



Future population and human capital in heterogeneous India

Samir KC^{a,b,1}, Marcus Wurzer^b, Markus Springer^b, and Wolfgang Lutz^{b,1}

^aAsian Demographic Research Institute, Shanghai University, Shanghai, 200444, China; and ^bWittgenstein Centre for Demography and Global Human Capital (IIASA, VID/OeAW, WU), International Institute for Applied Systems Analysis, Laxenburg, 2361, Austria

Contributed by Wolfgang Lutz, June 27, 2018 (sent for review January 31, 2018; reviewed by Josh Goldstein and K. S. James)

Within the next decade India is expected to surpass China as the world's most populous country due to still higher fertility and a younger population. Around 2025 each country will be home to around 1.5 billion people. India is demographically very heterogeneous with some rural illiterate populations still having more than four children on average while educated urban women have fewer than 1.5 children and with great differences between states. We show that the population outlook greatly depends on the degree to which this heterogeneity is explicitly incorporated into the population projection model used. The conventional projection model, considering only the age and sex structures of the population at the national level, results in a lower projected population than the same model applied at the level of states because over time the high-fertility states gain more weight, thus applying the higher rates to more people. The opposite outcome results from an explicit consideration of education differentials because over time the proportion of more educated women with lower fertility increases, thus leading to lower predicted growth than in the conventional model. To comprehensively address this issue, we develop a five-dimensional model of India's population by state, rural/urban place of residence, age, sex, and level of education and show the impacts of different degrees of aggregation. We also provide human capital scenarios for all Indian states that suggest that India will rapidly catch up with other more developed countries in Asia if the recent pace of education expansion is maintained.

India | population projections | human capital | subnational | heterogeneity

At the time of independence in 1947, India's total population was around 370 million and Indian women on average had six children. The age structure was very young, and over 80% of the population was illiterate (1). As a consequence, the population grew very rapidly, raising early concerns about the sufficiency of food supply and development prospects in general. Given these fears, in the late 1960s the Ford Foundation commissioned the "Second India" study to understand how India would fare under an expected doubling of its population (hence the name of the study) (2). In 1965 India's population was 500 million, and shortly before 2000 it reached the 1 billion mark. Revisiting the Second India around that time, Cassen found a rather mixed record. Some issues such as food production turned out to be better than feared, while others such as lack of education and poverty were worse than hoped (2). Both authors pointed at the great heterogeneity of the subcontinent, illustrated by the fertility rates in the early 1990s, which had already declined to 1.8 children per woman in Kerala but still stood at 5.1 in Uttar Pradesh (3).

The great heterogeneity of the Indian population is also the main focus of this paper. We will show how different ways of explicitly addressing heterogeneity in our demographic models will produce different outlooks for India's future population, human capital, and thus development. Fig. 1 shows the evolution of one century of India's population by level of education as observed since 1970 and forecast under a model described in this paper. It shows that in the 1970s still far more than half of the entire adult

population had never received any formal education and that this unfavorable situation has changed only very slowly. Still, by 1990, half of the adult population had never been to school.

Educational attainment of women has been much worse than that of men. Fig. 2 shows the age and education pyramids for 1970 and 2015. It shows that in 1970 about three quarters of Indian adult women had never been to school. Only a very tiny elite had the privilege of education. Among the younger cohorts, the proportion with at least primary education starts to slowly increase. For males, education levels are remarkably higher with only fewer than half of all adult men never having been to school. Because of higher fertility levels—during the 1960s Indian women had on average almost six children—the population age structure in 1970 was still extremely young. This very young age structure, together with only slow declines in birth rates, resulted in an increase of India's population from 554 million in 1970 to 1.3 billion in 2015. Today the younger cohorts are significantly better educated, but the legacy of low levels of female education is still visible in the low educational attainment of older cohorts, particularly of women. In association with the improving education of younger women, national-level fertility rates have also declined to 2.2, which is just around a third of their levels in the 1960s.

In this paper we will address the likely future population trends of India while systematically accounting for India's great population heterogeneity. Earlier projections of India that tried to go beyond conventional aggregate projections by age and sex

Significance

India will soon be the world's most populous country, but in terms of human capital and, consequently, Gross Domestic Product per capita, it has been trailing behind China. While some economists believe that India's younger population will be an advantage over China's aging one, here we show that much will depend on future investments in education and health and thus human capital. In terms of methodology, this paper addresses the question of what sources of observable population heterogeneity should be explicitly incorporated in population projections. It suggests that the dominant model of considering only the age and sex structures at the national level should be complemented by multidimensional models depending on the importance of heterogeneity and substantive user interest in the additional dimensions.

Author contributions: S.K. and W.L. designed research; S.K., M.W., and M.S. performed research; S.K., M.W., and M.S. analyzed data; and S.K. and W.L. wrote the paper.

Reviewers: J.G., University of California, Berkeley; and K.S.J., Institute for Social and Economic Change.

The authors declare no conflict of interest.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹To whom correspondence may be addressed. Email: kcsamir@gmail.com or lutz@iiasa.ac.at.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1722359115/-DCSupplemental.

Published online July 30, 2018.

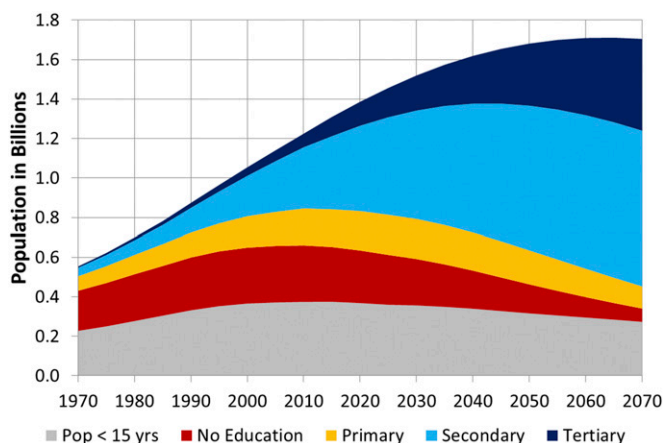


Fig. 1. India's total population 1970–2070 by level of education (23).

revealed an interesting phenomenon, namely that projections turn out to be significantly higher or lower depending on what additional sources of heterogeneity are taken into account. One study (4) showed that, if the projection is carried out at the level of India's 35 states, then the sum of state projections turns out to be significantly higher due to the fact that the high-fertility states over time receive more weight, and thus the higher fertility rates are applied to relatively more women. In contrast, projections that differentiated by level of educational attainment (at the national level) (5) resulted in lower forecasts because over time the younger, more educated cohorts of women entered the main reproductive ages, and since higher education is associated with lower fertility, this led to lower overall fertility.

These seemingly contradictory results, which depend on which source of evident population heterogeneity is included in the model, lead to the more general methodological debate in population forecasting and even more broadly for all social and economic forecasting models. What is the most appropriate way to account for the observable heterogeneity of agents in forecasts? While unobservable population heterogeneity also matters (6), the options to account for it are limited, a fact that suggests caution when interpreting results. Observed population heterogeneity, on the other hand, could readily be incorporated into multidimensional models, but there has been an interesting debate about whether this should always be done, most prominently in a set of papers in 1995 on the question of whether

simple models outperform complex ones (7). This discussion focused primarily on the question of whether forecasting total population size directly by applying assumed growth rates has given more accurate projections than the more complex cohort-component methods projecting individual age cohorts. In this context, Long (8) stresses that one needs to distinguish between two different questions: (i) whether one is only interested in the difference it makes for total population size forecasts and (ii) whether the additional dimension considered is of interest in its own right. We will add to this methodological discussion through an *ex ante* analysis of the sensitivity of Indian population forecasts to different sources of heterogeneity in the context of a multidimensional model, which, in addition to the conventional age and sex structure, also explicitly differentiates by level of educational attainment, urban/rural place of residence, and state of residence with differential fertility and mortality rates.

Heterogeneous India

India is a subcontinent that includes many population groups differing by language, ethnicity, religion, and caste (3). While some of this heterogeneity is stratified spatially and can be captured by differentiating between states and urban and rural areas, other factors (such as caste) exist in almost every location. Since statistical information tends to be collected along administrative boundaries, regional differentiation can be captured more easily from official aggregate statistical sources. Some of the other sources of heterogeneity can be derived only from individual-level data or more detailed cross-tabulations of census data. As has been argued earlier (9) and recently by Lutz and KC (10), the level of educational attainment and urban/rural place of residence are the two most important demographic dimensions of population heterogeneity after age and sex that cover relevant sociodemographic differentiations and should be used when measuring and modeling population dynamics. Following this approach, this study uses data that differentiate the populations of each of the 35 Indian states by all four dimensions (age, sex, level of education, and urban/rural place of residence).

The data used in this study come from detailed tabulations of the two most recent Indian censuses that were conducted in 2001 and 2011. These census tabulations were complemented with respect to vital rates by tabulations from the Sample Registration Survey (SRS) with annual information for the years 1999–2013. This allows us not only to study cross-sectional information, but also to analyze the trends over time since 1999. A more detailed specification of the data sources is given in *SI Appendix*.

Since fertility levels are the most important source of differential population growth, we studied the regional demographic

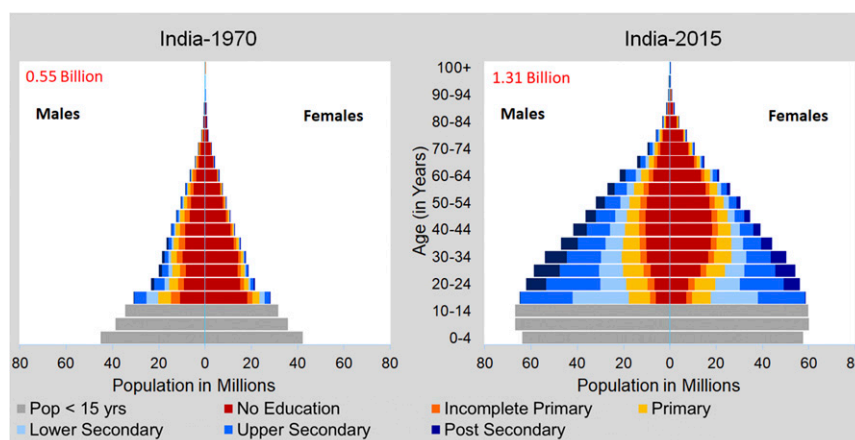


Fig. 2. Age and education pyramids for India (national level) for 1970 and 2015 (23).

population momentum and the only gradually increasing differences in rates, all hitting the 1.6 billion mark between 2036 and 2046 (Fig. 7). However, in terms of births the differences in the trajectories start earlier with 2.3% in 2011–2016 between the age-and-sex-only model and the age-sex-education model, increasing to around 7% for the next 25 y.

After 2040 the paths in total population size diverge, with India's population peaking at quite different levels and at different points in time. The red line in Fig. 7 gives the conventional national-level projection in which only the age and sex structures are considered. Here the population will peak at 1.71 billion in 2056–2066 and then enter a slow decline. Fig. 7 also gives the medium variant of the UN projections for India (thick broken line), which is based on an age- and sex-only model but assumes slightly lower fertility than our age- and sex-only model. This is why it results in somewhat lower projections that after 2070 are almost identical with our Indian education trend scenario. This also reminds us of the fact that the heterogeneity effect discussed in this paper is only one dimension of uncertainty while different assumptions on future fertility levels may even have bigger impacts on the outcomes.

As expected, the lowest projection comes from the model that considers only age, sex, and level of education, showing that the population will peak at 1.66 billion. If only the states are being considered and education is disregarded, the results peak at almost 1.8. The full model—considering all five dimensions—first (dark blue line) produces a lower trajectory than the conventional age-and-sex-only model (red line) due to a dominating education effect, but at 2061 the two lines cross and it climbs higher due to the state effect dominating. Finally, Fig. 7 also shows the line for the full scenario in which medium fertility, mortality, and migration assumptions are combined with the assumption of constant school enrollment rates. Because of the great momentum of changes in the educational composition by cohort, this results in the highest population growth only toward the end of the century. But, as Fig. 8 clearly shows, the two different education scenarios show quite different education distributions for the younger age groups in 2061. Since the Indian education expansion has not yet reached all parts of the population, cessation of further expansions would result in a sizable segment of the population with very low or no education.

Outlook and Conclusions

This study has provided insights with implications for the future of India as well as for the future of producing population projections

around the world. We have shown how different degrees of accounting for measurable heterogeneity within populations changes the way in which we see the future. No universally valid recommendation can be derived, and we suggest following Long's (8) pragmatic suggestion to include those dimensions that are informative for the users and for which an empirical basis exists. While age and sex are explicitly included by most producers of population forecasts, we have concluded that education should also be routinely included because it has well-established implications for fertility and mortality (10, 16), all methods and data are readily available, and the future educational attainment distributions are of great interest in their own right as indicators of a country's future human capital and development potential (17, 18).

Since independence, India has seen tremendous expansion in its population size, which has increased by a factor of 3.6 up to today. In the past, only elites were educated, with the majority of the population and in particular women never receiving any schooling. Still, in 1990, 70% of adult women had never attended any school, a proportion that subsequently has declined to 46% today. In parallel, the proportion of adult women with some tertiary education increased from 3 to 7%. Hence, recent years have seen a rapid improvement in education, and a look at younger cohorts shows that India is set for further rapid expansion. Among women aged 15–19 today, only 14% are without formal schooling, and already 65% have completed junior secondary or higher levels. Given the consistent evidence of the importance of broad-based education, benefits ranging from poverty eradication and economic growth to health and well-being to quality of institutions and even democracy (18–22) suggest likely future improvements in human development. But our analysis also shows that, if the education expansion should stall in the near future, some of this potential benefit might be lost.

Where does this leave us with respect to comparison of the world's two billion-plus populations? Because China has massively invested in universal education since the 1950s, it is about three to four decades ahead of India in terms of human capital. Actually, the education pyramid of India today looks similar to that of China around 1980. And the one projection given here for India in 2050 looks similar to that of China today. While cultural and institutional factors may differ between the two countries, and there can be no perfect analogy, this comparison makes it look likely that India will experience similarly rapid human-capital-driven development as China has over the past three to four decades.

- National Informatics Centre (2007) Economic Survey 2006–2007: 9.4 State-wise literacy rates (1951–2001) (National Informatics Centre, Ministry of Finance, New Delhi). Available at <https://www.indiabudget.gov.in/es2006-07/chapt2007/tab94.pdf>. Accessed December 7, 2017.
- Cassen R (1995) Review of the "Second India" revisited: Population, poverty, and environmental stress over two decades. *Popul Dev Rev* 21:163–170.
- James KS (2011) India's demographic change: Opportunities and challenges. *Science* 333:576–580.
- Haub C, Sharma OP (2006) *India's Population Reality: Reconciling Change and Tradition* (Population Reference Bureau, Washington, DC).
- Lutz W, Scherbov S (2004) Probabilistic population projections for India with explicit consideration of the education-fertility link. *Int Stat Rev* 72:81–92.
- Vaupel JW, Yashin AI (1985) Heterogeneity's ruses: Some surprising effects of selection on population dynamics. *Am Stat* 39:176–185.
- Rogers A (1995) Population forecasting: Do simple models outperform complex models? *Math Popul Stud* 5:187–202, 291.
- Long JF (1995) Complexity, accuracy, and utility of official population projections. *Math Popul Stud* 5:203–216.
- Lutz W, Goujon A, Doblhammer-Reiter G (1998) Demographic dimensions in forecasting: Adding education to age and sex. *Popul Dev Rev* 24:42–58.
- Lutz W, KC S (2010) Dimensions of global population projections: What do we know about future population trends and structures? *Philos Trans R Soc Lond B Biol Sci* 365:2779–2791.
- Bhat PM (2002) Returning a favor: Reciprocity between female education and fertility in India. *World Dev* 30:1791–1803.
- KC S, Springer M, Wurzer M (2017) 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 (International Institute for Applied Systems Analysis, Laxenburg, Austria). Available at pure.iiasa.ac.at/id/eprint/14516/1/WP-17-004.pdf.
- Sobotka T, Lutz W (2011) Misleading policy messages derived from the period TFR: Should we stop using it? *Comp Popul Stud-Z Für Bevölkerungswissenschaft* 35:637–664.
- ORGI (2014) Sample registration system statistical report 2013 (Office of Registrar General of India, New Delhi). Available at www.censusindia.gov.in/vital_statistics/SRS_Reports_2013.html. Accessed February 10, 2017.
- USAID (2006) Demographic and health survey India 2005/2006 (ICF International, Rockville, MD). Available at www.dhsprogram.com/Data/. Accessed June 4, 2016.
- Lutz W, Butz WP, KC S, eds (2014) *World Population and Human Capital in the Twenty-First Century* (Oxford Univ Press, Oxford).
- Lutz W (2017) Global sustainable development priorities 500 y after Luther: *Sola schola et sanitas*. *Proc Natl Acad Sci USA* 114:6904–6913.
- Lutz W, Cuaresma JC, Sanderson W (2008) Economics. The demography of educational attainment and economic growth. *Science* 319:1047–1048.
- Lutz W, Cuaresma JC, Abbasi-Shavazi MJ (2010) Demography, education, and democracy: Global trends and the case of Iran. *Popul Dev Rev* 36:253–281.
- Lutz W, Muttarak R (2017) Forecasting societies' adaptive capacities through a demographic metabolism model. *Nat Clim Chang* 7:177–184.
- Pamuk ER, Fuchs R, Lutz W (2011) Comparing relative effects of education and economic resources on infant mortality in developing countries. *Popul Dev Rev* 37:637–664.
- Cutler DM, Lleras-Muney A (2010) Understanding differences in health behaviors by education. *J Health Econ* 29:1–28.
- Wittgenstein Centre for Demography and Global Human Capital (2015) Wittgenstein Centre Data Explorer. Version 1.2. Available at www.wittgensteincentre.org/dataexplorer/. Accessed September 26, 2017.
- ORGI (2017) Sample Registration System. Available at www.censusindia.gov.in/2011-Common/Sample_Registration_System.html. Accessed December 22, 2017.