Trends of Education Inequality and Gender Disparities: Mean Education Levels, Distributions, and Metrics of Inequality for 195 Countries

Hunter York, Emmanuela Gakidou

Background

Educational attainment is an essential social determinant of health outcomes. The educational level of women of reproductive age has been shown to influence child mortality, maternal mortality, and fertility (Cleland et al., 1988; Gakidou et al., 2010). Past research has established educational attainment, among other aspects of human capital, as a key driver in the economic productivity of a country and of household income levels (Lim et al., in press; Lucas, 1988; Case et al., 2002). There have been several attempts to characterize the educational composition of countries via school enrollment ratios, censuses, and surveys (Barro, 1991; Barro & Lee, 2012; Cohen and Soto, 2007). While these papers have allowed for comparisons of *between*-country inequality in access to education, few attempts have been made to produce comprehensive measures of inequalities *within* countries, and no attempts have been made to produce mean attainment estimates for individuals under 15 years of age.

Sustainable Develoment Goal (SDG) 4 is outlined in the 2015 report by United Nations to "ensure inclusive and equitable quality education […] for all." To date, no published estimates of education distributions accurately capture single-year distributions of education. Instead, they report the percentage of the population in larger bins of education (completed primary, incomplete primary, completed secondary, etc.) for ages 15 and above. For more accurate estimates of education inequality and gender disparities including assessment of SDG target 4.1 to "ensure that all girls and boys complete free, equitable and quality primary and secondary education […]," it is necessary to capture the granularity afforded by single year bins. Granularity at the single year level of distributions of educational attainment ensures the accurate capture of dropout rates to allow for precision changes in policy to achieve increased school completion rates. Further, estimates of mean educational attainment for younger age groups allow for more immediately actionable areas of policy improvement.

This study seeks to a) quantify disparities in educational attainment at the single-year level across all years 1950-2017, b) compare disparities within and between genders across all countries, and c) produce the first estimates of mean educational attainment for age groups 5-9 and 10-14. Using directly modeled estimates of mean years of educational attainment and percentage of the population with no education, this study leverages a machine learning algorithm to derive full distributions of single-year proportions of educational attainment for 195 countries, ages 5-99, both sexes, and years 1950-2017 in order to produce the most exhaustive metrics of education inequality to date.

Methods

Estimates of mean levels of educational attainment were based on a compilation of 7,486 unique source-years comprising of censuses and nationally and sub-nationally representative household surveys. Each source included information on educational attainment by country, year, sex, and 5-year age group, as well as data regarding the distribution of attainment within the population of interest. Ideally, each data source provided information on the distribution of education in single years of attainment. Where years of schooling data were aggregated into multi-year bins, we utilized a database of 2,031 sources reporting single-years of schooling to split these binned data into complete single-year distributions from 0 to 18 years based on the

average of the 12 closest distributions in space and time. Data are top-coded to 18 years, as it is a common choice among providers of single-year education data. We used means of these source-specific distributions and percentages with no education as the direct inputs in our estimation procedure.

A multi-stage modeling framework was used to model mean levels of educational attainment and proportion of the population with no education over space and time. Age-cohort imputation was used to project observed cohorts through time. An age-period model was then used to extrapolate outside of observed birth cohorts and to complete time series. Gaussian process regression was used to synthesize these models with the input data, ensuring a good model fit where data are available, and producing estimate uncertainty.

Using modeled mean years of education and proportion with 0 years of education for each location, age, sex, and year combination, a K nearest neighbors machine learning algorithm was implemented. This algorithm compares the modeled means and proportions with no education to the aforementioned database of educational distributions in order to construct a time series of modeled distributions for years with missing data.

[Methods will be expanded greatly in final paper]

Results [In progress]

Concerning the global trend, mean years of education are increasing for both males and females. The same holds true for all 22 regions used by the Global Burden of Disease study. Estimates of inequality show similar progress, with Gini coefficients decreasing ubiquitously for all regions between 1950 and 2017 (figure 1).

Ratios of inequality between men and women likewise show regional variation but a broader range of progress. Though all super-regions are showing progress towards gender equality in age-standardized measurements of mean years of education, several super-regions show that inequality is actually improving for women more slowly than it is for men, causing the ratio of inequality (as measured by the Gini coefficient) for women to men to increase in some locations (figure 2).

These gender disparities are increasingly due to disparities in access to education in secondary and above. As countries are converging on gender equality for primary school as indicated by progress in mean educational attainment for ages 5-9, gender equality for ages 10-14 and 15-19 has failed to follow suit at a similar pace (figure 3). Super-regions such as Sub-Saharan Africa and North Africa and the Middle East show stymied progress in secondary and

tertiary attainment as girls continue to lag behind boys from age 10 onwards.

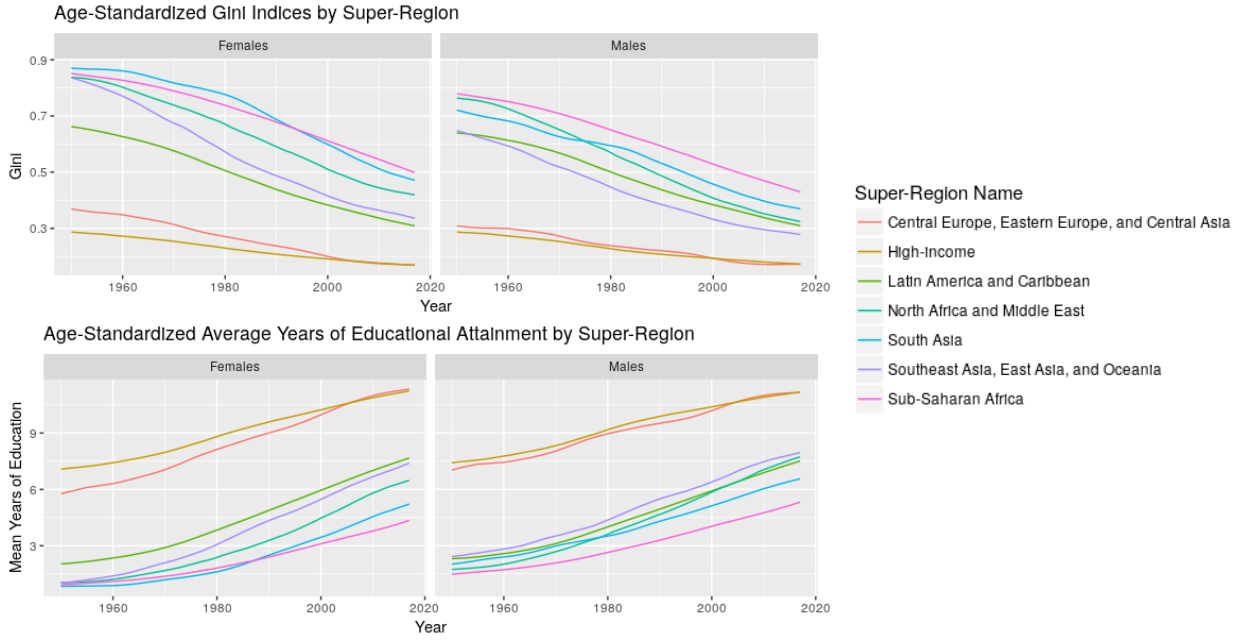


Figure 1. Age-Standardized Gini Coefficients and Average Levels of Educational Attainment by Super-Region, ages 15-99.

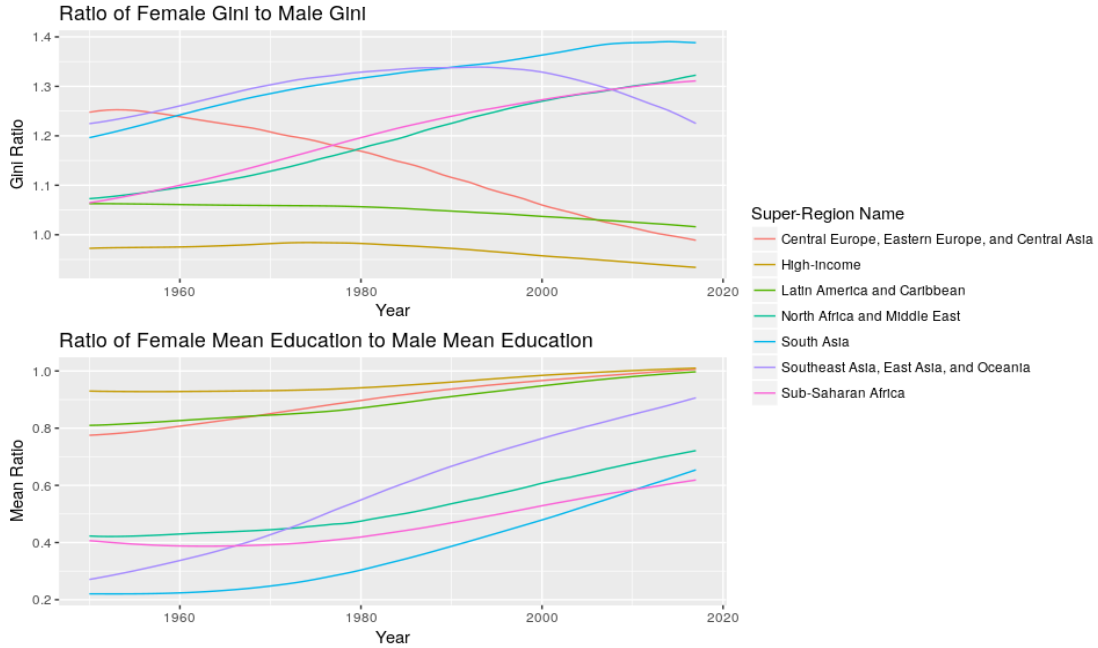
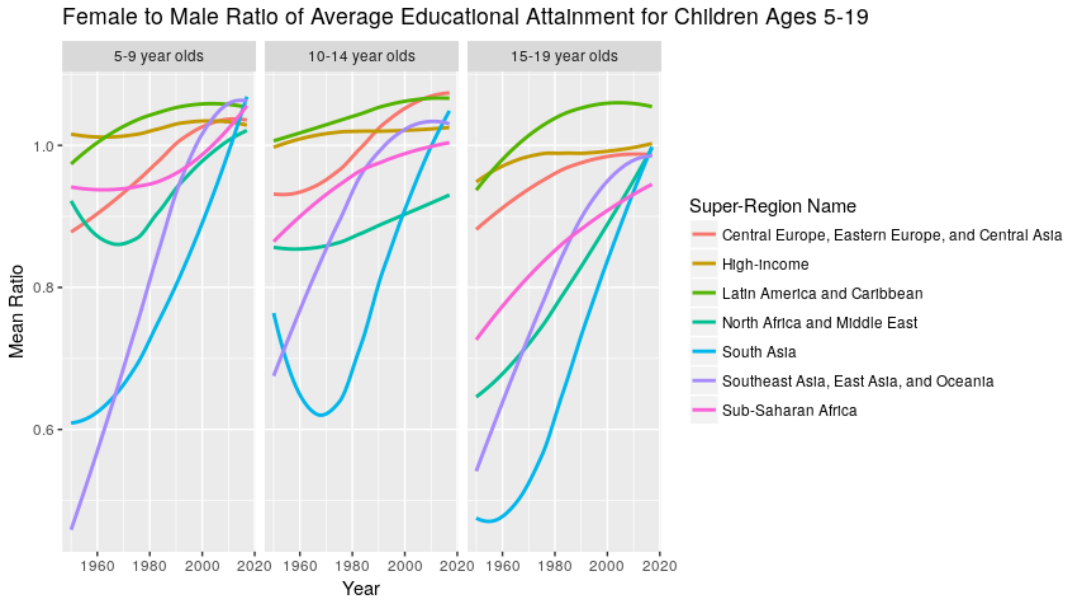


Figure 2. Age-Standardized Ratios of Mean Years of Educational Attainment and Gini Coefficients, ages 15-99.



Interpretation

Globally, access to education has steadily improved since 1950, but the story of how these gains have been made is less straightforward. Though improvements of mean educational attainment and reductions in inequality are ubiquitous at the macro level, the pace at which advances are being seen are not equitable across regions, countries, sexes, and time. In particular, disparities remain greater for women for many countries as compared to men, and there is a wide variety of the rates of change in gender disparity metrics, even at the super-region level.

While this paper has provided the most comprehensive set of estimates of average years of education and distributions of education, the metrics of inequality used to describe these distributions do suffer from some limitations. The Gini coefficient suffers from unstable estimates for demographic groups where mean years of schooling is near 0 or 18 due to it being a measure of relative inequality. Alternatives such as variance and quantile-based measurements might be more appropriate, and there are over 12 variations of Gini indices that may be better suited to measuring education inequalities (Yitzhaki, 1998). Further, our estimates for ages 5-14 are still being refined. As education continues to be accumulated by most individuals for these ages, single-age estimates may prove to be more useful than 5-year bins. Further, estimates at the single-age level will allow for metrics of inequality to be produced.

This paper calls into question commonly used metrics of gender disparities such as ratios of mean educational attainment by demographic group. Though inequality metrics, and the Gini coefficient in particular, have existed for over a century, studies rarely capitalize on their strengths to produce comparisons of inequality across demographic groups. While traditional metrics can elucidate between-group disparities, they often leave out vital information on equitable access to education within groups. Despite much progress having been made on the global stage towards equal access to education, these estimates, and especially those for ages 5-14, demonstrate that there are still barriers to accessing basic human right, leading to between- and within-country inequalities.

References

Barro, R. J. (1991). Economic Growth in a Cross Section of Countries. *The Quarterly Journal of Economics*, *106*(2), 407–443. https://doi.org/10.2307/2937943

Barro, R. J., & Lee, J. W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, *104*, 184–198. https://doi.org/10.1016/j.jdeveco.2012.10.001

Case, A., Lubotsky, D., & Paxson, C. (2002). Economic Status and Health in Childhood: The Origins of the Gradient. *American Economic Review*, *92*(5), 1308–1334. https://doi.org/10.1257/000282802762024520

Cleland, J., & Rodríguez, G. (1988). The Effect of Parental Education on Marital Fertility in Developing Countries. *Population Studies*, *42*(3), 419–442. https://doi.org/10.1080/0032472031000143566

Cohen, D., & Soto, M. (2007). Growth and human capital: good data, good results. *Journal of Economic Growth*, *12*(1), 51–76. https://doi.org/10.1007/s10887-007-9011-5

Gakidou, E., Cowling, K., Lozano, R., & Murray, C. J. (2010). Increased educational attainment and its effect on child mortality in 175 countries between 1970 and 2009: a systematic analysis. *The Lancet*, *376*(9745), 959–974. https://doi.org/10.1016/S0140-6736(10)61257-3

Gore, F. M., Bloem, P. J., Patton, G. C., Ferguson, J., Joseph, V., Coffey, C., … Mathers, C. D. (2011). Global burden of disease in young people aged 10–24 years: a systematic analysis. *The Lancet*, *377*(9783), 2093–2102. https://doi.org/10.1016/S0140-6736(11)60512-6

Human Capital Collaborators (In Press). Measuring Human Capital: a Systematic Analysis of 195 Countries and Territories, 1990-2016. *Lancet*.

UN. Transforming our world: the 2030 Agenda for Sustainable Development. New York, NY, USA: United Nations, 2015.

Yitzhaki, S. (1998). More than a Dozen Alternative Ways of Spelling Gini 1. Retrieved September 13, 2018, from /paper/More-than-a-Dozen-Alternative-Ways-of-Spelling-Gini-Yitzhaki/07e17b07caed70d513379129a3fe664cdb606efd