Approximating Children's Potential Environmental Exposures with Cross-Sectional Surveys

Messy matching across space and time with MICS and EM-DAT

Emily Hannum and Yujie Zhang

Penn Arts & Sciences
University of Pennsylvania

Stata Sociology Virtual Symposium

20 October 2026



This talk

- **The problem:** research on children's cumulative environmental and climatic exposures requires life-course data linked to location. In LMICs, such data are rare.
- **The approach:** approximate cumulative potential exposures from geocoded **cross-sectional surveys**, which are widely available, by
  1. aligning each child's life course with **dated hazard events**;
  2. summing exposure over **developmental windows**;
  3. **testing the key assumption** that the child has not moved.
- **The tools:** the **Stata commands** we used for the linkage of survey and disaster records.
- **The illustration:** disasters and education in seven Asian countries (MICS6 and EM-DAT).

Outline

1. Climate hazards and children
    a. Global context
    b. Climate hazards and education
    c. The CECI project
    d. A data challenge in LMICs
2. This paper: disasters and education in Asian LMICs
3. Constructing and checking the exposure measures
4. Uses, limitations, and conclusions

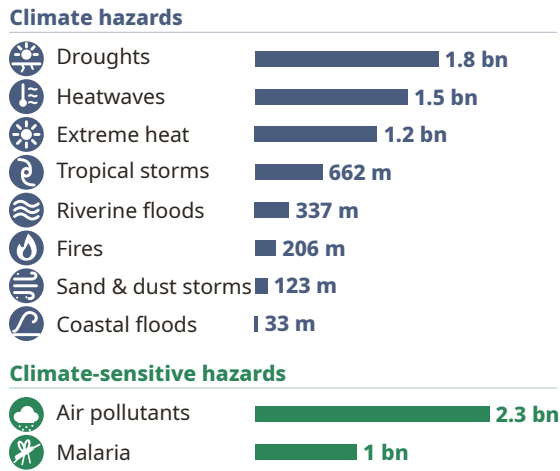
Climate hazards and children

a Global context

Children are highly exposed to climatic hazards

More than a billion children face at least three overlapping climate hazards.

Source: UNICEF, The Children's Climate Risk Report 2026, p. 12.



Approximating exposure from cross-sections

Children are highly vulnerable when exposed to hazards

1. **Physically:** less able to withstand and survive shocks such as floods, droughts, severe weather, and heatwaves.
2. **Physiologically:** toxic substances, such as lead and other pollution, affect children more than adults, even at lower doses.
3. **Disease:** at greater risk of death than adults from diseases likely to be worsened by climate change, such as malaria and dengue.
4. **Over the life course:** [deprivations from] climate and environmental degradation at a young age can mean a lifetime of lost opportunity.

Source: UNICEF (2021), The Climate Crisis Is a Child Rights Crisis.

LMICs are particularly important to study

"

[There is an expectation that] the largest health risks of climate change will occur in LMICs, where rising ambient temperatures, changing precipitation patterns, and rising sea levels are increasingly affecting air, water, and food systems. Such changes will affect the geographic and seasonal patterns of major causes of morbidity and mortality—particularly undernutrition, diarrheal diseases, malaria, dengue, and injuries—and will disproportionately affect the most vulnerable population groups…

—Ebi & Luchters (2021), "Invited perspective: Most affected by climate change; least studied," Environmental Health Perspectives 129(11): 111301. doi.org/10.1289/EHP10384

Climate hazards and children

b Climate hazards and education

Educational disruption is poorly understood

> *Children and youth are vulnerable to the threat of climate change, yet potential non-health impacts, such as those on global education outcomes, are not well understood. [This cross-disciplinary review finds] that climate change is probably already undermining education outcomes for many populations worldwide via multiple direct and indirect pathways.*

—Prentice et al. (2024, p. 214), "Education outcomes in the era of global climate change," Nature Climate Change 14: 214–224. doi.org/10.1038/s41558-024-01945-z

UNICEF's *Learning Interrupted Report* (2025)

At least 1 in 7 students



had schooling disrupted by climate hazards in 2024.

74% of affected students



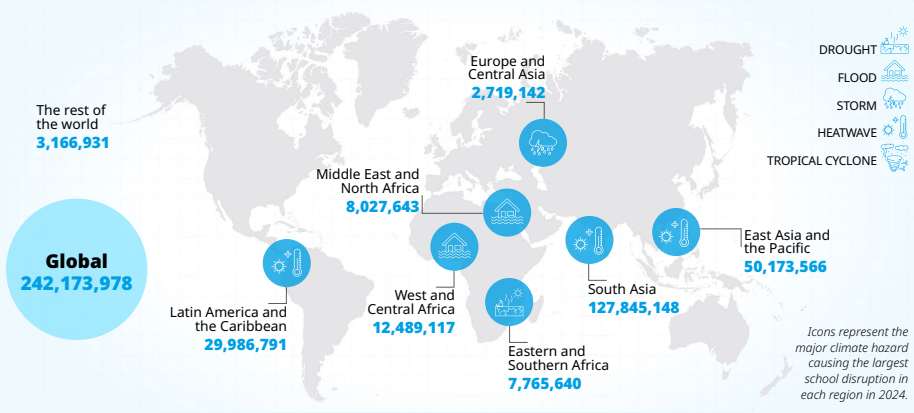
are in low- and lower-middle-income countries.

Source: UNICEF (2025), Learning Interrupted, p. 1.

- Globally, at least **242 million students**—from pre-primary to upper secondary—experienced school disruptions due to climate events in 2024.
- **85 countries or territories** saw schools affected; 23 had multiple rounds of disruption.
- At least **20 countries** had nationwide school disruptions due to climate-induced disasters.

Approximating exposure from cross-sections

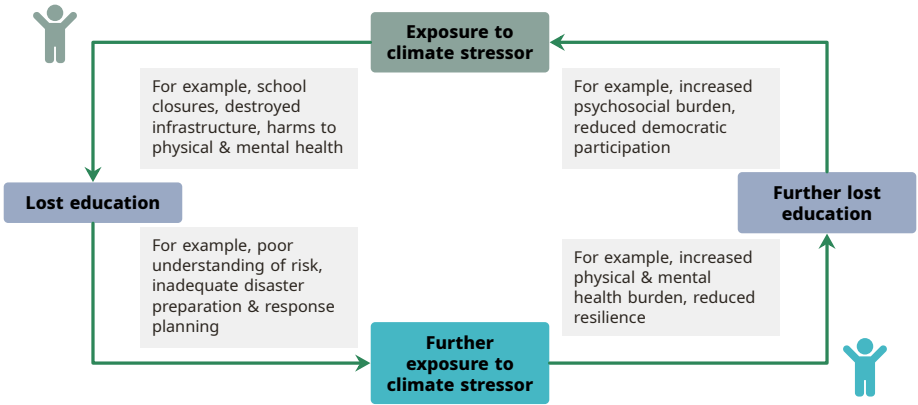
Number of students affected by climate-related school disruptions in 2024



Source: UNICEF (2025), Learning Interrupted, p. 4, Figure 2. Note: By UNICEF programme region.

Approximating exposure from cross-sections

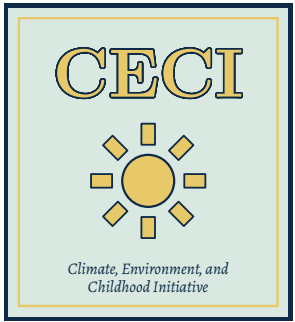
Climate stressors and education: a reinforcing cycle



Source: adapted from Prentice et al. (2024, p. 221), "Education outcomes in the era of global climate change," Nature Climate Change 14: 214–224. doi.org/10.1038/s41558-024-01945-z

Climate hazards and children

c The CECI project



Climate, Environment, and Childhood Initiative

Investigating climatic and environmental impacts on children.

ceci.sas.upenn.edu

Penn Global

NSF #2230615

SSRC / Sloan #20257924

Climate, Environment, and Childhood Initiative

Link child population and environmental data, to:

1

Investigate children's **exposures** to climatic/environmental risks, with attention to social stratification.

2

Investigate children's **ill effects** when exposed to climatic/environmental risks, with attention to social stratification.

3

Investigate **inputs that buffer** impacts of climatic/environmental risks, and how these are shaped by childhood inequalities.

Climate, Environment, and Childhood Initiative

Today's talk: Objective 2

1

Investigate children's **exposures** to climatic/environmental risks, with attention to social stratification.

2

Investigate children's **ill effects** when exposed to climatic/environmental risks, with attention to social stratification.

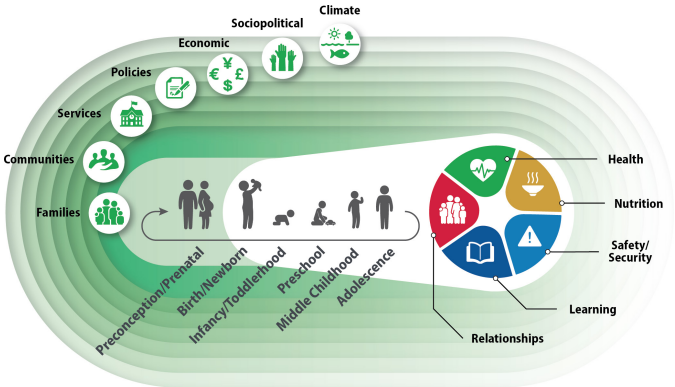
3

Investigate **inputs that buffer** impacts of climatic/environmental risks, and how these are shaped by childhood inequalities.

Climate hazards and children

d A data challenge in LMICs

A longitudinal, life-course view of development…



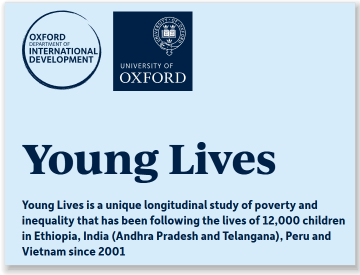
Source: Black et al. (2021), Figure 2, BMJ Global Health (CC BY).

…but few longitudinal child datasets exist in LMICs

Study (country)	Children born
INCAP (Guatemala)	1962–77
New Delhi (India)	1969–72
Pelotas (Brazil)	1982, 1993, 2004, 2015
Cebu (Philippines)	1983–84
Birth to 20 (S. Africa)	1990
Gansu GSCF (China)	1987–91
Young Lives Ethiopia, India, Peru, Vietnam	**1994–95, 2001–02**

Sources: Richter et al. (2012), "Cohort profile: The Consortium of Health-Orientated Research in Transitioning Societies," International Journal of Epidemiology 41(3): 621–626. doi.org/10.1093/ije/dyq251; Young Lives; study websites.

A multi-country exception



younglives.org.uk

Meanwhile, cross-sectional data from LMICs is plentiful

- Cross-sectional household surveys of children are far more common, and many record each child's age and location.
- Most not originally designed for exposure research, but geocoding and location information enable linkage to geo-referenced hazard data.

Source: UNICEF, A Brief Introduction to MICS.

UNICEF Multiple Indicator Cluster Surveys (MICS)

Household surveys focused on children and women

Time: 7 rounds since the mid-1990s

Place: 418 surveys, 123 countries

Scale: nationally representative, with regional and urban/rural results

mics.unicef.org

The approach

- A cross-sectional survey records each child's **age** and **location**.
- Disaster records give the **dates and places** of hazard events.
- Linking the two gives each child a **potential exposure history**: which hazards struck where they lived, and at what ages.
- The key assumption is that children have not moved since conception; the survey contains mothers' residential mobility information that sheds light on this assumption.

Outline

1. Climate hazards and children

2. This paper: disasters and education in Asian LMICs
   a. Why disasters, and why Asia?
   b. Paper summary

3. Constructing and checking the exposure measures

4. Uses, limitations, and conclusions

This paper: disasters and education in Asian LMICs

a Why disasters, and why Asia?

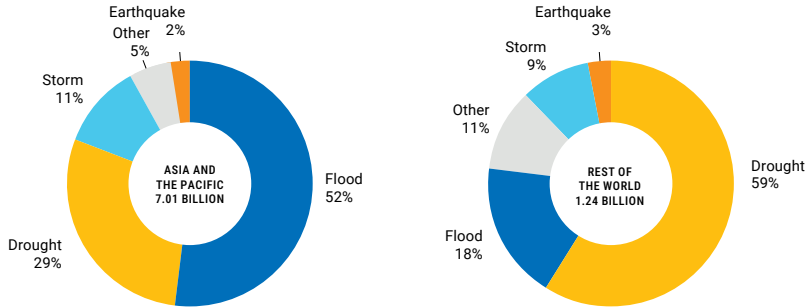
Climate change and increasingly extreme weather events…have caused a surge in natural disasters over the past 50 years[,] disproportionately impacting poorer countries…

—UN News (2021)

Asia-Pacific is the most disaster-prone region

In 2022: over 140 disasters, over 64 million people affected

Number of people affected from disasters in the Asia-Pacific region and the rest of the world, 1970–2022



Source: UN ESCAP (2023), Asia-Pacific Disaster Report 2023, p. v; Figure 1.3, p. 6.

Approximating exposure from cross-sections

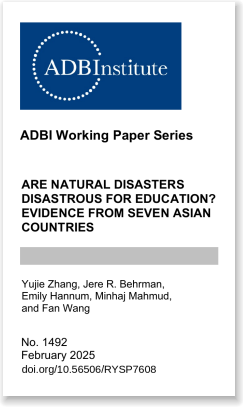
This paper: disasters and education in Asian LMICs

b Paper summary

Source for this presentation:

A study of disaster exposure and education in seven Asian countries

- **Data:** MICS6 surveys of children aged 5–17 in seven Asian countries, linked to EM-DAT disaster records.
- **Exposure:** the first 1,000 days from conception, the most recent years, and the period in between.
- **Hazards:** all disasters, floods, and severe disasters.
- **Outcomes:** school enrollment and math test scores.



Research questions and goals

Questions

1. What are the implications of exposure to natural disasters for children's educational outcomes?
2. Are there heterogeneous impacts on school enrollment and math skills for children along **gender and age** gradients?

Goals

- Provide one of the first **cross-nation and cross-disaster** analyses of the effects of multiple disruptive natural disasters on educational outcomes.
- Consider children's potential exposure histories, and distinguish the effects of **recent and early-life (first 1,000 days)** disaster shocks.

Main findings of the paper

1. **Early-life** exposure to natural disasters significantly reduces school enrollment; early exposure to floods has persistent negative effects on math scores. **Recent** floods lower enrollment, while in some contexts, enrollment among younger children increases after severe disasters.
2. Effects vary by **gender and age**: boys exposed to early-life disasters experience persistent negative enrollment effects during upper-primary-school ages, regardless of disaster types.
3. Effects also vary by **country**; pooled estimates mostly reflect Bangladesh and Pakistan, which make up about 80% of the sample.

Source: Zhang, Behrman, Hannum, Mahmud, and Wang (2025), ADBI Working Paper 1492.

Outline

1. Climate hazards and children

2. This paper: disasters and education in Asian LMICs

3. Constructing and checking the exposure measures
   a. The data
   b. Messy matching
   c. Building exposure windows
   d. Assessing the no-move assumption
   e. Assessing various disaster types

4. Uses, limitations, and conclusions

Constructing and checking the exposure measures

a The data

Two data sources

- **MICS6**: children aged 5–17 in **seven Asian countries**, each located to a subnational unit.
- **EM-DAT** (Emergency Events Database): time- and geo-coded records of **355 disasters** causing substantial human losses, 1999–2019 (floods, storms, droughts, earthquakes, extreme temperatures).
  - For each event: **start and end dates** and a list of **affected subnational areas**.

Linked: a **child-level panel of disaster exposure** by time, age, and location, constructed from cross-sectional data.

Sources: MICS (UNICEF); EM-DAT (Delforge et al. 2023).

EM-DAT raw data example

DisNo.	Disaster Type	Location	Start Year	Start Month	Start Day	End Year	End Month	End Day
2001-0563-BGD	Storm	Lalmonirhat, Nilphamari, Gaibandha, Rangpur districts (Rangpur province)	2001	10	5	2001	10	5
2005-0590-BGD	Flood	Naogaon, Joypurhat district (Rajshahi province), Gaibandha, Rangpur, Nilphamari, Dinajpur districts (Rangpur province)	2005	10	3	2005	10	20
2012-0589-NPL	Landslide	Gandaki district (Western province)	2012	9	23	2012	9	23

Each disaster affects locations at different administrative levels.

MICS6 seven-country sample

Region	Country	Survey yr.	Geo level	N
South Asia	Bangladesh	2019	District	40,617
	Nepal	2019	Region	7,824
	Pakistan	2017–19	District	71,121
East & SE Asia	Mongolia	2018	Region	7,628
	Thailand	2019	Changwat	9,608
Central Asia	Kyrgyz Republic	2018	Oblast	3,897
	Turkmenistan	2019	Region	3,776
Total children 5–17				**144,471**

Geo-identifiers differ across countries (e.g., 64 districts in Bangladesh, 97 in Pakistan).

Constructing and checking the exposure measures

b Messy matching

Linking a survey to hazard records can be “messy”

Spatial matching

- **inconsistent place names** §3b
- coarse or uneven geocoding
- **assumed residence history** §3d

Temporal matching

- inexact caregiver-reported birth dates
- hazard start and end dates

Bold: addressed in this talk (section noted). Others: see Limitations.

Constructed exposures are **cumulative potential exposures**: they rely on where the child lives now and on events recorded for large areas, so they describe conditions in the child’s area, not what happened to the child.

Geographic and temporal integration

For each location-month, construct the disaster history using:

- **Location linkage file** — one row per district (admin 2), nested in its state (admin 1); columns: country, state, and district codes and names.
- **Location-month skeleton file** — one row per district $\times$ month, from oldest child's conception to the latest interview; columns: year, month, district code and name.
- **Matching location names** — fuzzy string matching that scores the similarity of two place names (Stata commands in §3b).

"District" and "state" are our generic labels; countries call them division, province, oblast, aimag, changwat …

Geographic and temporal integration: matching steps

1. **Split** — one row per disaster $\rightarrow$ one row per disaster $\times$ place name; tag each name as state or district.
2. **Match** — with location linkage file, state names first, then district names.
3. **Review** — check every match with score < 1 by hand, modify location names and repeat steps 1 and 2.
4. **Expand** — each disaster $\times$ district over its start–end months $\rightarrow$ a file unique by disaster $\times$ district $\times$ month.

Merged, this yields measures such as

- a dummy for **any disaster** in district d in month t;
- the **number of months** district d was flooded over several years.



Stata tools for messy linkage

Command	Source	Use in this linkage
`matchit`	SSC	Fuzzy matching of two lists of names, with a **similarity score** in $[0, 1]$; matches EM-DAT place names to the linkage file
`reclink`	SSC	Record linkage on several fields at once; an alternative to `matchit`
`split`	built-in	Splits EM-DAT's location field into separate place names
`expand`	built-in	Repeats each disaster $\times$ district row over its start–end months
`merge`	built-in	Joins disaster-months onto the district-by-month skeleton

Constructing and checking the exposure measures

c Building exposure windows

Exposure is summed within life-course windows defined by developmental stage.

Three exposure windows

For each child, we sum months of disaster exposure within three life-course windows.



Why the first 1,000 days? The prenatal and early postnatal periods are conventionally viewed as sensitive periods for child development, for reasons including physical and physiological vulnerability; central-nervous-system and brain growth are rapid between 8 and 25 weeks post-conception.

Constructing and checking the exposure measures

d Assessing the no-move assumption

Testing the no-move assumption

Building a history from a snapshot assumes children have **not moved** out of their disaster-linked administrative unit since conception.

- MICS6 asks mothers a **residence-duration** question ("how long have you lived here?"). Among the linkable sample:
  - mothers resided in the current locality for **92–98%** of the years since the sampled child's birth;
  - **83–96%** of mothers had **not moved** since that birth.
- Main analyses use the full sample; **non-mover subsamples** serve as a robustness check.
- This test is possible because MICS6 asks about **residence duration**.

Source: Zhang et al. (2025), §2.2.

Constructing and checking the exposure measures

e Assessing various disaster types

The effects from natural disaster exposures on educational outcomes may be heterogeneous across disaster types.

Educational outcomes and various disasters

Exposure window	Disaster type	Enrollment	Math score
First 1,000 days	Any disaster	-0.1 pp/mo.	negative
First 1,000 days	Floods	-0.2 pp/mo.	negative
Recent / interv.	Any / severe	mixed	mixed
Recent year	Floods	>1.2 pp lower	not significant

Key findings. Early-life exposure, especially to floods, leads to lower enrollment and math scores, with differences by age, gender, and country. Recent severe events sometimes increase enrollment among young children, which may reflect schools' role as shelters or relief sites.

Selected estimates from Zhang et al. (2025), Tables 5–6. Full estimates in backup.



Outline

1. Climate hazards and children
2. This paper: disasters and education in Asian LMICs
3. Constructing and checking the exposure measures
4. Uses, limitations, and conclusions

Other uses of the approach

- Life-course, exposure-history research on children **where geocoded longitudinal data do not exist**, using cross-sectional surveys, such as:
  - MICS, as in this illustration;
  - the Demographic and Health Surveys (DHS).
- The same exposure measures can be used to study **other outcomes**, including health, nutrition, and child development.
- **Other environmental data** can replace EM-DAT. Gridded data on temperature, rainfall, or air pollution are not limited to administrative units or to events with large human losses.
- With richer location histories, it could be extended to longer windows and to **adults**, for whom a fixed-location assumption is harder to sustain.

Limitations

- **No-move assumption**: cannot capture temporary or return migration; richer location histories would sharpen exposure.
- **EM-DAT resolution**: affected areas recorded at admin levels 1–2; events are keyed to human losses rather than crop or infrastructure damage.
- **Within-country power**: differing sample composition and disaster histories limit fine, country-specific estimates.
- **Timing**: birth dates are reported by caregivers, and some EM-DAT events lack a start or end day, so exposure windows may be misaligned by a few months.

Source for points 1–3: Zhang et al. (2025), §5.

Summary

1. Where geocoded longitudinal data are missing, a geocoded **cross-sectional survey** can be used to approximate children's cumulative potential exposure to environmental hazards.
2. The steps: align each child's life course with **dated events**, sum exposure within **developmental windows**, and **test the no-move assumption**.
3. In MICS6, residence-duration data show that most mothers had not moved since the child's birth.
4. Whether the no-move assumption can be tested depends on the **residence questions a survey asks**.

Thank you!

Emily Hannum, hannumem@upenn.edu
Yujie Zhang, yjz@sas.upenn.edu