

Human Capital Index Plus (HCI+)

Reproducibility Package for the Human Capital Index Plus

Overview

The Human Capital Index Plus (HCI+) converts health, education, and labor market indicators into a common metric of expected lifetime productivity, measured as the expected contribution to log earnings. The index combines three components:

1. **Health Component** - Captures productivity effects of surviving to working age and achieving adequate early-life physical growth
2. **Education Component** - Aggregates human capital accumulated during formal schooling
3. **On-the-Job Learning Component** - Measures human capital accumulation after age 18 from work experience

Directory Structure

1. **01_raw_data** contains the raw data for the project for each indicator.
2. **02_programs** contains the main replication file for the project. Most of the code is written in Stata. The entire project can be run by running the file `run_HCIPlus.do`. Paths can be automatically set by running the `profile.do` file, which allows user to set the path by locating a `.here` file in their local machine in the root directory of the project.
3. **03_output**. This folder contains a number of final output files in either `.csv` or Stata `.dta` format. The final dataset is `03_output/hci_plus/hci_plus_index.dta`. The dataset is also available in `.xlsx` format.

Instructions to Replicators

- Clone the repository to your local machine.
- Please run `run_HCIPlus.do` to generate the data. This file will run all of the code to generate the data. The replicator should expect the code to run for around 5-10 minutes.

- There should be no need to change the working directory. The code should provide a prompt to change the working directory if necessary.

HCI+ Mathematical Formula

We write lifetime human capital as the sum of three components using a log-form representation consistent with Mincer-style interpretations:

$$HCI^+ = HCI_{Health} + HCI_{Education} + HCI_{OTJ}$$

Values for males and females combined ($_mf$) are shown in the formulas below. The data also contains separate values for males ($_m$) and females ($_f$).

Health & Nutrition Component

$$HCI_{Health} = 0.5 \times 0.6528 \times Surv_{15-60} + 0.5 \times 0.3468 \times Not_Stunt$$

We use adult survival (ages 15–60) and stunting as health measures. Both are measures of underlying latent health. Because both capture latent health and adult productivity, the two are each given 50% weight as in Kraay (2019).

Rationale: Following Kraay (2019) and Weil (2007), the earnings penalty per centimeter of height loss is approximately 3.4%. Mortality and stunting are associated with height deficits of 19.2 cm and 10.2 cm, respectively, implying log earnings penalties of 0.65 and 0.35.

Education Component

$$HCI_{Education} = 0.12 \times EYS_{pp} \times \frac{HLO}{625} + 0.12 \times EYS \times \frac{HLO}{625} + 0.16 \times 4 \times Tertiary$$

This can be written more compactly by factoring out HLO/625:

$$HCI_{Education} = \frac{HLO}{625} \times (0.12 \times EYS_{pp} + 0.12 \times EYS) + 0.16 \times 4 \times Tertiary$$

Note: For tertiary returns, we assume zero additional return after a country reaches 50% tertiary completion.

On-the-Job Learning Component

Employment indicators are split for youth (18-24) and working age (25-64) groups. In both age groups, the following are considered: labor force participation rates, unemployment rates, and the share in wage employment.

Depreciation rate (loss of skills when not working): $\delta = 0.0125$ (1.25% per year)

Youth Factor (18-24)

Define effective labor force participation and non-participation:

$$P_{ya} = LFP_{ya} \times (1 - U_{ya}) \quad (\text{Effective participation})$$

$$N_{ya} = 1 - P_{ya} - Tertiary \quad (\text{Non-participation in either work or schooling})$$

Returns during youth are:

$$\theta_{ya} = 0.039 \times wage_emp_{ya} + 0.02 \times (1 - wage_emp_{ya})$$

$$HCl_{Youth} = 7 \times (\theta_{ya} \times P_{ya} - \delta \times N_{ya})$$

Working Age Factor (25-64)

Define effective labor force participation and non-participation:

$$P_{wa} = LFP_{wa} \times (1 - U_{wa}) \quad (\text{Effective participation})$$

$$N_{wa} = 1 - P_{wa} \quad (\text{Non-participation})$$

Returns during working age are:

$$\theta_{wa} = 0.03 \times wage_emp_{wa} + 0.018 \times (1 - wage_emp_{wa})$$

The working-age component is:

$$HCl_{WA} = T \times (\theta_{wa} \times P_{wa} - \delta \times N_{wa})$$

Definition of T:

$$T = 0.5 \times LE_{25-64}$$

T measures the expected working-life for adults aged 25–64. The maximum lifespan during that period is 40 years. Assuming the population is uniformly distributed, a randomly selected 25-64 year old in a country with no mortality will be 45 years old and have up to 20 years of experience.

To avoid using two alternative statistics on adult mortality, the country-specific value for LE_{25-64} is computed using the following approximation:

$$LE_{25-64} = e^{0.5 \times ASR + 3.18}$$

which is obtained by regressing life expectancy on the adult survival rate data in a Poisson regression (pseudo $R^2 = 0.989$).

Summary of Returns Used in HCI+ Scoring

Component	Return / Parameter	Source / Notes
Health	3.4% earnings penalty per cm height loss	Kraay (2019); Weil (2007)
Health	19.2 cm height loss (mortality), 10.2 cm (stunting)	Kraay (2019); Weil (2007)
Education	12% per learning-adjusted year (pre-primary + LAYS)	World Bank analysis based on Schoellman (2012), Jedwab et al. (2023), and Gethin's (2025) meta-analysis
Education	16% per tertiary year ($\times 4$ years)	Psacharopoulos & Patrinos (2018)
On-the-job learning (youth)	3.9% per year (wage); 2.0% (non-wage)	World Bank analysis using I2D2 & GLD collections
On-the-job learning (adults)	3.0% per year (wage); 1.8% (non-wage)	World Bank analysis using I2D2 & GLD collections
On-the-job learning	1.25% annual depreciation if not working	Dinerstein et al. (2022)

Comparing the Original HCI and HCI+ Scores

The HCI+ scale is closely related to the scale of the previously published World Bank HCI.

The original HCI compared the level of human capital of a country to the benchmark of an ideal society with perfect health and full schooling to age 18. Because human capital accumulates, each component was multiplied over time:

$$\text{Human Capital} = e^{\beta_1 \cdot \text{Health}} \times e^{\beta_2 \cdot \text{Education}} \times e^{\beta_3 \cdot \text{OTJ}}$$

Converting to logarithmic form:

$$\log(\text{Human Capital}) = \beta_1 \cdot \text{Health} + \beta_2 \cdot \text{Education} + \beta_3 \cdot \text{OTJ}$$

This transformation means that contributions of health, education, and on-the-job learning are **additive**. This is the scale used in the HCI+.

Converting HCI+ to HCI scale: An interested user can easily convert the HCI+ into units of the original HCI (0-1 scale). Suppose a country has an HCI+ score of 200. The ideal society has a score of 325. The comparable score on the HCI 0-1 scale is:

$$\frac{e^{2.00}}{e^{3.25}} = 0.29$$

Key Columns in the Data

Key dataset is 03_output/hci_plus/hci_plus_index.dta. This dataset contains the Human Capital Index Plus (HCI+) for each country. The HCI+ is the sum of the component scores for health, education, and on-the-job learning.

Only total values are shown in the table below. However, values for females/males are also available in the data. The naming convention for these columns is as follows:

- `_mf` refers to the total (male + female)
- `_m` refers to the value for males
- `_f` refers to the value for females

Health Component Variables

Column Name	Description	Min	Max
surv_15to60_mf_2025	Adult survival rate (ages 15-60)	0	1

Column Name	Description	Min	Max
nostu_mf_2025	Probability of not being stunted (under 5)	0	1

Education Component Variables

Column Name	Description	Min	Max
eys_pp_mf_fill_2025	Expected years of pre-primary schooling (EYS)	0	3
eys_sa_mf_fill_2025	Expected years of schooling (primary + secondary)	0	12
hlo_mf_fill_2025	Harmonized learning outcomes (HLO)	200	625
lays_sa_mf_fill_2025	Learning-adjusted years of schooling (LAYS)	0	12
ter_ya_mf_fill_2025	Fraction completing tertiary education	0	0.5

On-the-Job Learning Component Variables

Youth (18-24):

Column Name	Description	Min	Max
lfp_ya_mf_fill_2025	Labor force participation rate (youth)	0	1
emp_ya_mf_fill_2025	Employment rate (youth)	0	1
shr_wemp_ya_mf_fill_2025	Share in wage employment (youth)	0	1

Working Age (25-64):

Column Name	Description	Min	Max
lfp_wa_mf_fill_2025	Labor force participation rate (working age)	0	1
emp_wa_mf_fill_2025	Employment rate (working age)	0	1
shr_wemp_wa_mf_fill_2025	Share in wage employment (working age)	0	1

HCI+ Component Scores

Column Name	Description
hci_health_mf_2025	Health component score
hci_education_mf_2025	Education component score
hci_otj_mf_2025	On-the-job learning component score
hcip_mf_2025	Human Capital Index Plus (HCI+) total score

Metadata Columns

Column Name	Description
wbcountryname	Country name
wbcode	Country code (ISO3)
year	Year of the data
wbregion	World Bank region
wbincomegroup	World Bank income group

Background: Three Components of HCI+

1. Health Component

The health component captures the productivity effects of surviving to working age and achieving adequate early-life physical growth. Two proxies are used:

- **Adult survival rate (ages 15–60):** Sourced from UN Population Division life tables
- **Share of children under age 5 who are not stunted:** From WHO–UNICEF–World Bank Joint Malnutrition Estimates

Both indicators rely on long-run evidence linking adult height and productivity. Following Kraay (2019) and Weil (2007), the earnings penalty per centimeter of height loss is approximately 3.4 percent. Mortality and stunting are associated with height deficits of 19.2 cm and 10.2 cm, respectively, implying log earnings penalties of 0.65 and 0.35. These are weighted equally in the HCI+ scoring formula.

2. Education Component

The education component aggregates human capital accumulated during formal schooling. Data inputs include:

- **Expected years of schooling (EYS):** UNESCO UIS
- **Harmonized learning outcomes (HLO):** World Bank EduAnalytics
- **Tertiary completion:** UIS and household surveys

Returns to pre-primary and LAYS are combined into a single estimate based on World Bank analysis indicating returns of 9–12 percent per effective learning-adjusted year. The HCI+ uses **12% per LAYS** based on:

- Schoellman (2012)
- Islam–Jedwab–Romer (2023)
- Gethin’s (2025) meta-analysis of high-quality IV estimates

Tertiary returns follow Psacharopoulos & Patrinos (2018), using a **16 percent return per year** for a four-year degree.

3. On-the-Job Learning Component

The HCI+ introduces a component measuring human capital accumulation after age 18 from work experience. Inputs include:

- Labor force participation rates
- Employment rates
- Share of workers in formal wage employment

Data sources: ILOSTAT and the I2D2/GLD survey database.

Returns to experience differ by job type:

- **3.9% per year** in formal wage jobs (youth 18-24)
- **2.0% per year** in informal/non-wage work (youth 18-24)
- **3.0% per year** in formal wage jobs (adults 25-64)
- **1.8% per year** in informal/non-wage work (adults 25-64)

Skills depreciation: A depreciation rate of **1.25% per year** is applied to periods out of work, following Dinerstein, Megalokonomou, and Yannelis (2022).

Limitations of the HCI+

Like any summary measure of complex phenomena, the HCI+ provides a simplified representation of human capital that cannot capture every dimension of health, learning, and work experience.

Independence Assumptions and Correlations

The mathematical structure of the HCI+ assumes that its components are largely independent of one another. In practice, these independence assumptions do not hold perfectly—children who are stunted are more likely to complete fewer years of schooling and score lower on learning assessments. The HCI+ does not explicitly account for these correlations because reliable, cross-country data on how human capital components interact is rarely collected systematically.

Decomposability and Subgroup Analysis

The additive structure of the HCI+ allows the index to be decomposed by pillar (health, education, on-the-job learning) and by individual components within each pillar. While the report presents separate male and female HCI+ scores, these are computed independently rather than decomposed from a single aggregate measure.

Job Quality Measurement

The HCI+ uses the share of the population in wage employment as a proxy for job quality. Other measures of job quality (firm size, job type) were not used due to data limitations.

Morbidity and Health Measurement

The HCI+ does not include a direct measure of adult morbidity. However, the index captures the effects of poor health on productivity through:

- Adult survival component (mortality reflects underlying health conditions)
- Labor force participation rates (illness/disability reduces participation)
- Unemployment rates (chronic conditions affect employability)

Data Quality and Coverage

The HCI+ relies on survey estimates from international organizations (UNESCO, WHO, ILO, UN Population Division) to ensure data is properly harmonized and complete country coverage. Where survey data are unavailable or outdated, modelled data are sometimes used to impute values.

License

The data are licensed under a Creative Commons/CC-BY-4.0 license.

Summary of Availability

- ☒ All data **are** publicly available.
- ☐ Some data **cannot be made** publicly available.
- ☐ **No data can be made** publicly available.

Data Availability Statement

All input datasets required to reproduce the HCI+ outputs are listed below with structured metadata: filename(s), source, URL, access date (when recorded in do files), suggested short citation, license/access status, access instructions or subset details used by scripts, and whether the file is included in the repository.

Filename(s)	Source	URL	Access date	Citation (short)	License / access	Access instructions / subset used	Included in repo?
hci_health_nutrition/WPP2024_MORT_F04_1_LIFE_TABLE_SURVIVORS_BOTH_	UN Population Division (World Population	https://population.un.org/wpp/ (mortality & demographic downloads)	Oct 15, 2025	United Nations, Department of Economic and Social	Public — UNPD terms	Imported sheets: “Estimates”; used to compute	Yes

Filename(s)	Source	URL	Access date	Citation (short)	License / access	Access instructions / subset used	Included in repo?
SEXES.xlsx ci_health_nutrition/WPP2024_MORT_F04_2_LIFE_TABLE_SURVIVORS_MALE.xlsx ci_health_nutrition/WPP2024_MORT_F04_3_LIFE_TABLE_SURVIVORS_FEMALE.xlsx ci_health_nutrition/WPP2024_GEN_F01_DEMOGRAPHIC_INDICATORS_COMPACT.xlsx	Prospects 2024)			Affairs, Population Division (2024). World Population Prospects 2024. Data files.		surv_15to60 .	
hci_health_nutrition/P_Data_Extract_From_Health_Nutrition_and_Population_Statistics.xlsx	World Bank Databank (Health, Nutrition and Population Statistics; UNICEF–WHO–WB	https://databank.worldbank.org/source/health-nutrition-and-population-statistics#	Jun 15, 2025	World Bank. Health, Nutrition and Population Statistics (UNICEF–WHO–World Bank JME).	Public — World Bank Data Terms	final outputs use 2009–2024.	Yes

Filename(s)	Source	URL	Access date	Citation (short)	License / access	Access instructions / subset used	Included in repo?
	JME)			Data extract Jun 2025.			
hci_on_job_learning/lfp_survey_v2.dta hci_on_job_learning/lfp_modelled_v2.dta hci_on_job_learning/unemp_survey_v2.dta hci_on_job_learning/unemp_modelled_v2.dta hci_on_job_learning/emp_structure_survey_v2.dta hci_on_job_learning/emp_structure_age_modelled_v2.dta hci_on_job_learning/workingage_pop.dta hci_on_job_learning/	ILOSTAT (International Labour Organization)	https://ilostat.ilo.org/	Dec 2025	ILO (ILOSTAT) database.	Public — ILOSTAT terms	Do files read the provided .dta survey and modeled files. Reproduce by downloading LFP/unemployment/employment-structure indicators	Yes

Filename(s)	Source	URL	Access date	Citation (short)	License / access	Access instructions / subset used	Included in repo?
ng/pop_data_modelled.dta							
hci_education/eys_data_for_HCI_Team_Dec_11_2025_v2.dta hci_education/hlo_combined_2025_2020.dta hci_education/rawfull_tertiary.dta	World Bank — Education Analytics team (internal dataset)	internal (contact Education Analytics team)	Dec 2025	World Bank, Education Analytics Team (2025). EYS/HLO/Tertiary datasets for HCI.	Internal — request access from Education Analytics team	Do files load these .dta files directly and select many EYS/HLO variables (see 02_programs/hci_education/1.eys.do, 1.hlo.do, 1.tertiary.do). Tertiary logic uses hierarchy TCR > TPR > GGR > GER (UIS indicators).	Yes (internal)
01_raw_data/misc/loggd_p_wdi.dta	World Bank (World Development Indicators)	https://databank.worldbank.org/source/world-development-indicators	July 2025	World Bank. World Development Indicators (GDP per	Public — World Bank Data Terms	Merged by wbcode and year in multiple do files to	Yes

Filename(s)	Source	URL	Access date	Citation (short)	License / access	Access instructions / subset used	Included in repo?
		indicators		capita).		provide log GDP per capita for years up to 2024.	
01_raw_data/misc/masterdata.dta	World Bank (project master country metadata)	internal (repo)	July 2025	Project master metadata file (World Bank).	Internal (project)	Used as canonical country-year panel for merges (wbcode, year); included in repo.	Yes
hci_education/rawfull_tertiary.dta (UIS indicators referenced: GGR.6T7, GER.5T8)	World Bank GMD survey data; UNESCO Institute for Statistics (UIS)	https://databrowser.uis.unesco.org/	Dec 2025	World Bank GMD survey data; UNESCO Institute for Statistics. UIS Data.	Public — UIS terms		Yes
hci_health_nutrition/LHCI_component_data_reviewed.dta hci_on_job_le	World Bank internal (country-team reviewed)	internal (repo)	July 2025	World Bank internal — country-team reviewed	Internal (project)	Internal review files included in repo; contact	Yes

Filename(s)	Source	URL	Access date	Citation (short)	License / access	Access instructions / subset used	Included in repo?
arning/LHCI _component_ data_review ed.dtaLHCI_ component_d ata_reviewe d.dta	updates)			updates.		authors for details on review notes and provenance.	
metadata/WB _CLASS_FY26 .xlsx	World Bank	internal (repo)	July 2025	World Bank classificatio n file.	Internal	Used for membership /classificatio n metadata.	Yes