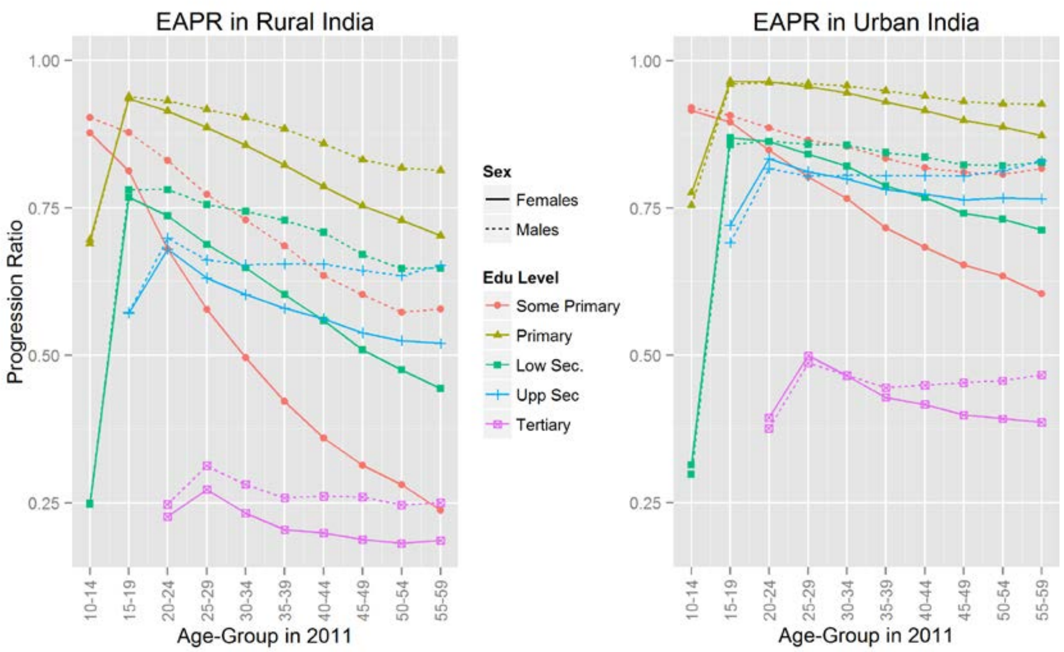


Figure 9 shows the EAPR in five-yearly cohorts for rural (left panel) and urban (right panel) regions of India for five educational attainment categories (five colors) for males (dashed lines) and female (solid lines) reported in the Census 2011.

An alternative way to look at the educational attainment is to consider proportions as well as numbers. The numbers are subjected to the effect of changing cohort size and hence won't tell us the direction of relative change. To study some specific educational attainment, where the number attaining is very small compare to the overall cohort size, and the high-cost investment in it, number is the right indicator. For example, in poor countries the degree in Engineering and Medicine is often restricted to a very small number, and policies and decisions are made on the numbers such as to increase the number of seats by 100 or so. The proportion as well does not provide full information about the base (lower level of education, i.e., transition) and only gives us information of the education level in question.

EAPR gives us a true sense of educational attainment (i.e., by making transition from one state to the next). It is also useful for policy makers, especially when comparisons in relative terms between places, different age groups, as well as different educational levels need to be made. We calculated the EAPR for five education categories, the first transition is entering or enrolling in a school for the first time; the next category is to primary completion, and so on until the post-secondary completion (at least first degree after the high school).

We analyzed each of the trends drawn from several cohorts and defined future education scenarios essentially by extrapolating the trend and, in some cases, by applying some 'expert' opinion. For example, while all other transitions were allowed to become universal, the transition from upper secondary to tertiary was limited to 70% in urban and to 50% in rural areas. Also, for those regions with slower speed of change than the national one (by states/UT, residence and sex), we allowed the speed (slope) to converge to the national one.

2.5.1 EAPR to some primary

The EAPR to at least 'some primary' represents the proportions of those who have ever been to school, we termed as EAPR1. India still faces the challenge to bring everyone into the school. Between states/UT by place of residence, the range among age-group 10-14 by sex is between 77% to more than 98%. Surprisingly, among rural females in Punjab, Uttarakhand, Karnataka, and Gujarat less than 84% have ever been to school but less developed states, such as Uttar Pradesh and Bihar, have almost universal enrolment (more than 95%).

At the national level, there is no gender gap between urban areas and very little gender gap in terms of favoring boys remains in rural areas. However, at the state/residence level, the gender gap in terms of favoring males among 10-14 year old's ranges from -4% to +9.

At the national level, the gap between rural and urban regions is almost zero. However, the gap among 10-14 year old's is much bigger between states/UT in the range of -6% to 15% favoring urban dwellers. The worst gap is in states such as Chhattisgarh, Punjab, Gujarat. The exceptions were observed notably in Tamil Nadu and Uttar Pradesh where the gap is in favor of rural residents.

In our projections, first we estimated the trend by linearly regressing the logit of EAPR on time. We used the logits of the EAPR because the transformed values were more linear and to make sure that the EAPR do not exceed a maximum value of 1. The trend line was estimated for each group defined by sex, type of residence and states/UT (140 lines for states/UT and 8 for India). For the first transition (EAPR to some primary) we used the data for those aged 15-39 years (5 data points). Using the trend line we extrapolated the EAPR into the future for our baseline scenario for each group (by sex/residence/states). We visually inspected each of the 148 graphs and found that some slopes were negative and few were too slow compared to the Indian average.

Therefore, in the second step, we decided to correct for the negative or slow growth by applying a convergence rule to those groups with speed (slope) less than the national slope (for the same sex and residence) to converge to the national value by 2051.

Again, we visually inspected all the lines and found that in few groups the predicted value for the next cohort in 2016, who were aged 10-14 in 2011, was less than the empirical EAPR of the same age group in 2011. Actually, 10-14 is the ultimate age by which the first transition would have taken place. However, for some groups the transition could occur during the age-group 10-14 as well. To correct for the early transition, as a third step, we first repeated the steps above for the age-group 10-34 and corrected the predicted values for the 'early transition' groups by replacing them with the new predicted values.

2.5.2 EAPR to primary

In Figure 9, the EAPR to completed primary (triangle shape in the figure) among those who went to school were the highest among all other EAPR. This shows that once a child gets into the school, the probability that the child completes education is very high. The transition values are slightly higher in urban areas compared to rural areas and no gender gap can be observed.

Between groups by states and residence (70 groups), the gender gap (females-males) among 15-19 year old's range from -7% (in rural Goa) to 9% (in rural Rajasthan, followed by 4% in rural Karnataka). The gap between rural and urban place of residence within states/UT is much larger with a range from -4% (among males in Uttar Pradesh) to 18% (among males in Chhattisgarh).

We applied the same method, as for EAPR to incomplete primary, to project the EAPR to completed primary using the data from the age-group 15-49. We also applied the same rule of convergence to those with slower slope than the national one to converge by 2051.

2.5.3 EAPR to lower secondary

The EAPR from completed primary to lower secondary has similar patterns in terms of the gender gap as for the EAPR to primary (i.e., the gap has closed). In terms of difference in the EAPR between rural and urban types of residence, the gap among 20-24 year old's is larger among females (13%) than males (8%).

Within groups (by states/residence), the gender gap among 20-24 year old's is quite a large range from -5% in rural Assam, followed by -4% in urban Uttar Pradesh, to 12% in Rural Sikkim. The residence gap (urban-rural) in EAPR to lower secondary is positive in the urban area (except among males in Uttar Pradesh, -2%). The largest gaps are observed almost exclusively among females by 22% in Madhya Pradesh, 21% in Chhattisgarh, Mizoram and West Bengal, and 20% in Karnataka. Among males, the worst states were Madhya Pradesh with 20%, Chhattisgarh with 19%, and Mizoram with 17%.

For the purpose of projections, we applied the same methods by using the data from age groups 20-49 (6 data points), and applied similar rules of convergence.

2.5.4 EAPR to upper secondary

The gender gap in the EAPR to upper secondary in the urban areas has become negative (see Figure 9, Panel 2), with more women (83.4%) than men (81.7%) making the transition to upper secondary among those with completed lower secondary. In rural areas the girls are speeding up to overtake boys in the nearest future. In the 27 mostly urban groups (by state and residence) the gender gap has reversed. The most extreme situation is in urban Uttar Pradesh where the EAPR to upper secondary is 77.4% for women compared to 65.4% among men. The highest range of the gender gap is from -12% to 9% in Kerala.

In India, the gap in EAPR to upper secondary between rural and urban region is still significant, 15% among females and 12% among males. Except among Uttar Pradesh males (-6%), the gap is positive with higher EAPR in urban areas than in rural areas of states/UT. The range is from 1% to 29% (among females in Delhi and West Bengal).

For the projections, we used data from the age group 20-40 (4 data points, including data for older ages that show a sudden jump). We applied the same methods as applied to other EAPR including the convergence rule. The range in the EAPR to upper secondary among 20-24 year old’s is from 41% (in rural Delhi) to 93% (in urban Himanchal Pradesh). Based on the currently observed maximum value, we allowed the future EAPR to become universal for all groups.

2.5.5 EAPR to tertiary

The final transition in our model is the EAPR to completed tertiary among high school graduates. While we did not impose any upper limits for the earlier four EAPR, which means eventually all cohorts will have at least upper secondary, we imposed an upper limit to the EAPR to tertiary of 50% for rural residents and 70% among urban residents.

In Figure 9, we observe that in urban India among 25-29 year old’s, 50% of the population with completed upper secondary further attain tertiary degree and recently the EAPR has become slightly higher for females than males. However, the EAPR is much lower in rural India with a small gender gap, 31% for males and 27% for females. The range in the EAPR to tertiary is very wide from 14% (among females in rural Uttar Pradesh) to 65% (among females in Urban Pondicherry). Females living in urban area are more likely to make the transition to tertiary. The highest value in rural area is in Maharashtra with 44% EAPR to tertiary.

Between the groups (by states/UT and residence), the range in the gender gap among 25-29 year old’s is quite large from -19% (in urban Manipur followed by mostly urban regions) to 11% (in rural Himanchal Pradesh followed by 10% in urban Uttarakhand, otherwise mostly rural regions). It shows a clear pattern that women residing in urban regions are more likely to complete tertiary than those living in rural regions.

The gap between urban and rural region (urban minus rural) in EAPR to tertiary is always positive except among males in Uttar Pradesh (-2%). The gap is very high among females, e.g., among females in Haryana the EAPR to tertiary value is 61% in urban areas and less than half (30%) in rural areas. Such situation is the reality in many states/UT. However, the gap among males is also significant in many states/UT, e.g., in Uttarakhand (26% in rural vs 51% in urban), Arunachal Pradesh (28% in urban vs 49% in rural) and so on.

For the projection, we applied the same method and the convergence rule by using the data for the age-group 25-49.

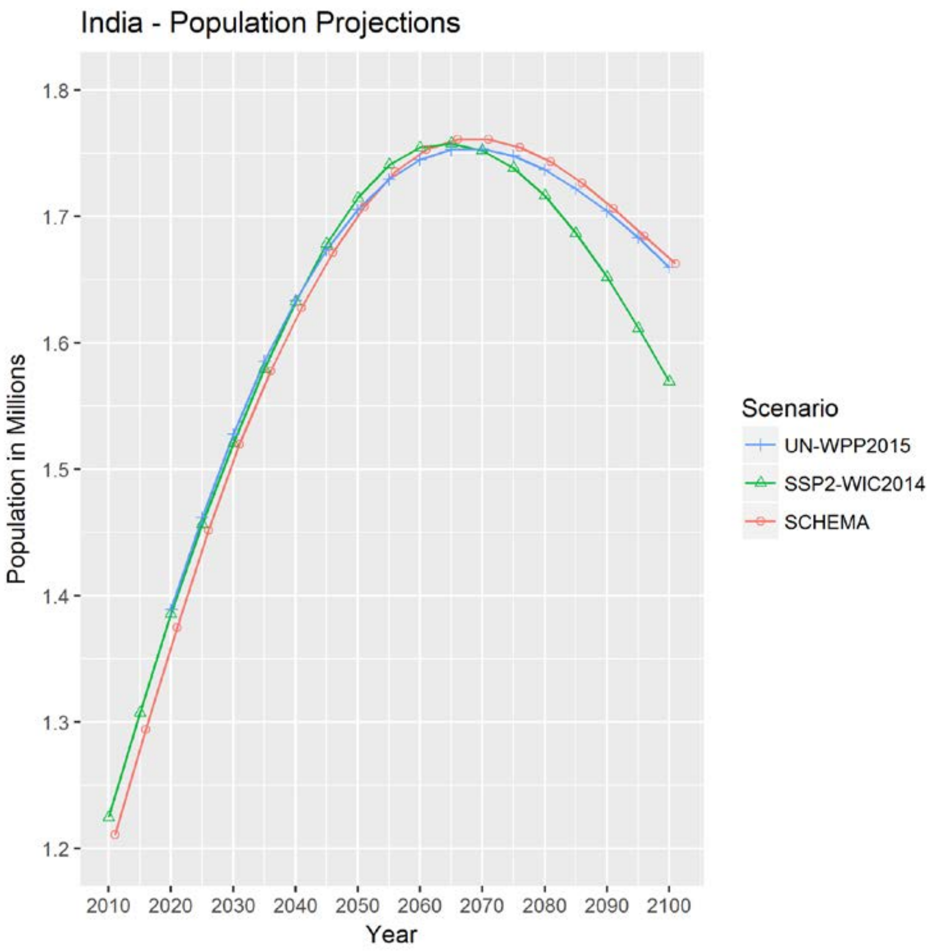
3 Results

In this paper for the purpose of illustrating our methodological and technical approach, we defined a single baseline scenario, where the assumptions are mostly based on the continuation of the current trend and authors opinions. In the future we plan to run sensitivity analyses on our key assumptions. We defined a baseline scenario for India and projected the population by age, sex, and educational attainment in each state/UT of India by rural and urban place of residence.

3.1 Total population

Our baseline scenario projection shows that the population of India will increase rapidly from 1.21 billion in 2011 to 1.71 billion by 2051, half a billion population in the course of 40 years, and will slowly peak at above 1.76 billion by 2071 (Figure 10). The population will then decline below 1.66 billion by the end of the century. While we were expecting our projections to be much different, it was surprising to find that they are very close to the UN 2015 assessment and WIC 2014 assessment until 2070, as our baseline scenario was developed independently from both and done by aggregating bottom-up projections of population. After 2070, while our projections continue to be very close to the UN projection, the WIC projections rapidly decline to almost 1.57 billion by 2100.

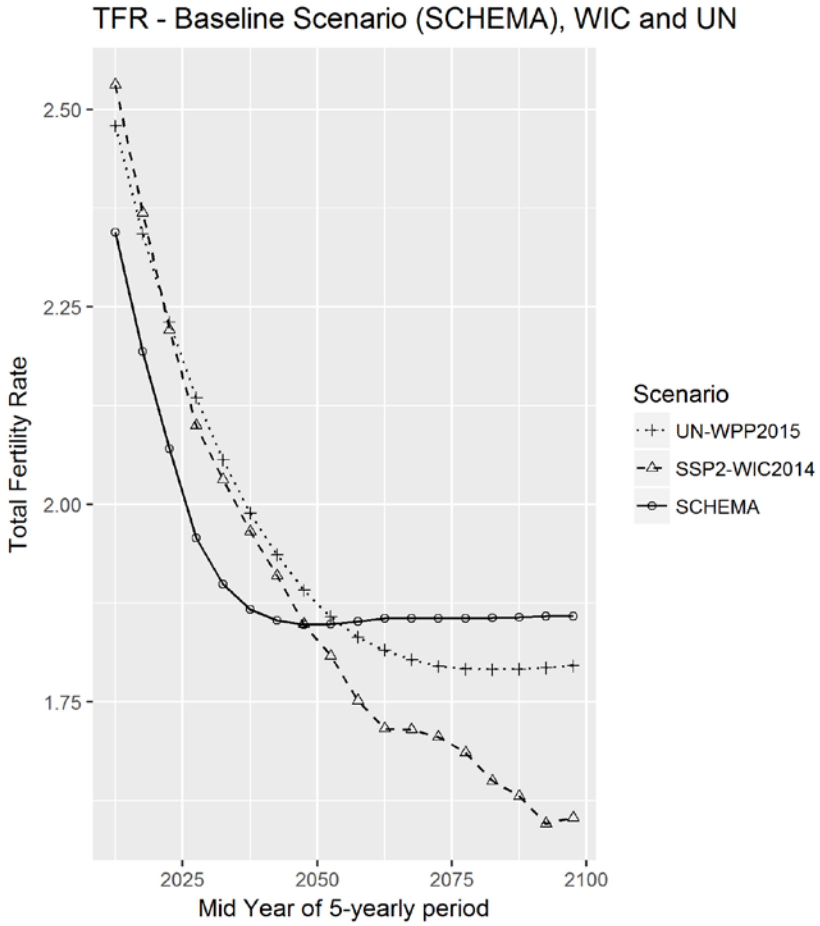
Figure 10. Population Projections for India, baseline scenarios (SCHEMA) along with UN medium variant (UN-WPP2015) and Wittgenstein Center projection (SSP2-WIC2014)



The difference in fertility assumptions between SCHEMA with the WIC and UN projections is large at the beginning of the projection period. While the UN and WIC have assumed total fertility for the whole of India, we calculated it by taking weighted averages of ASFR by states/UT, types of residence, and level of educational attainment among women. The UN assumed a TFR value of 2.48 children per woman for the period 2010-2015 and WIC assumed it to be 2.53 (see Figure 11). Our projection starts with the value

2.35 children per woman for the period 2011-2016 which is significantly lower than both the UN and WIC values. We have discussed the higher estimates of fertility by the UNPD and WIC in Section 2.1. This lower TFR value at the beginning should have resulted in lower population growth in our projections, by 65 million less births by 2050/51, which indicates that our assumption for mortality and international migration might have been different from that of the UN and WIC.

Figure 11. Total Fertility Rate in the baseline line scenario (SCHEMA), Wittgenstein Center (WIC) and United Nations (UN) medium projections



Our assumed life expectancies are higher than those of UN and WIC, postponing deaths to older ages and contributing to increases in population size. Comparing total number of deaths, UN calculated more than 7 million more deaths than resulted in our projections in the first period (47 million vs 40 million) and the difference remained higher for a long period with more than 70 million extra deaths in UN projections compared to ours by 2050. This shows that if we would have employed the UN or WIC mortality assumptions our projections would have resulted in a population of almost 72 million less.

Regarding international migration, UN and WIC assume negative net-migration for India with UN assuming -1.9 million during 2010-2015, which will slowly decline to less than -1.0 million by the end of the century. On the other hand, WIC assumptions are a bit higher (-2.3 million) for the period 2010-2015 that would decline to zero by the end of the century. This indicates that if we would include international migration in our projections following the UN and WIC, the gap would further widen by about 20 million by 2051 and more by the end of the century. Hence, combined effect of different mortality and migration assumptions between UN/WIC and ours would bring down our projection by 92 million by 2050/51.

3.2 Internal migration and urbanization

We observe urban population increased by more than 43 million during 2011-2016, which is less than half of (91 million) what was estimated for the period 2001-2011 (Pradhan, 2013). Of the total increase almost 50% is due to the natural increase (births-deaths), about 21% due to internal migration and the remaining 29% due to reclassification of the former rural areas into urban areas. The rate of increase in the size of urban population will decline in the future from 43 million during 2011-2016 to 26 million during 2046-2051, and will remain positive (half a million) until the end of the century. While the expected number of births in urban regions is stable at above 32 million per period, the number of deaths will increase dramatically from 10 million in 2011-2016 to 29 million during 2046-2051, and to 50 million during 2096-2101.

We find the rate of urbanization in the baseline scenario is slow. It will take another forty years for India to increase the share of urban population by 9 percentage points, from 31% in 2011 to 40% in 2051. UN, on the other hand, assumes a higher urbanization rate for India, reaching 50% urban population by 2050. This large difference in urbanization levels can be explained as follows: while the UN method extrapolates the urbanization trend at the highest level of aggregation, we apply the bottom-up approach making assumptions at more granular level. The UN approach is to apply the experience from other countries, which might not be appropriate for India due to its large and diverse population.

In our projections, the urbanization is affected by rural-urban migration assumptions and the reclassification rule. We have kept both rural-to-urban and urban-to-rural migration rates by age and sex constant. With a growing urban population, the number of population moving from urban to rural area will also grow. A five-yearly net flow from rural to urban will peak to about 9.3 million during 2016-2021 and will decline to around 7.5 million during 2046-2051 to 3.6 million by the end of the century.

By allowing the same rate of urban-to-rural migration, we are not implicitly speeding the process of sub-urbanization (i.e., people migrating to rural region but essentially commuting to work in non-agricultural jobs in urban areas). However, we do not capture this process properly and large portion of this sub-urbanization is becoming ‘rural’ population in our model. The future rural population will be quite different from the current one. While many urban-to-rural populations are contributing to the process of sub-urbanization, many will be contributing to the formation of new CT. In some areas of India, around big cities (e.g., Delhi, Bengaluru, etc.) we can already see this process unfolding. In our projections, we need also to include sub-urbanization not captured by the definition of CT employed in the Census 2011. Rural to urban (Census and ST) reclassification rates might increase in the future compared to what we have assumed in this paper based on migration data from the Census 2001 (the migration result from the Census 2011, once available, will provide more cues).

In addition to sub-urbanization and reclassification issues, the third reason for the low urbanization rate in our baseline scenario is the population weights carried by larger states, such as Uttar Pradesh and Bihar, with a very large proportion of rural less educated population. On top of it, the rural-to-urban migration rate among women is lower than among men, which essentially means that while men from rural areas go to work in other states, women (wives) are left behind to bear and rear the children as well as to take care of household, farm and the elderly. When these two states are excluded the speed of urbanization in the rest of the 33 states/UT is much faster from 35% in 2011 to 47% by 2051, compared to the whole of India (from 31% to 40%).

3.3 Births and fertility (population growth)

The number of births during a period of projection is a result of the assumed age-education-specific fertility rates and the size and age distribution of women in reproductive ages. While latter is already known at the beginning of the period, affected slightly by the assumptions of mortality and migration, the fertility assumption is crucial.

In our baseline scenario, we projected the education-specific fertility in each of the 70 sub-national units plus the overall rural and urban region of India by defining fertility pathways separately for rural and urban regions. The projected fertility rates in the population projections and the resulting overall TFR for India are shown in Figure 11, along with UN medium variant (United Nations, 2015) and WIC medium scenario (also called SSP2 scenario) (Lutz et al., 2014). A TFR of 2.35 children per woman during the period 2011-2016 will decline to below 2 children per woman already by 2025-31 and will remain at around 1.85 after 2041 for the rest of the century.

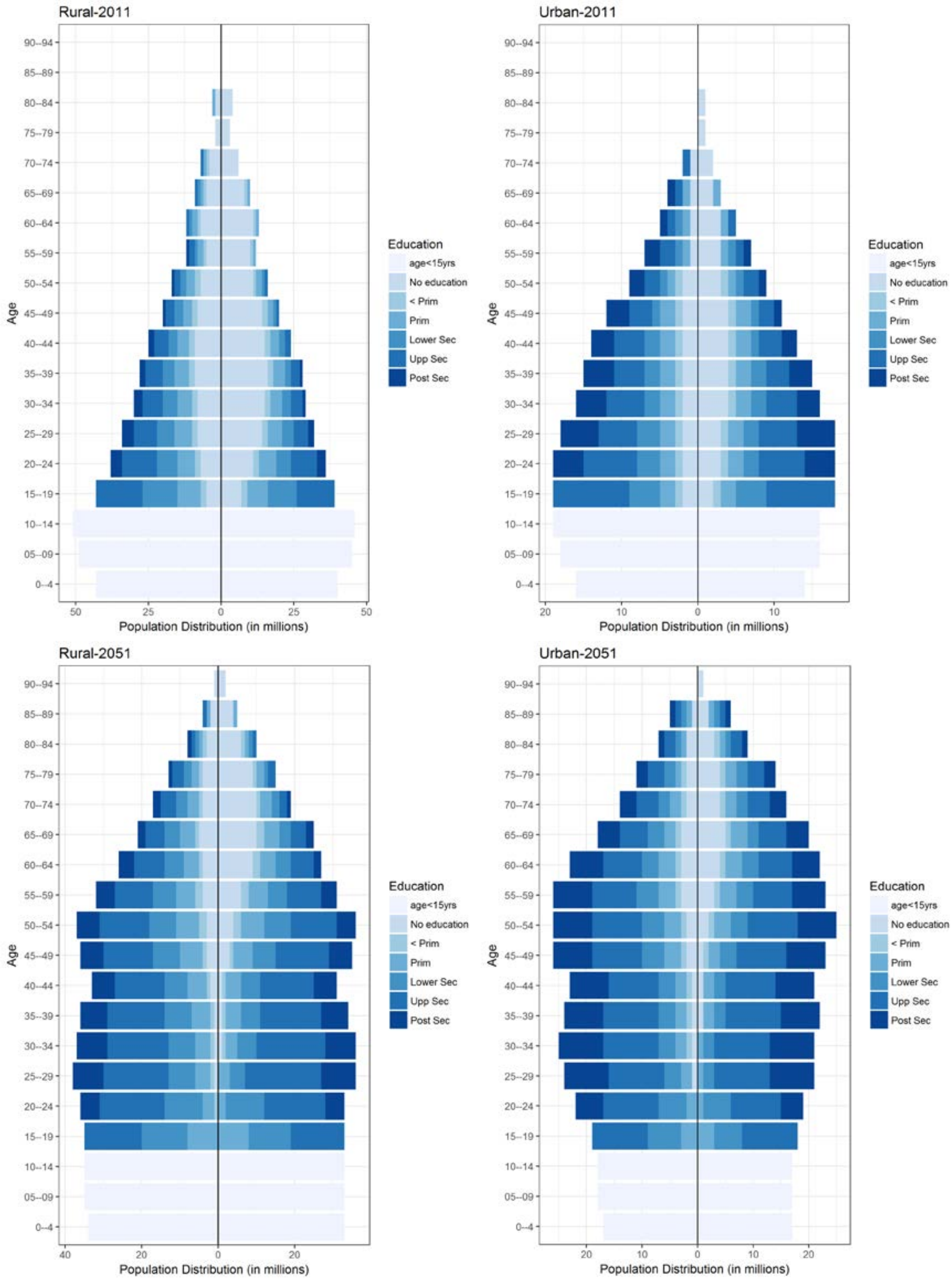
Compared to the UN and WIC assumptions, the TFR in our baseline scenarios stabilize at a higher level, at around 1.85 (Figure 11). We can see that with a slow increase in urbanization, the TFR of 1.85 lies between ultimate TFR levels in rural (2.08) and urban (1.75) areas and is sensitive to the future level of urbanization in India. While the levelling of fertility in the UN medium scenario is a median of thousands of random trajectories following experience of other countries, our ultimate level (around 1.85) is based purely on the past education specific fertility patterns among sub-populations within India. For a short run, our projections can be taken as a prediction that the TFR level in India will reach below replacement level, 2.1 children per woman, during 2021-26.

3.4 Educational attainment

Among the adult population aged 15-64, in 2011 about 47% have at least completed lower secondary education. Between states/UT, the inequality is very high and ranges from 35% in Bihar to 71% in Chandigarh. The larger states, mostly in northern India (Bihar, Madhya Pradesh, Uttar Pradesh, West Bengal and Gujarat) have low level of human capital. Most of the southern states have a higher level of human capital. The situation will slowly change as the process of replacing the older less educated cohorts with younger better educated cohorts will take some time. In the baseline scenario, in India the proportion will increase to 64% by 2031 and to 79% by 2051. Between provinces, the proportion ranges between 54-86% by 2031 to 70-94% by 2051.

The population pyramid of India for rural and urban regions in 2011 and 2051 (Figure 12) shows a demographic and human capital transformation in the next 40 years. Darker color in the pyramid represents better education. A larger proportion of rural population (more women than men) has never been to school in 2011 compared to those living in urban India. We expect the situation to change in the future (in 2051).

Figure 12. Population pyramid by educational attainment in Rural and Urban India in 2011 and 2051 (baseline Scenario)



3.5 Population dynamics at sub-national level

The next 20 years (2011-2031) can be considered a period of rapid population growth in India as the population will increase by 25% (309 million). 20.9% of all growth will occur in Uttar Pradesh (64.5 million), followed by Bihar (10.7%, 33.1 million), and Maharashtra (9.6%, 29.5 million). During the same period, all states/UT will experience population growth in the range of 11% in Kerala (33 million in 2011 to 37 million in 2031) to 100% in the tiny UT of Daman and Diu. Among the larger states (more than 1 million), Chandigarh (64%) and Delhi (56%) will see a very high growth rate. Both Uttar Pradesh and Bihar will experience a moderate growth rate of 32%.

Between 2031-2051 the two-decadal rate of population growth will halve to 12% but the absolute number is still high (188 million). Again the growth will occur mostly in Uttar Pradesh (20.6%), Bihar (12.1%), and Maharashtra (11.3%). During this period, Kerala will be the first and only state to see a decline in population. The population growth rate is declining in all states. By the end of the century, many states will see population decline but the states with major cities will continue to grow mainly due to migrants coming from other parts of India.

In terms of population share by state/UT, we found that it will remain unchanged (insignificant) for the rest of the century. This balance is caused by a combination of migration and fertility (and to some extend by mortality). Richer states with high urbanization level often have low fertility which is compensated by attracting migrants from poorer states that have a high proportion of people living in rural areas with higher fertility rates. This implies that if the calculation of the share is done based on place (state) of birth, the balance (unchanged proportion) will not be there anymore with increasing proportions of population born in poorer states and declining proportions of those born in richer states.

In 2011, the urbanization level within each state/UT ranges from 10% (in Himanchal Pradesh) to 98% in Delhi and 97% in Chandigarh. Among the larger states, Tamil Nadu and Kerala have almost 50% population living in urban areas (see Appendix, Table 1). By 2031, our assumptions of internal migration and reclassification rates increase the level of urbanization in almost all states except in Sikkim (from 25% to 23%) and Chandigarh (from 97% to 96%). The highest urban growth rate will occur in Kerala (from 48% to 70%) followed by Goa (62% to 78%). Maharashtra, West Bengal, Tamil Nadu, Punjab, and Karnataka will also experience moderate urban growth mainly because of big urban cities attracting migrants from other poorer states. Between 2011 and 2031 traditional (migrant) sending states will see a low increase in urbanization rates (Bihar with 1%, Uttar Pradesh with 2%, and Madhya Pradesh with 3%) due to high rural fertility combined with a low rate of reclassification and a high rate of migration to other richer states. This implies that the slower urbanization rate in India is based on the slow urbanization rates in these big states due to low rural to urban migration rates within states and assumed low reclassification rates. Therefore future urbanization rates of India would largely depend on what will happen (e.g., in terms of policy) in these states.

4 Conclusion

We developed a multi-regional multistate population projection model for India that can simultaneously simulate population heterogeneity in the demographic (age-sex), socioeconomic (educational attainment), and spatial (states/UT and rural/urban) dimensions. This is the first model of its kind for India. We populated the model with data and parameters and defined a baseline scenario based on the data analysis from the Census and SRS on five dimensions (fertility by education, mortality, internal migration, education, and rural reclassification).

It was surprising to see that our population projections for India with baseline scenario was consistent with the UN medium variant and WIC SSP2 until 2070. We found that while our fertility assumptions are lower, our mortality assumptions were also lower and compensated for the lower number of births (and no international migration) with higher number of survivors.

The results show that the overall fertility for India is lower than estimated/assumed by UN and WIC due to lower starting values in our projection as well as due to explicit consideration of education in the model. This results in a rapid TFR decline to about 1.85 children per woman in the next two decades and stabilization for the rest of the century.

The projection resulted in slower rate of urbanization in India from 31% in 2011 to 40% in 2051, compared to the UN urbanization projection and we presented several explanations for that. The most important reasons being the largely rural, less educated, large populations in Bihar and Uttar Pradesh (and other such states) that are slowing down the momentum of urbanization in India.

In terms of educational attainment, it will take some time before the adult population of India will attain universal basic education. While the younger cohorts are rapidly progressing to attain higher education in most regions of the country, there remain areas with big rural population posing challenges for India's human capital formation.

We plan further to conduct several sensitivity tests and define alternative scenarios with relevant narratives for India (e.g., SDG scenario). We are also working on developing SSP narratives at the sub-national level in India.

5 References

Barthelemy, Johan, and Thomas Suesse. 2016. *Mipfp: Multidimensional Iterative Proportional Fitting and Alternative Models. R Package Version 3.1.* R. https://CRAN.R-project.org/package=mipfp.

Basten, Stuart, Tomas Sobotka, and Krystof Zeman. 2014. "Future Fertility in Low Fertility Countries." In *World Population and Human Capital in the 21st Century*, edited by Wolfgang Lutz, William P. Butz, and Samir KC, 39–146. Oxford: Oxford University Press. http://ukcatalogue.oup.com/product/9780198703167.do.

Bhagat, R.B. 2011. "Emerging Pattern of Urbanisation in India." *Economic & Political Weekly* 46(34): p.10–12.

Census India. 2011. "Census of India 2011. Provisional Population Totals. Urban Agglomerations and Cities." Delhi, India: Office of the Registrar General &

Census Commissioner, India(RGI). http://censusindia.gov.in/2011-prov-results/paper2/data_files/India2/1.%20Data%20Highlight.pdf.

KC, Samir, Bilal Barakat, Anne Goujon, Vegard Skirbekk, Warren C. Sanderson, and Wolfgang Lutz. 2010. "Projection of Populations by Level of Educational Attainment, Age, and Sex for 120 Countries for 2005-2050." *Demographic Research* 22 (Article 15): 383–472. doi:10.4054/DemRes.2010.22.15.

Lutz, Wolfgang, William P. Butz, and Samir KC, eds. 2014. *World Population and Human Capital in the 21st Century*. Oxford, UK: Oxford University Press. http://ukcatalogue.oup.com/product/9780198703167.do.

Oeppen, Jim, and James W. Vaupel. 2017. "Broken Limits to Life Expectancy." *Science's Compass: Policy Forum*, p.3.

ORGI. 2006. "Population Projections for India and States 2001-2026. Report of the Technical Group on Population Projections Constituted by the National Commission on Population." Delhi, India: Office of the Registrar General and Census Commissioner of India. https://nrhm-mis.nic.in/Part%20B%20Demographic%20and%20Vital%20Indicators/Population%20Projection%20Report%202006%20by%20RGI.pdf.

———. 2014. "Sample Registration System Statistical Report 2013." REPORT NO. 1 OF 2014. Delhi, India. http://www.censusindia.gov.in/vital_statistics/SRS_Reports_2013.html.

———. 2017. "Sample Registration System." http://www.censusindia.gov.in/2011-Common/Sample_Registration_System.html.

Pradhan, Kanhu Charan. 2013. "Unacknowledged Urbanisation: New Census Towns of India." *Economic and Political Weekly* xlviii(36): p.43–51.

UNESCO. 2006. "International Standard Classification of Education: ISCED 1997 (Reprint)." Montreal, Canada: UNESCO Institute for Statistics. http://www.uis.unesco.org/Library/Documents/isced97-en.pdf.

United Nations. 2015. "World Population Prospects. The 2015 Revision. Volume 1: Comprehensive Tables." ST/ESA/SER.A/379. New York: United Nations Population Division. https://esa.un.org/unpd/wpp/Publications/Files/WPP2015_Volume-I_Comprehensive-Tables.pdf.

WIC. 2015. "Wittgenstein Centre Data Explorer Version 1.2." www.wittgensteincentre.org/dataexplorer.

Appendix

Table 1 Total population (in 1.000) and proportion urban population (in %) in India and states/UTs, 2011-2101 (authors calculations)

Region / state	Total population (in 1.000)				Proportion urban population (in %)			
	2011	2031	2051	2101	2011	2031	2051	2101
India	**1210827**	**1519555**	**1707699**	**1662514**	**31.1**	**36.0**	**39.9**	**47.1**
Andaman & Nicobar Is.	381	525	636	710	37.7	41.2	43.1	45.5
Andhra Pradesh	84573	97902	103718	90026	33.4	40.2	45.8	54.9
Arunachal Pradesh	1384	1946	2377	2478	22.9	28.6	32.0	37.3
Assam	31206	39501	44341	44840	14.1	17.0	19.8	25.2
Bihar	104098	137187	159987	163402	11.3	11.9	12.6	14.1
Chandigarh	1055	1732	2314	2915	97.3	96.0	96.6	97.7
Chhattisgarh	25545	32148	36152	36398	23.2	26.3	29.4	36.3
Dadra & Nagar Haveli	344	649	928	1212	46.7	60.7	68.7	78.5
Daman & Diu	243	486	722	963	75.2	76.4	82.3	89.5
Goa	1459	1896	2239	2383	62.2	77.5	84.9	90.0
Gujarat	60439	76643	86569	84949	42.6	48.9	54.7	65.6
Haryana	25351	34324	41663	47138	34.9	41.4	46.7	55.4
Himachal Pradesh	6865	8218	9073	8696	10.0	12.7	14.6	17.4
Jammu & Kashmir	12541	15054	16559	14105	27.4	30.9	34.3	42.6
Jharkhand	32988	42452	48637	48099	24.1	30.3	36.2	49.0
Karnataka	61095	72577	78641	72532	38.7	44.9	50.4	60.8
Kerala	33406	37024	36823	28316	47.7	69.9	80.7	89.0
Lakshadweep	64	93	116	127	78.1	80.9	82.3	84.9
Madhya Pradesh	72627	95203	110468	115683	27.6	30.3	33.3	40.6
Maharashtra	112373	141874	163212	167350	45.2	53.1	59.5	71.5
Manipur	2856	3185	3186	1905	29.2	37.7	45.1	59.6

Region / state	Total population (in 1.000)				Proportion urban population (in %)			
	2011	2031	2051	2101	2011	2031	2051	2101
Meghalaya	2967	3864	4487	4284	20.1	21.8	22.9	24.6
Mizoram	1097	1238	1281	881	52.1	61.7	68.6	77.9
Nagaland	1978	2342	2500	1769	28.9	34.2	38.8	50.8
Nct of Delhi	16788	26224	34967	45640	97.5	98.4	98.7	99.0
Odisha	41974	49487	53411	50839	16.7	21.0	24.8	31.8
Puducherry	1248	1675	1965	2093	68.3	74.9	79.7	86.5
Punjab	27743	33693	37896	37530	37.5	44.0	49.2	57.9
Rajasthan	68547	91330	106766	111488	24.9	28.1	31.1	37.2
Sikkim	611	786	892	805	25.2	22.7	21.4	19.9
Tamil Nadu	72147	81154	83443	70830	48.4	54.7	60.2	70.3
Tripura	3674	4260	4385	2796	26.2	36.7	44.8	56.9
Uttar Pradesh	199799	264261	303006	296957	22.3	24.3	26.2	31.1
Uttarakhand	10086	12920	15108	15819	30.2	37.1	42.8	52.5
West Bengal	91276	105704	109235	86555	31.9	39.5	45.8	54.7

Table 2 Proportion of women aged 20 to 39 years with at least lower secondary education (in %) and proportion of population aged 15 to 64 years with at least lower secondary education (in %) in India and states/UTs, 2011-2101 (authors calculations)

Region / state	Proportion of women aged 20 to 39 years with at least lower secondary education (in %)				Proportion of population aged 15 to 64 years with at least lower secondary education (in %)			
	2011	2031	2051	2101	2011	2031	2051	2101
India	**43.2**	**73.1**	**90.2**	**98.9**	**46.9**	**63.6**	**78.5**	**94.0**
Andaman & Nicobar Is.	66.6	85.1	94.9	99.6	61.6	74.1	84.6	95.9
Andhra Pradesh	36.9	75.1	91.7	99.3	41.2	61.7	81.2	96.3
Arunachal Pradesh	39.3	69.1	88.0	98.5	42.9	60.7	75.0	93.4
Assam	42.9	68.3	86.9	98.5	43.4	58.3	73.5	92.9
Bihar	25.7	62.8	86.5	98.4	34.7	53.8	71.6	92.7
Chandigarh	69.7	81.7	92.3	98.8	71.5	75.2	82.5	93.6
Chhattisgarh	37.1	71.9	89.4	98.7	41.5	61.6	78.2	93.9
Dadra & Nagar Haveli	42.1	72.2	90.1	98.7	48.9	65.3	78.1	93.4
Daman & Diu	57.7	77.7	91.6	99.1	60.0	68.3	77.8	93.4
Goa	68.7	84.5	93.3	98.9	64.9	75.8	84.6	94.9
Gujarat	40.8	67.9	87.8	98.5	45.2	60.9	75.9	93.0
Haryana	53.3	78.7	92.3	99.1	56.0	70.0	81.5	93.9
Himachal Pradesh	68.7	90.7	97.7	99.8	63.1	79.0	89.4	97.0
Jammu & Kashmir	48.9	81.9	94.9	99.6	55.1	73.0	86.0	96.9
Jharkhand	31.3	66.9	88.0	98.6	39.2	57.7	74.4	92.9
Karnataka	46.9	78.8	92.7	99.2	48.2	66.6	82.8	96.1
Kerala	82.9	93.4	97.8	99.8	70.3	86.0	94.0	98.6
Lakshadweep	64.5	90.5	97.5	99.9	59.4	78.9	90.5	97.8
Madhya Pradesh	33.5	69.4	88.7	98.5	40.3	59.2	76.2	92.8
Maharashtra	55.5	81.5	93.8	99.3	56.1	71.8	84.3	95.5
Manipur	62.1	80.4	92.3	98.8	62.2	74.2	83.4	95.4

Region / state	Proportion of women aged 20 to 39 years with at least lower secondary education (in %)				Proportion of population aged 15 to 64 years with at least lower secondary education (in %)			
	2011	2031	2051	2101	2011	2031	2051	2101
Meghalaya	38.7	65.3	87.0	98.5	37.3	54.1	69.7	91.9
Mizoram	58.5	76.7	90.8	98.9	55.1	68.4	78.4	93.8
Nagaland	52.1	71.8	89.1	98.8	51.5	63.9	74.9	93.5
Nct of Delhi	66.9	79.8	91.4	98.7	69.0	73.8	81.2	93.3
Odisha	39.7	73.2	91.8	99.5	40.9	60.7	78.1	94.7
Puducherry	72.5	92.1	98.1	99.9	68.8	82.1	92.8	98.7
Punjab	61.4	83.1	93.9	99.4	57.6	71.1	82.1	94.3
Rajasthan	27.4	64.6	87.6	98.5	38.9	57.6	74.5	92.7
Sikkim	47.0	66.3	86.7	98.5	44.2	57.0	70.6	92.6
Tamil Nadu	59.0	89.3	97.5	99.9	56.6	75.0	90.7	98.6
Tripura	44.7	69.8	88.1	98.5	46.4	59.7	73.3	93.1
Uttar Pradesh	37.1	72.5	89.6	98.7	45.6	63.1	77.8	93.2
Uttarakhand	58.3	84.1	94.9	99.5	59.6	74.3	84.9	95.0
West Bengal	39.1	64.7	86.6	98.5	41.9	54.1	70.1	92.7