

Mapping disparities in education across low- and middle-income countries

<https://doi.org/10.1038/s41586-019-1872-1>

Local Burden of Disease Educational Attainment Collaborators*

Received: 9 November 2018

Accepted: 12 November 2019

Published online: 25 December 2019

Open access

Educational attainment is an important social determinant of maternal, newborn, and child health^{1–3}. As a tool for promoting gender equity, it has gained increasing traction in popular media, international aid strategies, and global agenda-setting^{4–6}. The global health agenda is increasingly focused on evidence of precision public health, which illustrates the subnational distribution of disease and illness^{7,8}; however, an agenda focused on future equity must integrate comparable evidence on the distribution of social determinants of health^{9–11}. Here we expand on the available precision SDG evidence by estimating the subnational distribution of educational attainment, including the proportions of individuals who have completed key levels of schooling, across all low- and middle-income countries from 2000 to 2017. Previous analyses have focused on geographical disparities in average attainment across Africa or for specific countries, but—to our knowledge—no analysis has examined the subnational proportions of individuals who completed specific levels of education across all low- and middle-income countries^{12–14}. By geolocating subnational data for more than 184 million person-years across 528 data sources, we precisely identify inequalities across geography as well as within populations.

Education, as a social determinant of health, is closely linked to several facets of the Sustainable Development Goals (SDGs) of the United Nations². In addition to the explicit focus of SDG 4 on educational attainment, improved gender equality (SDG 5) and maternal, newborn, and child health (SDG 3) have well-documented associations with increased schooling^{15–17}. In 2016, after years of deprioritization, aid to education reached its highest level since 2002¹⁸. Despite this shift, only 22% of aid to basic education—defined as primary and lower-secondary—went to low-income countries in 2016 compared to 36% in 2002¹⁹. This reflects a persistent pattern in which the distribution of aid does not align with the greatest need, even at the national level. Beyond international aid, domestic policy is also a crucial tool for expanding access to education, especially at higher levels. However, policy-makers often do not have access to a rigorous evidence base at a subnational level. This analysis presents the subnational distribution of education to support the growing evidence base of precision public health data, which shows widespread disparity of health outcomes as well as their social determinants.

Mapping education across gender

Despite widespread improvement in educational attainment since 2000, gender disparity persists in 2017 in many regions. Figure 1 illustrates the mean number of years of education and the proportion of individuals with no primary school attainment for men and women of reproductive age (15–49 years) in 2017. The average educational attainment is very low across much of the Sahel region of sub-Saharan Africa, consistent with previously published data¹⁴. In 2017, there was a large gender disparity in many regions, with men attaining higher average

education across central and western sub-Saharan Africa and South Asia. Considerable variation remains between the highest- and lowest-performing administrative units within countries in 2017. For Uganda in 2017, this indicator ranged from 1.9 years of education (95% uncertainty interval, 0.8–3.0 years) in rural Kotido to 11.1 years (10.1–12 years) in Kampala, the capital city. Figure 1b, d displays the proportion of men and women aged 15–49 years who have not completed primary school. By considering the variation within populations in different locations, these maps help to identify areas with large populations in the vulnerable lower end of the attainment distribution. We estimated large improvements in the proportions of individuals who have completed primary school in Mexico and China. However, across much of the world women in this age group failed to complete primary school at a much higher rate than their male counterparts.

Despite continued lack of gender parity in education among the reproductive age group, vast progress towards parity has been made among the 20–24 age group. Extended Data Fig. 2 further examines gender parity in 2000 and 2017. This figure highlights two additional advantages of our analytic framework. First, we examined a younger group aged 20–24 years. Although education in this group is less directly relevant to maternal, newborn, and child health than education in the full window of reproductive age, these estimates allowed us to capture how the landscape of education has shifted over time (that is, across successive cohorts) and is therefore more likely to pick up improvements to access and retention in education systems that have been made since 2000. Second, we illustrate the probability that this estimated ratio is credibly different from 1 (parity between sexes) given the full uncertainty in our data and model. In 2000, we estimated that men completed schooling at a higher rate than women across much

*A list of participants and their affiliations appears in the online version of the paper.

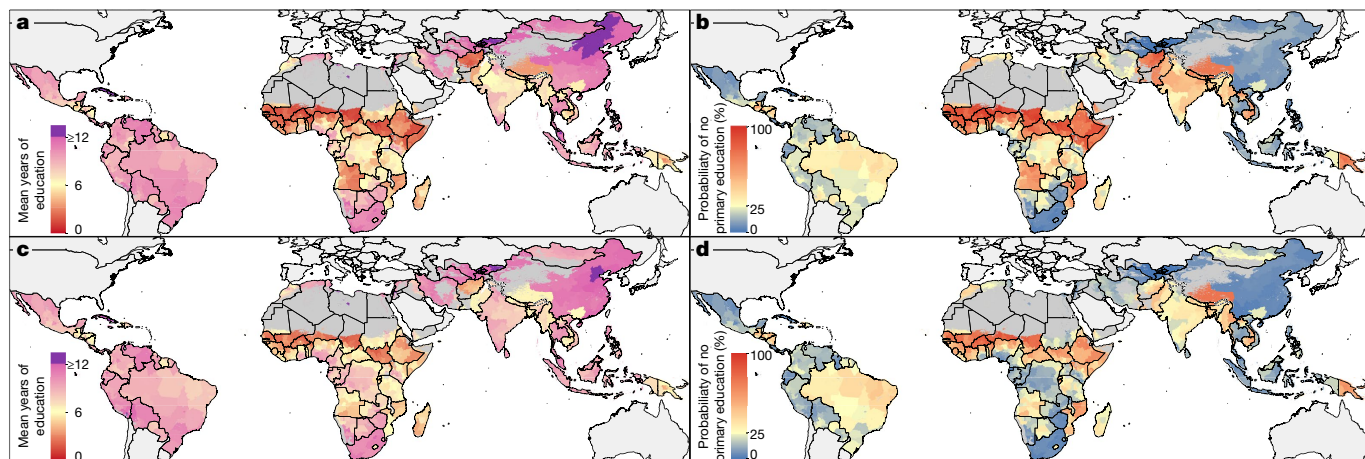


Fig. 1 | Average educational attainment and proportion of individuals with no completed primary education at the first administrative level and absolute difference between women and men aged 15–49 years. a–d, Mean

educational attainment for women (a) and men (c) and the proportion of individuals with no primary school education for women (b) and men (d) aged 15–49 years in 2017. Maps were produced using ArcGIS Desktop 10.6.

of the world, particularly for primary school education (that is, the probability that the parity ratio is greater than 1 was over 95%). This was true in most countries for both primary and secondary completion rates, but especially so in Burundi, Angola, Uganda, and Afghanistan

(Extended Data Fig. 2a, c). By 2017, many countries moved significantly towards parity in both secondary and primary completion rates with the exception of large regions within central and western sub-Saharan Africa (Extended Data Fig. 2b, d).

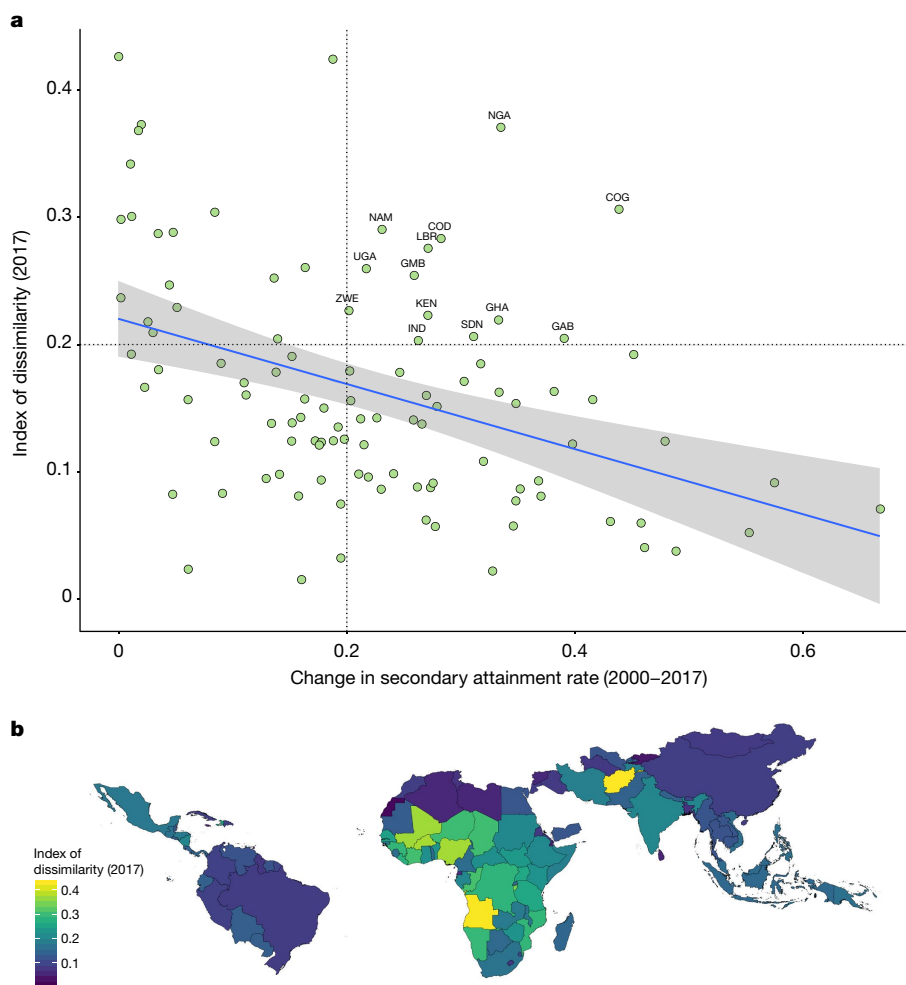


Fig. 2 | National progress in secondary attainment rates for women aged 20–24 years compared with the national index of dissimilarity in 2017. a, Change in secondary attainment rates for women age 20–24 years between 2000 and

2017 compared with the national index of dissimilarity in 2017 (simple linear regression lines are included). b, Map of the national index of dissimilarity in 2017. Maps were produced using ArcGIS Desktop 10.6.

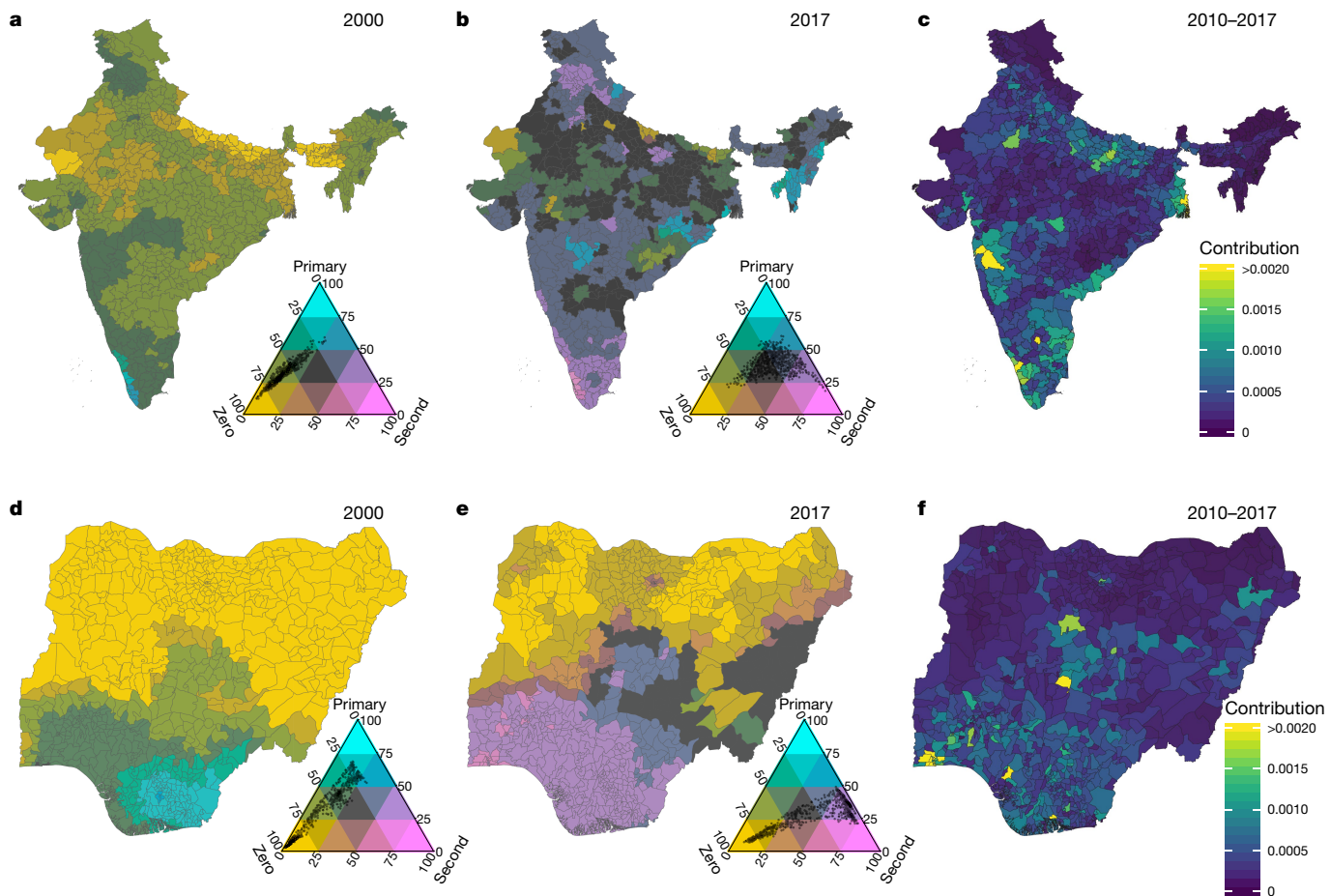


Fig. 3 | Attainment rates and contributions to national change in secondary rates for women aged 20–24 years in India and Nigeria, 2000–2017. **a, b**, Attainment rates for women aged 20–24 years in 2000 (**a**) and 2017 (**b**) at the second administrative level in India. **c**, Additive contributions of changes in the attainment rates at the second administrative level to change in the rate at the national level between 2000 and 2017 in India. **d, e**, Attainment rates for

women aged 20–24 years in 2000 and 2017 at the second administrative level in Nigeria. **f**, Additive contributions of changes in the attainment rates at the second administrative level to change in the rate at the national level between 2000 and 2017 in Nigeria. On all ternary maps, the ‘Zero’ category includes all individuals with either no schooling or some primary schooling without completion. Maps were produced using ArcGIS Desktop 10.6.

Inequalities within and between countries

The subnational estimates of attainment presented here enable a closer examination of within-country inequality and associated trends over time. Figure 2 plots the national change in secondary attainment rates for women aged 20–24 years with the index of dissimilarity across second administrative-level units in 2017. The index of dissimilarity is an intuitive measure of geographical inequality that can be interpreted as the percentage of women with secondary attainment that would have to move in order to equalize secondary rates across all subnational districts. We estimated that countries that experienced more national progress over the period tended to be more spatially equal in 2017. However, the top-right quadrant of the graph highlights several countries that experienced substantial national progress yet remain some of the most geographically unequal countries today.

We further examined national progress between 2000 and 2017 in two such countries, India and Nigeria, where rates of secondary attainment increased from 10.9% (8.5–12.5%) to 37.2% (33.6–41.1%) and from 11.5% (6.2–18.3%) to 45.0% (37.0–52.5%), respectively (Fig. 3). The geographical distribution between two cohorts—women aged 20–24 years in 2000 and 2017—was analysed by examining all proportions simultaneously (Fig. 3a, b). We estimate that there has been a massive shift towards primary and secondary completion coupled with greater geographical variability in completion rates (that is, spread of the dots that represent subnational units in the legend). The majority of the 2017 cohort living

in the northwest and northeast of India never completed secondary school. Urban centres in the south, such as Bangalore and Mumbai, have seen considerable progress compared with more rural regions. In Nigeria, we estimate substantial national improvement; however, the country remained one of the most spatially unequal in 2017 (Fig. 3d, e). The more-urban south, particularly around Lagos, experienced much faster progress than the more-rural north. The implications of the population distribution were explored by decomposing the improvement in the national rate of secondary completion since 2000 for each country into the additive contributions of rate changes at the second administrative level (Fig. 3c, f). This demonstrates that national progress was largely driven by improvements in populous urban regions (particularly Maharashtra, India, and Lagos, Nigeria), underscoring the importance of how subnational progress (or lack thereof) contributes differentially to narratives surrounding national change.

Discussion and limitations

We have built on previous modelling efforts that focused on the geographical distribution of average education¹⁴ by extending our estimation to the distribution of attainment, highlighting not only average attainment but also the proportions of individuals who completed key levels of schooling that are central to policy efforts. As we demonstrate, throughout much of the world women lag behind their male

counterparts, and there is significant heterogeneity across subnational regions. Countries such as South Africa, Peru, and Colombia have seen tremendous improvement since 2000 in the proportion of the young adult population who have completed secondary school. As this trend continues, it will be important to focus not only on attainment but also on quality of education. However, many young women across the world still faced obstacles to attaining even a basic level of education in 2017 (Extended Data Fig. 3). This represents a missed opportunity for the global health community to focus on a well-studied determinant of maternal, newborn, and child health. Even with only marginal returns to health in the short term, studies suggest that, on average, communities will also see increased human capital, social mobility, and less engagement in child marriage or early childbearing^{20,21}.

Children and adolescents do not complete formal schooling for many reasons. Many factors differentially affect girls, such as cost, late or no school enrolment, forced withdrawal of married adolescents, and the social influence of family members concerning the traditional roles of girls and women^{4,20,22,23}. A critical step is acknowledging that commercialization in the area of education typically leads to higher inequity²⁴. Treating public education as a societal good by increasing access, particularly in underserved rural communities, reduces inequality. Identifying areas that are stagnating or worsening, particularly in the realm of basic education for young women across the world, is an important first step to targeted, long-term reform efforts that will ultimately have widespread benefits for equity in health and development.

Many recent international calls to improve the social determinants of health have stated that measurement of inequity within countries is critical to understanding and tracking the problem, noting that geography is an increasingly important dimension of inequity^{24–26}. Where people are born greatly determines their life chances, and continuing to consider development and human capital formation on a national level is insufficient²⁴. The goal of this analysis is to identify local areas that may have experienced negligible improvements, but further rigorous research is required to contextualize these patterns within the unique mix of structural obstacles that each community faces. There are many indirect costs for attending school and each disadvantaged area that we identify in our analysis may experience them in different ways. These include the demand for children to work, the opportunity or monetary costs of attending school, distance to school, lack of compulsory education requirements, high fees for attendance, political instability, and many other forces. Overcoming these obstacles to improve educational attainment alone will not necessarily result in a more-educated and healthy population for each country as highly educated individuals may be more likely to emigrate, resulting in ‘brain drain’. This is especially true for countries that have been economically crippled over the past two decades and may lack the economic capacity to absorb a more highly educated labour force. Opening access to education will need to be coupled with economic reforms, both internationally and domestically, if countries are to fully experience dividends in human capital and health.

Over the next decade of the SDG agenda, it will be important to maintain the progress that has been made to reprioritise investment in education systems. There remains an alarming lack of distributional accountability in aid, especially to basic education, for which most funding is not going to the countries that need it most¹⁹. Connections between educational attainment and health offer promising opportunities for co-financing initiatives. For example, USAID recently invested US\$90 million in HIV funding to the construction of secondary schools in sub-Saharan Africa. Global health leaders have noted the need to invest in precise data systems and eliminate data gaps to effectively target resources, develop equitable policy, and track accountability⁷. Our analysis provides a robust evidence base for such decision-making and advocacy. Decades of research on the effect of basic education on maternal, newborn, and child health positions this issue squarely in the purview of the global health agenda. It is crucial for the global health community to invest in long-term, sustainable improvement in

the underlying distribution of human capital, as this is the only way to truly influence health equity across generations.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1872-1>.

1. UNESCO. Meeting our commitments to gender equality in education. *Global Education Monitoring Report*. <https://unesdoc.unesco.org/ark:/48223/pf0000261593> (2018).
2. United Nations. *Transforming our World: the 2030 Agenda for Sustainable Development* (UN, 2015).
3. Lim, S. S. et al. Measuring human capital: a systematic analysis of 195 countries and territories, 1990–2016. *Lancet* **392**, 1217–1234 (2018).
4. Yousafzai, M. & Lamb, C. *I am Malala: The Girl Who Stood Up for Education and Was Shot by The Taliban* (Weidenfeld & Nicolson, 2013).
5. Gates, M. *The Moment of Lift: How Empowering Women Changes The World* (Flatiron Books, 2019).
6. United Nations. *Youth and the 2030 Agenda for Sustainable Development* (UN, 2018).
7. Annan, K. Data can help to end malnutrition across Africa. *Nature* **555**, 7 (2018).
8. Horton, R. Offline: in defence of precision public health. *Lancet* **392**, 1504 (2018).
9. Dowell, S. F., Blazes, D. & Desmond-Hellmann, S. Four steps to precision public health. *Nature* **540**, 189–191 (2016).
10. Osgood-Zimmerman, A. et al. Mapping child growth failure in Africa between 2000 and 2015. *Nature* **555**, 41–47 (2018).
11. Golding, N. et al. Mapping under-5 and neonatal mortality in Africa, 2000–15: a baseline analysis for the Sustainable Development Goals. *Lancet* **390**, 2171–2182 (2017).
12. Bosco, C. et al. Exploring the high-resolution mapping of gender-disaggregated development indicators. *J. R. Soc. Interface* **14**, 20160825 (2017).
13. Roberts, D. A. et al. Benchmarking health system performance across regions in Uganda: a systematic analysis of levels and trends in key maternal and child health interventions, 1990–2011. *BMC Med.* **13**, 285 (2015).
14. Graetz, N. et al. Mapping local variation in educational attainment across Africa. *Nature* **555**, 48–53 (2018).
15. Caldwell, J. C. How is greater maternal education translated into lower child mortality? *Health Transit. Rev.* **4**, 224–229 (1994).
16. Caldwell, J. C. Education as a factor in mortality decline: an examination of Nigerian data. *Popul. Stud.* **33**, 395–413 (1979).
17. Gakidou, E., Cowling, K., Lozano, R. & Murray, C. J. Increased educational attainment and its effect on child mortality in 175 countries between 1970 and 2009: a systematic analysis. *Lancet* **376**, 959–974 (2010).
18. UNESCO. *UNESCO Operational Definition Of Basic Education. Thematic Framework* (UNESCO, 2007).
19. UNESCO. *Aid to Education: A Return to Growth?* (UNESCO, 2018).
20. LeVine, R. A., LeVine, S., Schnell-Anzola, B., Rowe, M. L. & Dexter, E. *Literacy and Mothering: How Women's Schooling Changes the Lives of the World's Children* (Oxford Univ. Press, 2012).
21. Abel, G. J., Barakat, B., Kc, S. & Lutz, W. Meeting the Sustainable Development Goals leads to lower world population growth. *Proc. Natl Acad. Sci. USA* **113**, 14294–14299 (2016).
22. UNESCO. *Reducing Global Poverty Through Universal Primary and Secondary Education. Out-Of-School Children, Adolescents And Youth: Global Status And Trends Policy Paper 32/Fact Sheet 44* (2017).
23. Jejeebhoy, S. J. *Women's Education, Autonomy, and Reproductive Behaviour: Experience From Developing Countries* (Clarendon, 1995).
24. Marmot, M., Friel, S., Bell, R., Houweling, T. A. & Taylor, S. Closing the gap in a generation: health equity through action on the social determinants of health. *Lancet* **372**, 1661–1669 (2008).
25. Kim, J. Y. *World Bank Group President Jim Yong Kim Speech at the 2017 Annual Meetings Plenary*. <http://www.worldbank.org/en/news/speech/2017/10/13/wbg-president-jim-yong-kim-speech-2017-annual-meetings-plenary-session> (13 October 2017).
26. UNESCO. *Is real progress being made in the equitable provision of education? #PISAresults* <http://www.iiep.unesco.org/en/real-progress-being-made-equitable-provision-education-pisaresults-3915> (UNESCO, 2017).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

Article

Board, Roberto Santos General Hospital, Salvador, Brazil. ⁷⁶Department of Health Metrics Sciences, School of Medicine, University of Washington, Seattle, WA, USA. ⁷⁷Epidemiology Department, Florida International University, Miami, FL, USA. ⁷⁸Department of Public Health Sciences, Karolinska Institutet, Stockholm, Sweden. ⁷⁹World Health Programme, Université du Québec en Abitibi-Témiscamingue, Rouyn-Noranda, Quebec, Canada. ⁸⁰Center of Communicable Disease Control, Ministry of Health and Medical Education, Tehran, Iran. ⁸¹School of Public Health, Arak University of Medical Sciences, Arak, Iran. ⁸²Babol University of Medical Sciences, Babol, Iran. ⁸³College of Medicine, Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia. ⁸⁴Department of Psychology, Federal University of Sergipe, Sao Cristovao, Brazil. ⁸⁵Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden. ⁸⁶Division of Neurology, University of Ottawa, Ottawa, Ontario, Canada. ⁸⁷REQUIMTE/LAQUV, University of Porto, Porto, Portugal. ⁸⁸Psychiatry Department, Kaiser Permanente, Fontana, CA, USA. ⁸⁹Department of Health Sciences, A. T. Still University, Mesa, AZ, USA. ⁹⁰Department of Population Medicine and Health Services Research, Bielefeld University, Bielefeld, Germany. ⁹¹Department of Dermatology, Kobe University, Kobe, Japan. ⁹²Gene Expression & Regulation Program, The Wistar Institute, Philadelphia, PA, USA. ⁹³Ramón de la Fuente Muñiz National Institute of Psychiatry, Mexico City, Mexico. ⁹⁴Unit of Academic Primary Care, University of Warwick, Coventry, UK. ⁹⁵Adelaide Medical School, University of Adelaide, Adelaide, South Australia, Australia. ⁹⁶Nursing and Health Sciences Department, University of Massachusetts Boston, Boston, MA, USA. ⁹⁷Department of Biostatistics and Epidemiology, University of Oklahoma, Oklahoma City, OK, USA. ⁹⁸Department of Health and Social Affairs, Government of the Federated States of Micronesia, Palikir, Federated States of Micronesia. ⁹⁹School of Medicine, Boston University, Boston, MA, USA. ¹⁰⁰School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia. ¹⁰¹Department of Epidemiology and Biostatistics, Zhengzhou University, Zhengzhou, China. ¹⁰²Academics and Research Department, Rajasthan University of Health Sciences, Jaipur, India. ¹⁰³Department of Medicine, Mahatma Gandhi University of Medical Sciences & Technology, Jaipur, India. ¹⁰⁴Department of Anthropology, University of Delhi, Delhi, India. ¹⁰⁵Department of Pharmacology, Tehran University of Medical Sciences, Tehran, Iran. ¹⁰⁶Obesity Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. ¹⁰⁷Department of Radiology, Johns Hopkins University, Baltimore, MD, USA. ¹⁰⁸Department of Family and Community Medicine, Arabian Gulf University, Manama, Bahrain. ¹⁰⁹School of Health and Environmental Studies, Hamdan Bin Mohammed Smart University, Dubai, United Arab Emirates. ¹¹⁰Department of Public Health, Mizan-Tepi University, Tepi, Ethiopia. ¹¹¹Unit of Epidemiology and Social Medicine, University Hospital Antwerp, Antwerp, Belgium. ¹¹²School of Public Health, Curtin University, Perth, Western Australia, Australia. ¹¹³Department of Pediatrics, University of Texas Austin, Austin, TX, USA. ¹¹⁴Department of Pharmacology and Therapeutics, Dhaka Medical College, Dhaka, Bangladesh. ¹¹⁵Department of Pharmacology, Bangladesh Industrial Gases Limited, Tangail, Bangladesh. ¹¹⁶Department of Computer Engineering, Islamic Azad University, Tehran, Iran. ¹¹⁷Computer Science Department, University of Human Development, Sulaimaniyah, Iraq. ¹¹⁸Department of Epidemiology and Health Statistics, Central South University, Changsha, China. ¹¹⁹Department of Community Medicine, University of Ibadan, Ibadan, Nigeria. ¹²⁰Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. ¹²¹Institute for Physical Activity and Nutrition, Deakin University, Burwood, Victoria, Australia. ¹²²Sydney Medical School, University of Sydney, Sydney, New South Wales, Australia. ¹²³Department of Epidemiology, Shahid Beheshti University of Medical Sciences, Tehran, Iran. ¹²⁴Department of Health Care and Public Health, Sechenov First Moscow State Medical University, Moscow, Russia. ¹²⁵Department of Community Medicine, Banaras Hindu University, Varanasi, India. ¹²⁶Environmental Research Center, Duke Kunshan University, Kunshan, China. ¹²⁷Nicholas School of the Environment, Duke University, Durham, NC, USA. ¹²⁸Department of Ophthalmology, Heidelberg University, Heidelberg, Germany. ¹²⁹Beijing Institute of Ophthalmology, Beijing Tongren Hospital, Beijing, China. ¹³⁰Social Determinants of Health Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran. ¹³¹Department of Family Medicine and Public Health, University of Opole, Opole, Poland. ¹³²Department of Forensic Medicine and Toxicology, All India Institute of Medical Sciences, Jodhpur, India. ¹³³Hematology-Oncology and Stem Cell Transplantation Research Center, Tehran University of Medical Sciences, Tehran, Iran. ¹³⁴Pars Advanced and Minimally Invasive Medical Manners Research Center, Iran University of Medical Sciences, Tehran, Iran. ¹³⁵Research Center for Environmental Determinants of Health, Kermanshah University of Medical Sciences, Kermanshah, Iran. ¹³⁶ODEL Campus, University of Nairobi, Nairobi, Kenya. ¹³⁷CSIR-Indian Institute of Toxicology Research, Council of Scientific & Industrial Research, Lucknow, India. ¹³⁸Department of Public Health, Jordan University of Science and Technology, Irbid, Jordan. ¹³⁹Social Determinants of Health Research Center, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran. ¹⁴⁰Epidemiology and Biostatistics Department, Health Services Academy, Islamabad, Pakistan. ¹⁴¹Department of Medical Parasitology, Cairo University, Cairo, Egypt. ¹⁴²Clinical Epidemiology Unit, Lund University, Lund, Sweden. ¹⁴³Research and Data Solutions, Synotech Consultants, Nairobi, Kenya. ¹⁴⁴School of Medicine, Xiamen University Malaysia, Sepang, Malaysia. ¹⁴⁵Department of Nutrition, Simmons University, Boston, MA, USA. ¹⁴⁶School of Health Sciences, Kristiania University College, Oslo, Norway. ¹⁴⁷Independent Consultant, Jakarta, Indonesia. ¹⁴⁸CIBERSAM, San Juan de Dios Sanitary Park, Sant Boi De Llobregat, Spain. ¹⁴⁹Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain. ¹⁵⁰Department of Anthropology, Panjab University, Chandigarh, India. ¹⁵¹Department of Social and Preventive Medicine, University of Montreal, Montreal, Quebec, Canada. ¹⁵²Department of Demography, University of Montreal, Montreal, Quebec, Canada. ¹⁵³Department of Psychiatry, University of Nairobi, Nairobi, Kenya. ¹⁵⁴Division of Psychology and Language Sciences, University College London, London, UK. ¹⁵⁵International Institute for Population Sciences, Mumbai, India. ¹⁵⁶Department of Community and Family Medicine, University of Baghdad, Baghdad, Iraq. ¹⁵⁷School of Nursing, Hong Kong Polytechnic University, Hong Kong, China. ¹⁵⁸Department of Medical Statistics and Epidemiology, Sun Yat-sen University, Guangzhou, China. ¹⁵⁹Alliance for Improving Health Outcomes Inc, Quezon City, The Philippines. ¹⁶⁰Department of Medicine, University of Malaya, Kuala Lumpur, Malaysia. ¹⁶¹Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Shatin, China. ¹⁶²Department of Dentistry, Radboud University, Nijmegen, The Netherlands. ¹⁶³Section for Translational Health Economics, Heidelberg University Hospital, Heidelberg, Germany. ¹⁶⁴Department of Epidemiology and Biostatistics, University of the Philippines Manila, Manila, The Philippines. ¹⁶⁵Department of Public Health, Trnava University, Trnava, Slovakia. ¹⁶⁶Community-Based Participatory-Research Center (CBPR), Tehran University of Medical Sciences, Tehran, Iran. ¹⁶⁷Knowledge Utilization Research Center (KURC), Tehran University of Medical Sciences, Tehran, Iran. ¹⁶⁸Department of Primary Care and Public Health, Imperial College London, London, UK. ¹⁶⁹Digestive Diseases Research Institute, Tehran University of Medical Sciences, Tehran, Iran. ¹⁷⁰Non-communicable Diseases Research Center, Shiraz University of Medical Sciences, Shiraz, Iran. ¹⁷¹Department of Epidemiology and Biostatistics, Tehran University of Medical Sciences, Tehran, Iran. ¹⁷²Campus Caucaia, Federal Institute of Education, Science and Technology of Ceará, Caucaia, Brazil. ¹⁷³Public Health Department, Botho University-Botswana, Gaborone, Botswana. ¹⁷⁴Division of Plastic Surgery, University of Washington, Seattle, WA, USA. ¹⁷⁵Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, USA. ¹⁷⁶Peru Country Office, United Nations Population Fund (UNFPA), Lima, Peru. ¹⁷⁷Center for Translation Research and Implementation Science, National Institutes of Health, Bethesda, MD, USA. ¹⁷⁸Department of Medicine, University of Cape Town, Cape Town, South Africa. ¹⁷⁹Breast Surgery Unit, Helsinki University Hospital, Helsinki, Finland. ¹⁸⁰University of Helsinki, Helsinki, Finland. ¹⁸¹Clinical Microbiology and Parasitology Unit, Dr Zora Profozic Polyclinic, Zagreb, Croatia. ¹⁸²University Centre Varazdin, University North, Varazdin, Croatia. ¹⁸³Pacific Institute for Research & Evaluation, Calverton, MD, USA. ¹⁸⁴Achutha Menon Centre for Health Science Studies, Sree Chitra Tirunal Institute for Medical Sciences and Technology, Trivandrum, India. ¹⁸⁵Global Institute of Public Health (GIPH), Ananthapuri Hospitals and Research Centre, Trivandrum, India. ¹⁸⁶Faculty of Internal Medicine, Kyrgyz State Medical Academy, Bishkek, Kyrgyzstan. ¹⁸⁷Department of Atherosclerosis and Coronary Heart Disease, National Center of Cardiology and Internal Disease, Bishkek, Kyrgyzstan. ¹⁸⁸Institute of Addiction Research (ISFF), Frankfurt University of Applied Sciences, Frankfurt, Germany. ¹⁸⁹Department of Food Technology, College of Agriculture, Salahaddin University-Erbil, Erbil, Iraq. ¹⁹⁰Department of Medicine Huddinge, Karolinska Institutet, Stockholm, Sweden. ¹⁹¹Department of Information Technology, University of Human Development, Sulaimaniyah, Iraq. ¹⁹²Health Systems and Policy Research Unit, Ahmadu Bello University, Zaria, Nigeria. ¹⁹³Non-communicable Diseases Research Center, Tehran University of Medical Sciences, Tehran, Iran. ¹⁹⁴Iran National Institute of Health Research, Tehran University of Medical Sciences, Tehran, Iran. ¹⁹⁵Clinical Epidemiology and Public Health Research Unit, Burlo Garofolo Institute for Maternal and Child Health, Trieste, Italy. ¹⁹⁶Department of Public Health Medicine, University of Kwazulu-Natal, Durban, South Africa. ¹⁹⁷Health Sciences Research Center, Mazandaran University of Medical Sciences, Sari, Iran. ¹⁹⁸Social Determinants of Health Research Center, Kurdistan University of Medical Sciences, Sanandaj, Iran. ¹⁹⁹Department of Epidemiology and Biostatistics, Kurdistan University of Medical Sciences, Sanandaj, Iran. ²⁰⁰Department of Mathematical Sciences, University of Bath, Bath, UK. ²⁰¹International Laboratory for Air Quality and Health, Queensland University of Technology, Brisbane, Queensland, Australia. ²⁰²Department of Surgery, University of Washington, Seattle, WA, USA. ²⁰³Department of Health Management and Economics, Tehran University of Medical Sciences, Tehran, Iran. ²⁰⁴Health Management Research Center, Baqiyatallah University of Medical Sciences, Tehran, Iran. ²⁰⁵Department of Pediatric Medicine, Nishtar Medical University, Multan, Pakistan. ²⁰⁶Department of Pediatrics & Pediatric Pulmonology, Institute of Mother & Child Care, Multan, Pakistan. ²⁰⁷Cancer Research Center, Tehran University of Medical Sciences, Tehran, Iran. ²⁰⁸Department of Epidemiology & Biostatistics, Kermanshah University of Medical Sciences, Kermanshah, Iran. ²⁰⁹Suraj Eye Institute, Nagpur, India. ²¹⁰Cochrane South Africa, South African Medical Research Council, Cape Town, South Africa. ²¹¹General Surgery, Carol Davila University of Medicine and Pharmacy Bucharest, Bucharest, Romania. ²¹²General Surgery, Emergency Hospital of Bucharest, Bucharest, Romania. ²¹³Anatomy and Embryology, Carol Davila University of Medicine and Pharmacy, Bucharest, Romania. ²¹⁴Cardiology, Cardio-Aid, Bucharest, Romania. ²¹⁵Department of Biological Sciences, University of Embu, Embu, Kenya. ²¹⁶Institute for Global Health Innovations, Duy Tan University, Hanoi, Vietnam. ²¹⁷Center of Excellence in Behavioral Medicine, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam. ²¹⁸Public Health Department, Universitas Negeri Semarang, Kota Semarang, Indonesia. ²¹⁹Graduate Institute of Biomedical Informatics, Taipei Medical University, Taipei City, Taiwan. ²²⁰Mazandaran University of Medical Sciences, Sari, Iran. ²²¹Faculty of Medicine & Health Sciences, Stellenbosch University, Cape Town, South Africa. ²²²UCIBIO, University of Porto, Porto, Portugal. ²²³Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Ontario, Canada. ²²⁴Department of Psychiatry, University of Lagos, Lagos, Nigeria. ²²⁵Centre for Healthy Start Initiative, Lagos, Nigeria. ²²⁶Department of Pharmacology and Therapeutics, University of Nigeria Nsukka, Enugu, Nigeria. ²²⁷Center for Population Health Research, National Institute of Public Health, Cuernavaca, Mexico. ²²⁸School of Health and Welfare, Jönköping University, Jönköping, Sweden. ²²⁹Division of Mental and Physical

Health, Norwegian Institute of Public Health, Bergen, Norway. ²³⁰Department of Psychosocial Science, University of Bergen, Bergen, Norway. ²³¹Department of Respiratory Medicine, Jagadguru Sri Shivarathreeswara Academy of Health Education and Research, Mysore, India. ²³²Health Outcomes, Center for Health Outcomes & Evaluation, Bucharest, Romania. ²³³Augenpraxis Jonas, Heidelberg University, Heidelberg, Germany. ²³⁴Regional Medical Research Centre, Indian Council of Medical Research, Bhubaneswar, India. ²³⁵Department of Paediatrics, University of Melbourne, Melbourne, Victoria, Australia. ²³⁶Population Health, Murdoch Children's Research Institute, Melbourne, Victoria, Australia. ²³⁷Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Bergamo, Italy. ²³⁸Department of Economics and Business, University of Groningen, Groningen, The Netherlands. ²³⁹University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. ²⁴⁰Department of Nephrology, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, India. ²⁴¹Population Studies, International Institute for Population Sciences, Mumbai, India. ²⁴²Non-communicable Diseases Research Center, Alborz University of Medical Sciences, Karaj, Iran. ²⁴³College of Medicine, University of Central Florida, Orlando, FL, USA. ²⁴⁴College of Graduate Health Sciences, A. T. Still University, Mesa, AZ, USA. ²⁴⁵Thalassemia and Hemoglobinopathy Research Center, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran. ²⁴⁶Metabolomics and Genomics Research Center, Tehran University of Medical Sciences, Tehran, Iran. ²⁴⁷Sina Trauma and Surgery Research Center, Tehran University of Medical Sciences, Tehran, Iran. ²⁴⁸Department of Public Health and Mortality Studies, International Institute for Population Sciences, Mumbai, India. ²⁴⁹Policy Research Institute, Kathmandu, Nepal. ²⁵⁰Institute for Poverty Alleviation and International Development, Yonsei University, Wonju, South Korea. ²⁵¹WHO Collaborating Centre for Public Health Education and Training, Imperial College London, London, UK. ²⁵²University College London Hospitals, London, UK. ²⁵³Academic Public Health, Public Health England, London, UK. ²⁵⁴Translational Health Research Institute, Western Sydney University, Penrith, New South Wales, Australia. ²⁵⁵School of Social Sciences and Psychology, Western Sydney University, Penrith, New South Wales, Australia. ²⁵⁶Research Directorate, Nihon Gakko University, Fernando De La Mora, Paraguay. ²⁵⁷Research Direction, Universidad Nacional de Caaguazú, Coronel Oviedo, Paraguay. ²⁵⁸Department of Clinical Research, Federal University of Uberlândia, Uberlândia, Brazil. ²⁵⁹Golestan Research Center of Gastroenterology and Hepatology, Golestan University of Medical Sciences, Gorgan, Iran. ²⁶⁰Infectious Diseases and Tropical Medicine Research Center, Babol University of Medical Sciences, Babol, Iran. ²⁶¹Centro de Investigación Palmira, Agrosavia, Palmira, Colombia. ²⁶²Department of Ocean Science and Engineering, Southern University of Science and Technology, Shenzhen, China. ²⁶³Kermanshah University of Medical Sciences, Kermanshah, Iran. ²⁶⁴Department of Psychiatry, All India Institute of Medical Sciences, New Delhi, India. ²⁶⁵Department of Pathology, Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia. ²⁶⁶Social Development and Health Promotion Research Center, Kermanshah University of Medical Sciences, Kermanshah, Iran. ²⁶⁷Department of Entomology, Ain Shams University, Cairo, Egypt. ²⁶⁸Department of Surgery, Marshall University, Huntington, WV, USA. ²⁶⁹Department of Nutrition and Preventive Medicine, Case Western Reserve University, Cleveland, OH, USA. ²⁷⁰Institute of Social Medicine, University of Belgrade, Belgrade, Serbia. ²⁷¹Centre-School of Public Health and Health Management, University of Belgrade, Belgrade, Serbia. ²⁷²Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK. ²⁷³Surgery Department, Hamad Medical Corporation, Doha, Qatar. ²⁷⁴Faculty of Health & Social Sciences, Bournemouth University, Bournemouth, UK. ²⁷⁵University of Alabama at Birmingham, Birmingham, AL, USA. ²⁷⁶Dr D. Y. Patil University, Pune, India. ²⁷⁷Department of Psychology, University of Alabama at Birmingham, Birmingham, AL, USA. ²⁷⁸Department of Food Science and Nutrition, Jigjiga University, Jigjiga, Ethiopia. ²⁷⁹Independent Consultant, Karachi, Pakistan. ²⁸⁰School of Medicine, Dezfoul University of Medical Sciences, Dezfoul, Iran. ²⁸¹School of Medicine, Alborz University of Medical Sciences, Karaj, Iran. ²⁸²Chronic Diseases (Home Care) Research Center, Hamadan University of Medical Sciences, Hamadan, Iran. ²⁸³University School of Management and Entrepreneurship, Delhi Technological University, New Delhi, India. ²⁸⁴Department of Pulmonary Medicine, Fudan University, Shanghai, China. ²⁸⁵Centre for Medical Informatics, University of Edinburgh, Edinburgh, UK. ²⁸⁶Division of General Internal Medicine, Harvard University, Boston, MA, USA. ²⁸⁷National Institute of Infectious Diseases, Tokyo, Japan. ²⁸⁸Department of Health Education & Promotion, Kermanshah University of Medical Sciences, Kermanshah, Iran. ²⁸⁹School of Health, University of Technology Sydney, Sydney, New South Wales, Australia. ²⁹⁰Brasília University, Brasília, Brazil. ²⁹¹Department of the Health Industrial Complex and Innovation in Health, Federal Ministry of Health, Brasília, Brazil. ²⁹²Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL, USA. ²⁹³Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA. ²⁹⁴Department of Epidemiology, School of Preventive Oncology, Patna, India. ²⁹⁵Department of Epidemiology, Healis Sekhsaria Institute for Public Health, Mumbai, India. ²⁹⁶Centre for Fertility and Health, Norwegian Institute of Public Health, Bergen, Norway. ²⁹⁷Department of Pediatrics, King Saud University, Riyadh, Saudi Arabia. ²⁹⁸Pediatric Department, King Khalid University Hospital, Riyadh, Saudi Arabia. ²⁹⁹Hospital Universitario de la Princesa, Autonomous University of Madrid, Madrid, Spain. ³⁰⁰Centro de Investigación Biomédica en Red Enfermedades Respiratorias (CIBERES), Madrid, Spain. ³⁰¹Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. ³⁰²Hull York Medical School, University of Hull, Hull, UK. ³⁰³Division of Community Medicine, International Medical University, Kuala Lumpur, Malaysia. ³⁰⁴Department of Nursing, Muhammadiyah University of Surakarta, Kartasura, Indonesia. ³⁰⁵Department of Public Health, China Medical University, Taichung, Taiwan. ³⁰⁶Department of Community Medicine, Ahmadu Bello University, Zaria, Nigeria. ³⁰⁷Neurology Department, Sree Chitra Tirunal Institute for Medical Sciences and Technology, Trivandrum, India. ³⁰⁸Sree Chitra Tirunal Institute for Medical Sciences and Technology, Trivandrum, India. ³⁰⁹Department of Medicine, University of Valencia, Valencia, Spain. ³¹⁰Carlos III Health Institute, Biomedical Research Networking Center for Mental Health Network (CIBERSAM), Madrid, Spain. ³¹¹Department of Pediatrics, Hawassa University, Hawassa, Ethiopia. ³¹²International Vaccine Institute, Seoul, South Korea. ³¹³College of Medicine, Alfaisal University, Riyadh, Saudi Arabia. ³¹⁴Department of Anesthesiology, Perioperative, and Pain Medicine, University of Virginia, Charlottesville, VA, USA. ³¹⁵Department of Anesthesiology, King Farah Medical City, Riyadh, Saudi Arabia. ³¹⁶Department of Medical Microbiology, University of Gondar, Gondar, Ethiopia. ³¹⁷Department of Epidemiology and Biostatistics, University of Gondar, Gondar, Ethiopia. ³¹⁸Department of Public Health and Community Medicine, Central University of Kerala, Kasaragod, India. ³¹⁹Faculty of Health Sciences, Jagiellonian University Medical College, Krakow, Poland. ³²⁰The Agency for Health Technology Assessment and Tariff System, Warsaw, Poland. ³²¹Department of Pathology and Legal Medicine, University of São Paulo, Ribeirão Preto, Brazil. ³²²Department of Health Economics, Hanoi Medical University, Hanoi, Vietnam. ³²³Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. ³²⁴Gomal Center of Biochemistry and Biotechnology, Gomal University, Dera Ismail Khan, Pakistan. ³²⁵TB Culture Laboratory, Mufti Mehmood Memorial Teaching Hospital, Dera Ismail Khan, Pakistan. ³²⁶Division of Health Sciences, University of Warwick, Coventry, UK. ³²⁷Argentine Society of Medicine, Ciudad de Buenos Aires, Argentina. ³²⁸Velez Sarsfield Hospital, Buenos Aires, Argentina. ³²⁹Psychosocial Injuries Research Center, Ilam University of Medical Sciences, Ilam, Iran. ³³⁰Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy. ³³¹Occupational Health Unit, Sant'Orsola Malpighi Hospital, Bologna, Italy. ³³²Department of Health Care Administration and Economics, National Research University Higher School of Economics, Moscow, Russia. ³³³Department of Global Health and Population, Harvard University, Boston, MA, USA. ³³⁴Department of Economics, University of Göttingen, Göttingen, Germany. ³³⁵Foundation University Medical College, Foundation University Islamabad, Islamabad, Pakistan. ³³⁶Department of Psychiatry, University of São Paulo, São Paulo, Brazil. ³³⁷Institute of Health and Society, University of Oslo, Oslo, Norway. ³³⁸Department of Neurology, Technical University of Munich, Munich, Germany. ³³⁹School of Population Health & Environmental Sciences, King's College London, London, UK. ³⁴⁰NIHR Biomedical Research Centre, Guy's and St Thomas' Hospital and Kings College London, London, UK. ³⁴¹Department of Diabetes and Metabolic Diseases, University of Tokyo, Tokyo, Japan. ³⁴²Wolkite University, Wolkite, Ethiopia. ³⁴³Centre for Suicide Research and Prevention, University of Hong Kong, Hong Kong, China. ³⁴⁴Department of Social Work and Social Administration, University of Hong Kong, Hong Kong, China. ³⁴⁵School of Allied Health Sciences, Addis Ababa University, Addis Ababa, Ethiopia. ³⁴⁶Department of Psychopharmacology, National Center of Neurology and Psychiatry, Tokyo, Japan. ³⁴⁷Health Economics & Finance, Global Health, Jackson State University, Jackson, MS, USA. ³⁴⁸School of Medicine, Tsinghua University, Peking, China. ³⁴⁹Prevention of Cardiovascular Disease Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran. ³⁵⁰Global Health Institute, Wuhan University, Wuhan, China. ³⁵¹Department of Epidemiology and Biostatistics, Wuhan University, Wuhan, China. ³⁵²Department of Medicine, Monash University, Melbourne, Victoria, Australia. ³⁵³Maternal and Child Health Division, International Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka, Bangladesh. ³⁵⁴George Warren Brown School, Washington University in St Louis, St Louis, MO, USA. ³⁵⁵School of Public Health, Wuhan University of Science and Technology, Wuhan, China. ³⁵⁶Hubei Province Key Laboratory of Occupational Hazard Identification and Control, Wuhan University of Science and Technology, Wuhan, China. ³⁵⁷These authors contributed equally: Nicholas Graetz, Lauren Woyczynski. ³⁵⁸These authors jointly supervised this work: Emmanuela Gakidou, Simon I. Hay. *e-mail: sihay@uw.edu

Overview

Using a Bayesian model-based geostatistical framework and synthesizing geolocated data from 528 household and census datasets, this analysis provides subnational estimates of mean numbers years of education and the proportion of the population who attained key levels of education for women of reproductive age (15–49 years), women aged 20–24 years, and equivalent male age bins between 2000 and 2017 in 105 countries across all low- and middle-income countries (LMICs). Countries were selected for inclusion in this analysis using the socio-demographic index (SDI) published in the Global Burden of Disease (GBD) study²⁷. The SDI is a measure of development that combines education, fertility, and poverty. Countries in the middle, lower-middle, or low SDI quintiles were included, with several exceptions. Albania, Bosnia, and Moldova were excluded despite middle SDI status due to geographical discontinuity with other included countries and lack of available survey data. Libya, Malaysia, Panama, and Turkmenistan were included despite higher-middle SDI status to create better geographical continuity. We did not analyse American Samoa, Federated States of Micronesia, Fiji, Kiribati, Marshall Islands, Samoa, Solomon Islands, or Tonga, where no available survey data could be sourced. Analytical steps are described below, and additional details can be found in the Supplementary Information.

Data

We compiled a database of survey and census datasets that contained geocoding of subnational administrative boundaries or GPS coordinates for sampled clusters. These included datasets from 528 sources (see Supplementary Table 2). These sources comprised at least one data source for all but two countries on our list of LMICs: Western Sahara and French Guiana. We chose to exclude these two countries from our analysis; 42 of 105 included countries have only subnational administrative level data. We extracted demographic, education, and sample design variables. The coding of educational attainment varies across survey families. In some surveys, the precise number of years of attainment is not provided, with attainment instead aggregated into categories such as ‘primary completion’ or ‘secondary completion’. In such cases, individuals who report ‘primary completion’ may have gone on to complete some portion of secondary education, but these additional years of education are not captured in the underlying dataset. Previous efforts to examine trends in mean years of education have either assumed that no additional years of education were completed (that is, primary education only) or have used the midpoint between primary and secondary education as a proxy²⁸. Trends in the single-year data, however, demonstrate that such assumptions introduce bias in the estimation of attainment trends over time and space, as differences in actual drop-out patterns or binning schema can lead to biased mean estimates²⁹.

For this analysis, we used a recently developed method that selects a training subset of similar surveys across time and space to estimate the unobserved single-year distribution of binned datasets²⁹. In comprehensive tests of cross-validation that leveraged data for which the single-year distributions are observed, this algorithmic approach significantly reduces bias in summary statistics estimated from datasets with binned coding schemes compared to alternatives such as the standard-duration method²⁸. The years in all coding schemes were mapped to the country- and year-specific references in the UNESCO International Standard Classification of Education (ISCED) for comparability³⁰. We used a top coding of 18 years on all data; this is a common threshold in many surveys that have a cap and it is reasonable to assume that the importance of education for health outcomes (and other related SDGs) greatly diminishes after what is the equivalent of 2 to 3 years of graduate education in most systems.

Data were aggregated to mean years of education attained and the proportions achieving key levels of education. The levels chosen were proportion with zero years, proportion with less than primary school (1–5 years of education), proportion with at least primary school (6–11 years of education), and proportion achieving secondary school or higher (12 or more years of education). A subset of the data for a smaller age bin (20–24 years) was also examined to more closely track temporal shifts. Equivalent age bins were aggregated for both women and men to examine disparities in mean years of attainment by sex. Where GPS coordinates were available, data were aggregated to a specific latitude and longitude assuming a simple-random sample, as the cluster is the primary sampling unit for the stratified design survey families, such as the Demographic and Health Survey (DHS) and Multiple Indicator Cluster Survey (MICS). Where only geographical information was available at the level of administrative units, data were aggregated with appropriate weighting according to their sample design. Design effects were estimated using a package for analysing complex survey data in R³¹.

Spatial covariates

To leverage strength from locations with observations to the entire spatiotemporal domain, we compiled several 5×5 -km² raster layers of possible socioeconomic and environmental correlates of education (Supplementary Table 5 and Supplementary Fig. 6). Acquisition of temporally dynamic datasets, where possible, was prioritized to best match our observations and thus predict the changing dynamics of educational attainment. We included nine covariates indexed at the 5×5 -km² level: access to roads, nighttime lights^{iv}, population^{iv}, growing season, aridity^{iv}, elevation, urbanicity^{iv}, irrigation, and year^{iv} (tv, time-varying covariates). More details, including plots of all covariates, can be found in the Supplementary Information.

Our primary goal is to provide educational attainment predictions across LMICs at a high (local) resolution, and our methods provide the best out-of-sample predictive performance at the expense of inferential understanding. To select covariates and capture possible nonlinear effects and complex interactions between them, an ensemble covariate modelling method was implemented³². For each region, three submodels were fitted to our outcomes using all of our covariate data: generalized additive models, boosted regression trees, and lasso regression. Each submodel was fit using fivefold cross-validation to avoid overfitting and the out-of-sample predictions from across the five folds were compiled into a single comprehensive set of predictions from that model. Additionally, the same submodels were also run using 100% of the data and a full set of in-sample predictions were created. The five sets of out-of-sample submodel predictions were fed into the full geostatistical model as predictors when performing the model fit. The in-sample predictions from the submodels were used as the covariates when generating predictions using the fitted full geostatistical model. This methodology maximizes out-of-sample predictive performance at the expense of the ability to provide statistical inference on the relationships between the predictors and the outcome. A recent study has shown that this ensemble approach can improve predictive validity by up to 25% over an individual model³². More details on this approach can be found in the Supplementary Information.

The primary goal of using the stacking procedure in our analyses was to maximize the predictive power of the raster covariates by capturing the nonlinear effects and complex interactions between covariates to optimize the model performance. It has previously been suggested³² that the primary purpose of the submodel predictions is to improve the mean function of the Gaussian process. Although we have determined a way to include the uncertainty from two of our submodels (lasso regression and generalized additive models (GAM)), we have not determined a way to include uncertainty from the boosted regression tree (BRT) submodel into our final estimates. Whereas GAM and lasso regression seek to fit a single model that best describes the relationship

between response variable and some set of predictors, BRT method fits a large number of relatively simple models for which the predictions are then combined to give robust estimates of the response. Although this feature of the BRT model makes it a powerful tool for analysing complex data, quantifying the relative uncertainty contributed by each simple model as well as uncertainty from the complex interactions of the predictor variables is challenging^{33,34}. It is worth noting, however, that our out-of-sample validation indicates that the 95% coverage is fairly accurate (for example, closely ranges around 95%) as shown in the figures and table of Supplementary Information section 4.3.2. This indicates that we are not misrepresenting the uncertainty in our final estimates.

Analysis

Geostatistical model. Gaussian and binomial data are modelled within a Bayesian hierarchical modelling framework using a spatially and temporally explicit hierarchical generalized linear regression model to fit the mean number years of education attainment and the proportion of the population who achieved key bins of school in 14 regions across all LMICs as defined in the GBD study (Extended Data Fig. 1). This means we fit 14 independent models for each indicator (for example, the proportion of women with zero years of schooling). GBD study design sought to create regions on the basis of three primary criteria: epidemiological homogeneity, sociodemographic similarity, and geographical contiguity²⁷. Fitting our models by these regions has the advantage of allowing for some non-stationarity and non-isotropy in the spatial error term, compared to if we modelled one spatiotemporal random-effect structure over the entire modelling region of all LMICs.

For each Gaussian indicator, we modelled the mean number of years of attainment in each survey cluster, d . Survey clusters are precisely located by their GPS coordinates and year of observation, which we map to a spatial raster location i at time t . We model the mean number of years of attainment as Gaussian data given fixed precision τ and a scaling parameter s_d (defined by the sample size in the observed cluster). As we may have observed multiple data clusters within a given location i at time t , we refer to the mean attainment, μ , within a given cluster d by its indexed location i , and time t as $\mu_{i(d),t(d)}$.

$\text{edu}_d | \mu_{i(d),t(d)}, s_d, \tau \sim \text{Normal}(\mu_{i(d),t(d)}, \tau s_d) \quad \forall \text{ observed clusters } d$

$$\mu_{i,t} = \beta_0 + \mathbf{X}_{i,t} \boldsymbol{\beta} + Z_{i,t} + \epsilon_{\text{ctr}(i)} + \epsilon_{i,t} \quad \forall i \in \text{spatial domain} \quad \forall t \in \text{time domain}$$

For each binomial indicator, we modelled the number of individuals at a given attainment level in each survey cluster, d . We observed the number of individuals reporting a given attainment level as binomial count data C_d among an observed sample size N_d . As we may have observed multiple data clusters within a given location i at time t , we refer to the probability of attaining that level, p , within a given cluster d by its indexed location i and time t as $p_{i(d),t(d)}$.

$C_d | p_{i(d),t(d)}, N_d \sim \text{Binomial}(p_{i(d),t(d)}, N_d) \quad \forall \text{ observed clusters } d$

$$\text{logit}(p_{i,t}) = \beta_0 + \mathbf{X}_{i,t} \boldsymbol{\beta} + Z_{i,t} + \epsilon_{\text{ctr}(i)} + \epsilon_{i,t} \quad \forall i \in \text{spatial domain} \quad \forall t \in \text{time domain}$$

We used a continuation-ratio modelling approach to account for the ordinal data structure of the binomial indicators³⁵. To do this, the proportion of the population with zero years of education was modelled using a binomial model. The proportion with less than primary education was modelled as those with less than primary education of those that have more than zero years of education. The same method followed for the proportion of population completing primary education. The

proportion achieving secondary school or higher was estimated as the complement of the sum of the three binomial models.

The remaining parameter specification was consistent between all indicators in both binomial and Gaussian models:

$$\sum_{h=1}^3 \beta_h = 1$$

$$\epsilon_{\text{ctr}} \sim \text{iid Normal}(0, \gamma^2)$$

$$\epsilon_{i,t} \sim \text{iid Normal}(0, \sigma^2)$$

$$\mathbf{Z} \sim \text{GP}(0, \Sigma^{\text{space}} \otimes \Sigma^{\text{time}})$$

$$\Sigma^{\text{space}} = \frac{\omega^2}{\Gamma(\nu) 2^{\nu-1}} \times (\kappa D)^\nu \times K_\nu(\kappa D)$$

$$\Sigma_{j,k}^{\text{time}} = \rho^{|k-j|}$$

For indices d , i , and t , $*$ (index) is the value of $*$ at that index. The probabilities $p_{i,t}$ represent both the annual proportions at the space-time location and the probability that an individual had that level of attainment given that they lived at that particular location. The annual probability $p_{i,t}$ of each indicator (or $\mu_{i,t}$ for the mean indicators) was modelled as a linear combination of the three submodels (GAM, BRT, and lasso regression), rasterized covariate values $X_{i,t}$, a correlated spatiotemporal error term $Z_{i,t}$, country random effects $\epsilon_{\text{ctr}(i)}$ with one unstructured country random effect fit for each country in the modelling region and all sharing a common variance parameter, γ^2 , and an independent nugget effect $\epsilon_{i,t}$ with variance parameter σ^2 . Coefficients β_h in the three submodels $h = 1, 2, 3$ represent their respective predictive weighting in the mean logit link, while the joint error term $Z_{i,t}$ accounts for residual spatiotemporal autocorrelation between individual data points that remains after accounting for the predictive effect of the submodel covariates, the country-level random effect $\epsilon_{\text{ctr}(i)}$, and the nugget independent error term, $\epsilon_{i,t}$. The purpose of the country-level random effect is to capture spatially unstructured, unobserved country-specific variables, as there are often sharp discontinuities in educational attainment between adjacent countries due to systematic differences in governance, infrastructure, and social policies.

The residuals $Z_{i,t}$ are modelled as a three-dimensional Gaussian process (GP) in space-time centred at zero and with a covariance matrix constructed from a Kronecker product of spatial and temporal covariance kernels. The spatial covariance Σ^{space} is modelled using an isotropic and stationary Matérn function³⁶, and temporal covariance Σ^{time} as an annual autoregressive (AR1) function over the 18 years represented in the model. In the stationary Matérn function, Γ is the Gamma function, K_ν is the modified Bessel function of order $\nu > 0$, $\kappa > 0$ is a scaling parameter, D denotes the Euclidean distance, and ω^2 is the marginal variance. The scaling parameter, κ , is defined to be $\kappa = \sqrt{8\nu}/\delta$ where δ is a range parameter (which is about the distance for which the covariance function approaches 0.1) and ν is a scaling constant, which is set to 2 rather than fit from the data^{37,38}. This parameter is difficult to reliably fit, as documented by many other analyses^{37,39,40} that set this parameter to 2. The number of rows and the number of columns of the spatial Matérn covariance matrix are equal to the number of spatial mesh points for a given modelling region. In the AR1 function, ρ is the autocorrelation function (ACF), and k and j are points in the time series where $|k-j|$ defines the lag. The number of rows and the number of columns of the AR1 covariance matrix are equal to the number of temporal mesh points (18). The number of rows and the number of columns of the space-time covariance matrix, $\Sigma^{\text{space}} \otimes \Sigma^{\text{time}}$, for a given modelling region are equal to: the number of spatial mesh points \times the number of temporal mesh points.

This approach leveraged the residual correlation structure of the data to more accurately predict estimates for locations with no data, while also propagating the dependence in the data through to uncertainty estimates⁴¹. The posterior distributions were fit using computationally efficient and accurate approximations in R-integrated nested Laplace approximation (INLA) with the stochastic partial differential equations (SPDE) approximation to the Gaussian process residuals using R project version 3.5.1^{42–45}. The SPDE approach using INLA has been demonstrated elsewhere, including the estimation of health indicators, particulate air matter, and population age structure^{10,11,46,47}. Uncertainty intervals were generated from 1,000 draws (that is, statistically plausible candidate maps)⁴⁸ created from the posterior-estimated distributions of modelled parameters. Additional details regarding model and estimation processes can be found in the Supplementary Information.

To transform grid cell-level estimates into a range of information that is useful to a wide constituency of potential users, these estimates were aggregated from the 1,000 candidate maps up to district, provincial, and national levels using $5 \times 5\text{-km}^2$ population data⁴⁹. This aggregation also enabled the calibration of estimates to national GBD estimates for 2000–2017. This was achieved by calculating the ratio of the posterior mean national-level estimate from each candidate map draw in the analysis to the posterior mean national estimates from GBD, and then multiplying each cell in the posterior sample by this ratio. National-level estimates from this analysis with GBD estimates can be found in Supplementary Table 44.

To illustrate how subnational progress has contributed differentially to national progress (Fig. 3), we decomposed the improvement in the national rate of secondary completion since 2000 for each country into the additive contributions of rate changes at the second administrative level, where C is the national secondary rate change, N is the total number of second-level administrative units, c_i is the population proportion in administrative unit i , and r_i is the rate of secondary attainment in administrative unit i .

$$C = \sum_{i=1}^N (c_{i,2017}r_{i,2017}) - (c_{i,2000}r_{i,2000})$$

Although the model can predict at all locations covered by available raster covariates, all final model outputs for which land cover was classified as ‘barren or sparsely vegetated’ were masked, on the basis of the most recently available Moderate Resolution Imaging Spectroradiometer (MODIS) satellite data (2013), as well as areas in which the total population density was less than 10 individuals per $1 \times 1\text{-km}^2$ pixel in 2015⁵⁰. This step has led to improved understanding when communicating with data specialists and policy-makers.

Model validation. Models were validated using source-stratified five-fold cross-validation. To offer a more stringent analysis by respecting some of the source and spatial correlation in the data, holdout sets were created by combining sets of data sources (for example, entire survey- or census-years). Model performance was summarized by the bias (mean error), total variance (root-mean-square error) and 95% data coverage within prediction intervals, and the correlation between observed data and predictions. All validation metrics were calculated on the predictions from the fivefold cross-validation. Where possible, estimates from these models were compared against other existing estimates. Furthermore, measures of spatial and temporal autocorrelation pre- and post-modelling were examined to verify correct recognition, fitting, and accounting for the complex spatiotemporal correlation structure in the data. All validation procedures and corresponding results are provided in the Supplementary Information.

Limitations. Our analysis is not without several important limitations. First, almost all data collection tools conflate gender and sex and we therefore do not capture the full distribution of sex or gender separately

in our data. We refer throughout to the measurement of ‘gender (in equality)’, following the usage in SDG 5. Second, it is extremely difficult to quantify quality of education on this scale in a comparable way. Quality is ultimately a large part of the SDG agenda and of utmost importance to achieving equity in opportunity for social mobility. However, many studies across diverse low- and middle-income settings have linked attainment, even very low levels, to measurable improvement in maternal and child health¹⁷. As our analysis highlights with the proportional indicators, there are still many subnational regions across the world where large proportions do not complete primary school. A third limitation is that we are unable to measure or account for migration. A concept note released from the forthcoming Global Education Monitoring Report 2019 focuses on how migration and displacement affects schooling⁵¹. Our estimates of the modelled outcome, educational attainment for a particular space–time–age–sex, are demonstrated to be statistically unbiased (Supplementary Information section 4.3); however, interpretation of any change in attainment as a change in the underlying education system could potentially be biased by the effects of migration. It is possible that geographical disparities reflect changes in population composition rather than changes in the underlying infrastructure or education system. Pathways for this change are complex and may be voluntary. Those who manage to receive an education in a low-attainment area may have an increased ability to migrate and choose to do so. This change may also be involuntary, particularly in politically unstable areas where displacement may make geographical changes over time difficult to estimate. A shifting population composition is a general limitation of many longitudinal ecological analyses, but the spatially granular nature of the analyses used here may be more sensitive to the effects of mobile populations.

Our analysis is purely predictive but draws heavily in its motivation from a rich history of literature on the role of education in reducing maternal mortality, improving child health, and increasing human capital. Studies have also demonstrated complex relationships between increased education and a myriad of positive health outcomes, such as HIV risk reductions and spillover effects to other household members^{52,53}. The vast majority of these studies are associational and recent attempts at causal analyses have provided more-mixed evidence^{54–56}. Although causal analyses of education are very difficult and often rely on situational quasi-experiments, associational analyses using the most comprehensive datasets demonstrate consistent support for the connection between education and health^{17,57}. Looking towards future analyses, it will be important to study patterns of change in these data and how they overlap with distributions of health. Lastly, our estimates cannot be seen as a replacement for proper data collection systems, especially for tracking contemporaneous change. Our analysis of uncertainty at a high-resolution may be used to inform investment in more robust data systems and collection efforts, especially if the ultimate goal is to measure and track progress in the quality of schooling.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The findings of this study are supported by data that are available in public online repositories, data that are publicly available upon request from the data provider, and data that are not publicly available owing to restrictions by the data provider, which were used under license for the current study, but may be available from the authors upon reasonable request and permission of the data provider. A detailed table of data sources and availability can be found in Supplementary Table 2. Interactive visualization tools are available at <https://vizhub.healthdata.org/lbd/education>. All maps presented in this study are generated by the authors; no permissions are required for publication. Administrative

boundaries were retrieved from the Global Administrative Unit Layers (GAUL) dataset, implemented by FAO within the CountrySTAT and Agricultural Market Information System (AMIS) projects⁵⁸. Land cover was retrieved from the online Data Pool, courtesy of the NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota⁵⁰. Lakes were retrieved from the Global Lakes and Wetlands Database (GLWD), courtesy of the World Wildlife Fund and the Center for Environmental Systems Research, University of Kassel^{59,60}. Populations were retrieved from WorldPop^{49,61}. All maps were produced using ArcGIS Desktop 10.6.

Code availability

Our study follows the Guidelines for Accurate and Transparent Health Estimates Reporting (GATHER). All code used for these analyses is available online at <http://ghdx.healthdata.org/record/ihme-data/lmic-education-geospatial-estimates-2000-2017>, and at <http://github.com/ihmeuw/lbd/tree/edu-lmic-2019>.

27. GBD 2016 Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **390**, 1151–1210 (2017).
28. Barro, R. J. & Lee, J. W. A new data set of educational attainment in the world, 1950–2010. *J. Dev. Econ.* **104**, 184–198 (2013).
29. Friedman, J., Graetz, N. & Gakidou, E. Improving the estimation of educational attainment: new methods for assessing average years of schooling from binned data. *PLoS ONE* **13**, e0208019 (2018).
30. UNESCO. *ISCED Mappings* (UNESCO, 2016); <http://uis.unesco.org/en/isced-mappings>
31. Lumley, T. survey: analysis of complex survey samples. R package v.3.36 <https://cran.r-project.org/web/packages/survey.pdf> (2019).
32. Bhatt, S. et al. Improved prediction accuracy for disease risk mapping using Gaussian process stacked generalization. *J. R. Soc. Interface* **14**, 20170520 (2017).
33. Elith, J., Leathwick, J. R. & Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **77**, 802–813 (2008).
34. Leathwick, J., Elith, J., Francis, M., Hastie, T. & Taylor, P. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Mar. Ecol. Prog. Ser.* **321**, 267–281 (2006).
35. Hosmer, D. W., & Lemeshow, S. in *Applied Logistic Regression* 289–305 (Wiley, 2013).
36. Stein, M. L. *Interpolation of Spatial Data* (Springer, 1999).
37. Lindgren, F. & Rue, H. Bayesian spatial modelling with R-INLA. *J. Stat. Softw.* **63**, 1–25 (2015).
38. Lindgren, F., Rue, H. & Lindström, J. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J. R. Stat. Soc. B* **73**, 423–498 (2011).
39. Rozanov, Y. A. in *Markov Random Fields* 55–102 (Springer, 1982).
40. Whittle, P. On stationary processes in the plane. *Biometrika* **41**, 434–449 (1954).
41. Diggle, P. & Ribeiro, P. J. *Model-based Geostatistics* (Springer, 2007).
42. Rue, H., Martino, S. & Chopin, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. B* **71**, 319–392 (2009).
43. Rue, H. et al. *Bayesian Computing with INLA* (2014); <http://www.r-inla.org/>
44. Blangiardo, M., Cameletti, M., Baio, G. & Rue, H. Spatial and spatio-temporal models with R-INLA. *Spat. Spatiotemporal Epidemiol.* **7**, 39–55 (2013).
45. Krainski, E. T., Lindgren, F., Simpson, D. & Rue, H. *The R-INLA Tutorial on SPDE Models* (2017); <https://inla.r-inla-download.org/r-inla.org/tutorials/spde/spde-tutorial.pdf>
46. Cameletti, M., Lindgren, F., Simpson, D. & Rue, H. Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *Adv. Stat. Anal.* **97**, 109–131 (2013).
47. Alegana, V. A. et al. Fine resolution mapping of population age-structures for health and development applications. *J. R. Soc. Interface* **12**, 20150073 (2015).
48. Patil, A. P., Gething, P. W., Piel, F. B. & Hay, S. I. Bayesian geostatistics in health cartography: the perspective of malaria. *Trends Parasitol.* **27**, 246–253 (2011).
49. Tatem, A. J. WorldPop, open data for spatial demography. *Sci. Data* **4**, 170004 (2017).
50. Friedl, M. et al. MCD12Q2 v006. MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500 m SIN Grid. (NASA EOSDIS Land Processes DAAC, 2019); <https://doi.org/10.5067/MODIS/MCD12Q1.006>
51. Global Education Monitoring Report. *Concept Note for the 2019 Global Education Monitoring Report on Education and Migration* (2017).
52. Behrman, J. A. The effect of increased primary schooling on adult women's HIV status in Malawi and Uganda: universal primary education as a natural experiment. *Soc. Sci. Med.* **127**, 108–115 (2015).
53. De Neve, J.-W., Fink, G., Subramanian, S. V., Moyo, S. & Bor, J. Length of secondary schooling and risk of HIV infection in Botswana: evidence from a natural experiment. *Lancet Glob. Health* **3**, e470–e477 (2015).
54. McCrary, J. & Royer, H. The effect of female education on fertility and infant health: evidence from school entry policies using exact date of birth. *Am. Econ. Rev.* **101**, 158–195 (2011).
55. Karlsson, O., De Neve, J.-W. & Subramanian, S. V. Weakening association of parental education: analysis of child health outcomes in 43 low- and middle-income countries. *Int. J. Epidemiol.* **48**, 83–97 (2019).
56. De Neve, J.-W. & Fink, G. Children's education and parental old age survival — quasi-experimental evidence on the intergenerational effects of human capital investment. *J. Health Econ.* **58**, 76–89 (2018).
57. Pamuk, E. R., Fuchs, R. & Lutz, W. Comparing relative effects of education and economic resources on infant mortality in developing countries. *Popul. Dev. Rev.* **37**, 637–664 (2011).
58. FAO-UN. *The Global Administrative Unit Layers (GAUL)* (2015); <http://www.fao.org/geonetwork/srv/en/metadata.show?id=12691>
59. World Wildlife Fund. *Global Lakes and Wetlands Database, Level 3* (2004); <https://www.worldwildlife.org/pages/global-lakes-and-wetlands-database>
60. Lehner, B. & Döll, P. Development and validation of a global database of lakes, reservoirs and wetlands. *J. Hydrol.* **296**, 1–22 (2004).
61. WorldPop. *Data Types* (accessed: 7th July 2017); <https://www.worldpop.org/project/list>
62. Channan, S., Collins, K. & Emanuel, W. Global mosaics of the standard MODIS land cover type data. University of Maryland and the Pacific Northwest National Laboratory, College Park, Maryland, USA. (2014)

Acknowledgements This work was primarily supported by grant OPP132415 from the Bill & Melinda Gates Foundation. N.G. is the recipient of a training grant from the National Institute of Child Health and Human Development (T32 HD-007242-36A1).

Author contributions S.I.H. and N.G. conceived and planned the study. K.W. and J.H. extracted, processed, and geo-positioned the data. L.W. and N.G. carried out the statistical analyses. All authors provided intellectual inputs into aspects of this study. N.G., L.W., J.H., and L.E. prepared figures and tables. N.G. wrote the manuscript with assistance by S.B.M., and all authors contributed to subsequent revisions.

Competing interests The authors declare no competing interests.

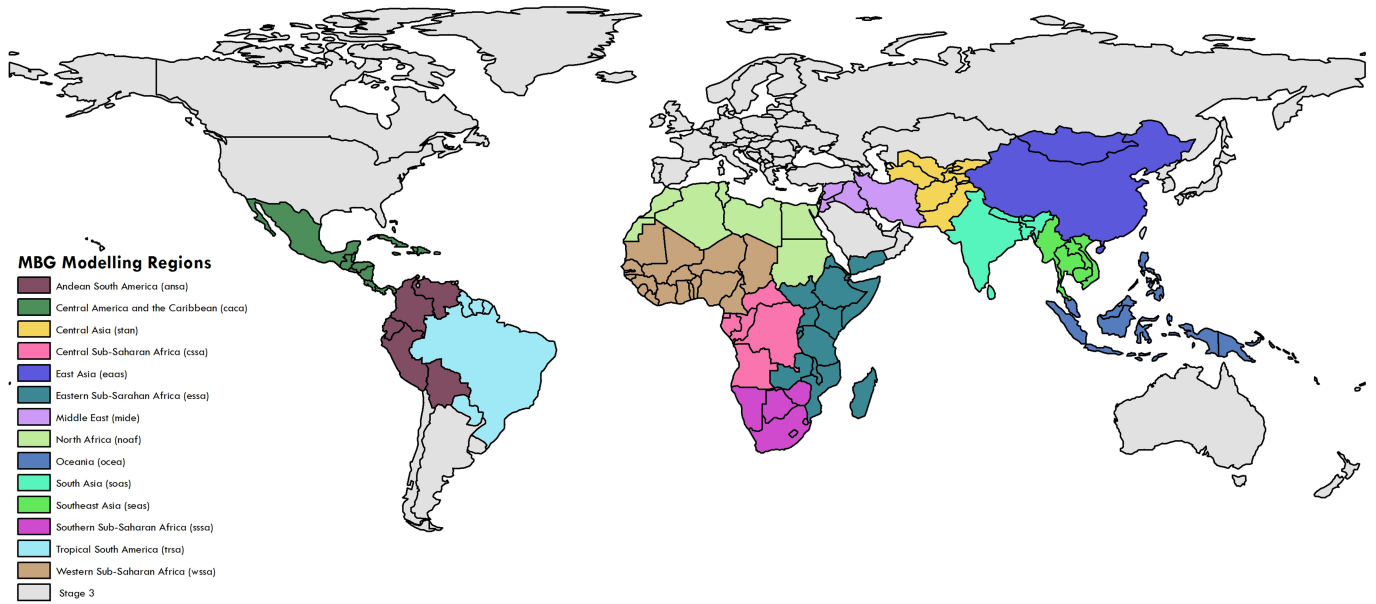
Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41586-019-1872-1>.

Correspondence and requests for materials should be addressed to S.I.H.

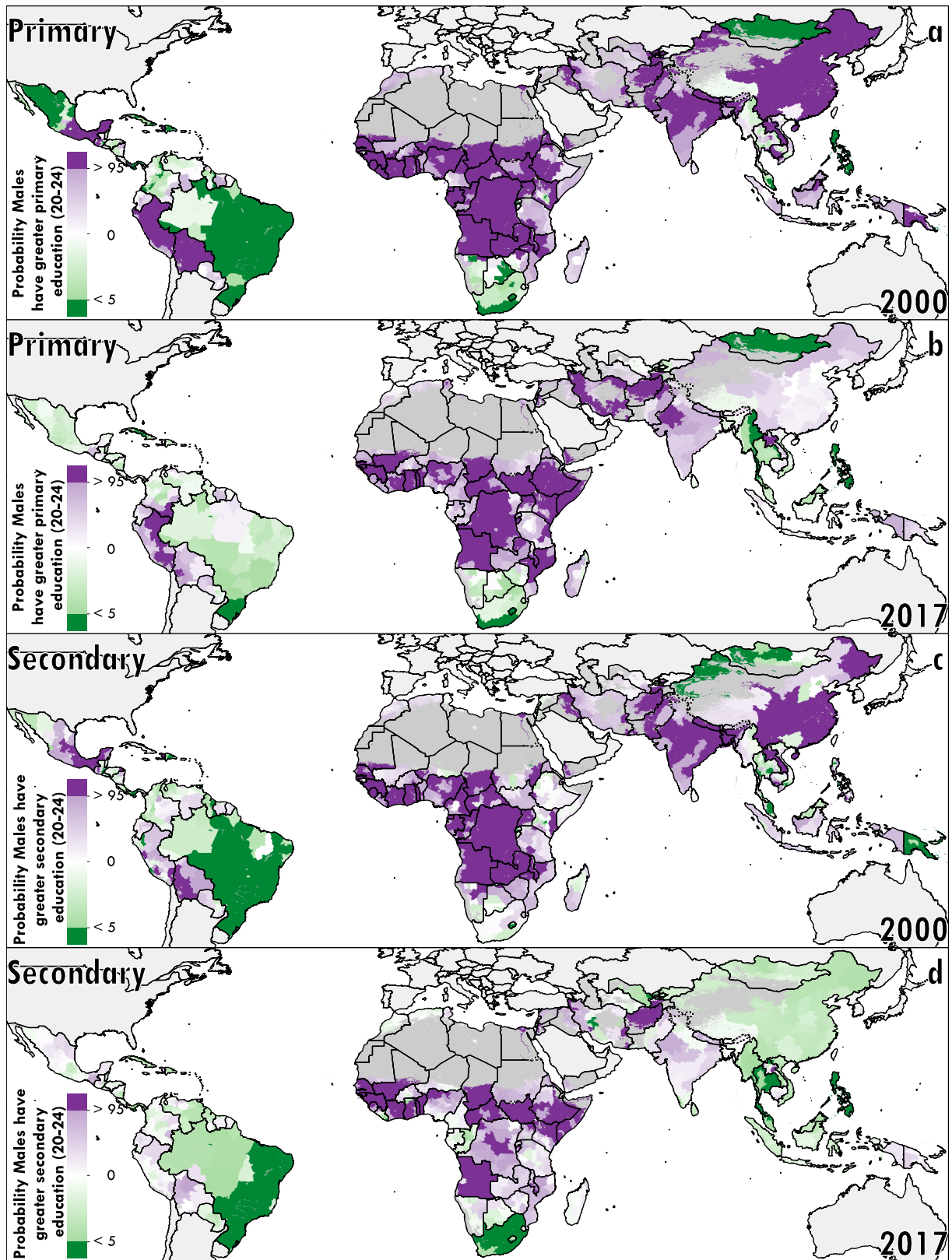
Peer review information *Nature* thanks M. Dolores Ugarte and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



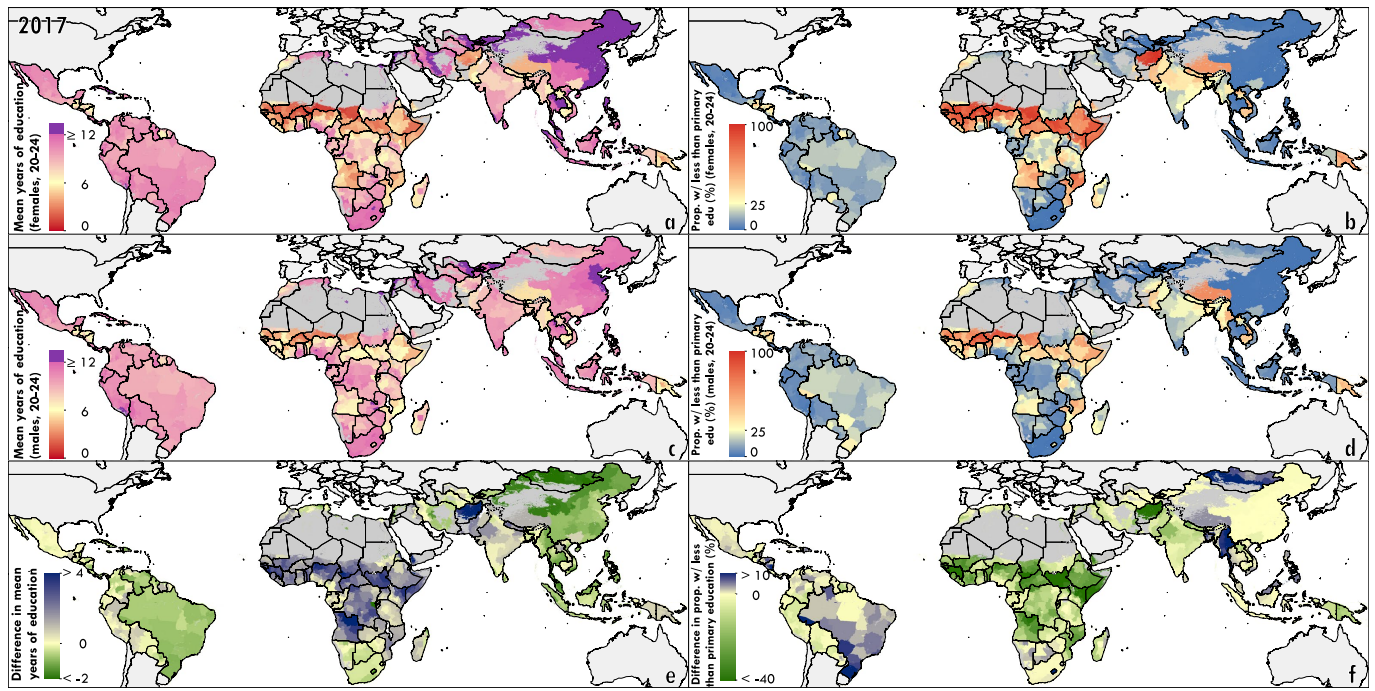
Extended Data Fig. 1 | Modelling regions based on geographical and SDI regions from the GBD. Modelling regions were defined as follows. Andean South America, Central America and the Caribbean, central sub-Saharan Africa, East Asia, eastern sub-Saharan Africa, Middle East, North Africa,

Oceania, Southeast Asia, South Asia, southern sub-Saharan Africa, Central Asia, Tropical South America, and western sub-Saharan Africa. Regions in grey were not included in our models due to high-middle and high SDIs²⁷. The map was produced using ArcGIS Desktop 10.6.



Extended Data Fig. 2 | Probability that the ratio of men to women aged 20–24 years who attained primary and secondary education is >1 in 2000 and 2017. a–d, Probability that ratio is >1 (for example, men complete at a

higher rate than women) for attaining primary education (a, b) and secondary education (c, d), aggregated to first administrative-level units in 2000 (a, c) and 2017 (b, d). Maps were produced using ArcGIS Desktop 10.6.



Extended Data Fig. 3 | Average educational attainment and proportion with no primary school at the first administrative level and absolute difference between women and men aged 20–24 years. a–d, Average educational attainment for women (a) and men (c) and proportion with no primary school for women (b) and men (d) aged 20–24 years in 2017. e, f, The absolute difference in average educational attainment between men and women aged 20–24 years in 2017 (e) and proportion of individuals with no primary school

education (f). Maps reflect administrative boundaries, land cover, lakes and population; grey-coloured grid cells were classified as ‘barren or sparsely vegetated’ and had fewer than ten people per $1 \times 1\text{-km}^2$ grid cell^{49,58–60,62}, or were not included in these analyses. Interactive visualization tools are available at <https://vizhub.healthdata.org/lbd/education>. Maps were produced using ArcGIS Desktop 10.6.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated
- Clearly defined error bars
State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on [statistics for biologists](#) may be useful.

Software and code

Policy information about [availability of computer code](#)

Data collection

No primary data collection was carried out for this analysis.

Data analysis

This analysis was carried out using R version 3.5.0. The main geostatistical models were fit using R-INLA version 18.07.12. All code used for these analyses is publicly available online at <http://ghdx.healthdata.org/>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The findings of this study are supported by data that are available in public online repositories, data that are publicly available upon request from the data provider, and data that are not publicly available due to restrictions by the data provider and which were used under license for the current study. A detailed table of data

sources and availability can be found in the Supplemental Information.

Administrative boundaries were retrieved from the Global Administrative Unit Layers (GAUL) dataset, implemented by FAO within the CountrySTAT and Agricultural Market Information System (AMIS) projects. Land cover was retrieved from the online Data Pool, courtesy of the NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota. Lakes were retrieved from the Global Lakes and Wetlands Database (GLWD), courtesy of the World Wildlife Fund and the Center for Environmental Systems Research, University of Kassel. Populations were retrieved from WorldPop.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was calculated as the number of unique data source location-pairs with observations of years of educational attainment for males and females. This sample size is reported in the methods section as "We compiled a database of survey and census datasets that contained geocoding of subnational administrative boundaries or GPS coordinates for sampled clusters. These included datasets from 528 sources (see Supplementary Table 2).
Data exclusions	As described in the methods section of the main text (with greater detail in the Supplementary Information) we did not include data from Western Sahara or French Guiana, due to lack of availability. 42 of 105 included countries have only subnational administrative level data. Employing a Bayesian model-based geostatistical framework and synthesizing geolocated data from 517 household and census datasets, this analysis provides subnational estimates of mean years of education and proportion of the population attaining key levels of education for women of reproductive age (15-49), women age 20-24, and equivalent male age-bins between 2000-2017 in low- and middle-income countries. This includes 105 countries across all low- and middle-income countries. Countries were selected for inclusion in this analysis using the Socio-demographic Index (SDI) published in the GBD46. The SDI is a measure of development that combines education, fertility, and poverty. Countries in the Middle, Lower-Middle, or Low SDI quintiles were included, with several exceptions. Albania, Bosnia, and Moldova were excluded despite Middle SDI status due to geographic discontinuity with other included countries and lack of available survey data. Libya, Malaysia, Panama, and Turkmenistan were included despite Higher-Middle SDI status to create better geographic continuity. We do not estimate for American Samoa, Federated States of Micronesia, Fiji, Kiribati, Marshall Islands, Samoa, Solomon Islands, or Tonga, where no available survey data could be sourced.
Replication	This is an observational study using many years of survey and surveillance data and could be replicated.
Randomization	This analysis is an observational mapping study and there were no experimental groups.
Blinding	Blinding was not relevant to this study, as it was an observational study using survey and surveillance data.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging