Misreporting Month of Birth: Implications for Research

Anna Folke Larsen (Dept. of Economics, University of Copenhagen)

Derek Headey (Poverty, Health & Nutrition Division, IFPRI)

William A. Masters (Friedman School of Nutrition, Tufts University)

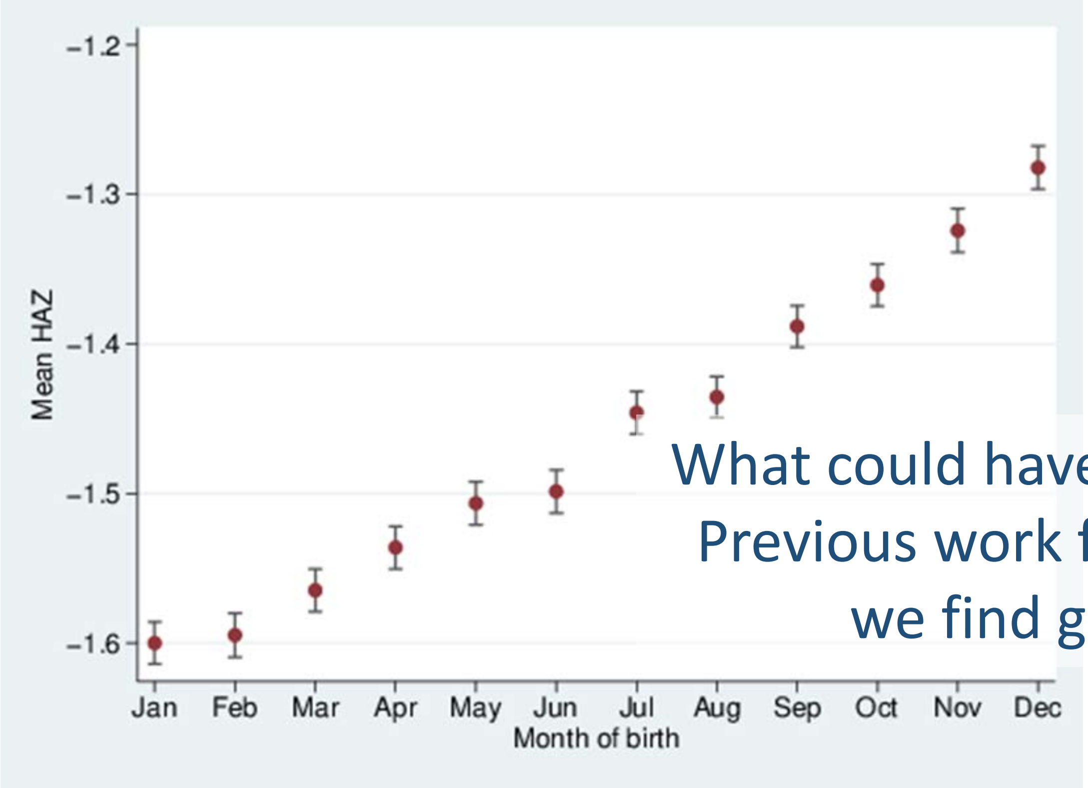
Selected paper presented at the annual meetings of the AAEA, 31 July 2017

Motivation

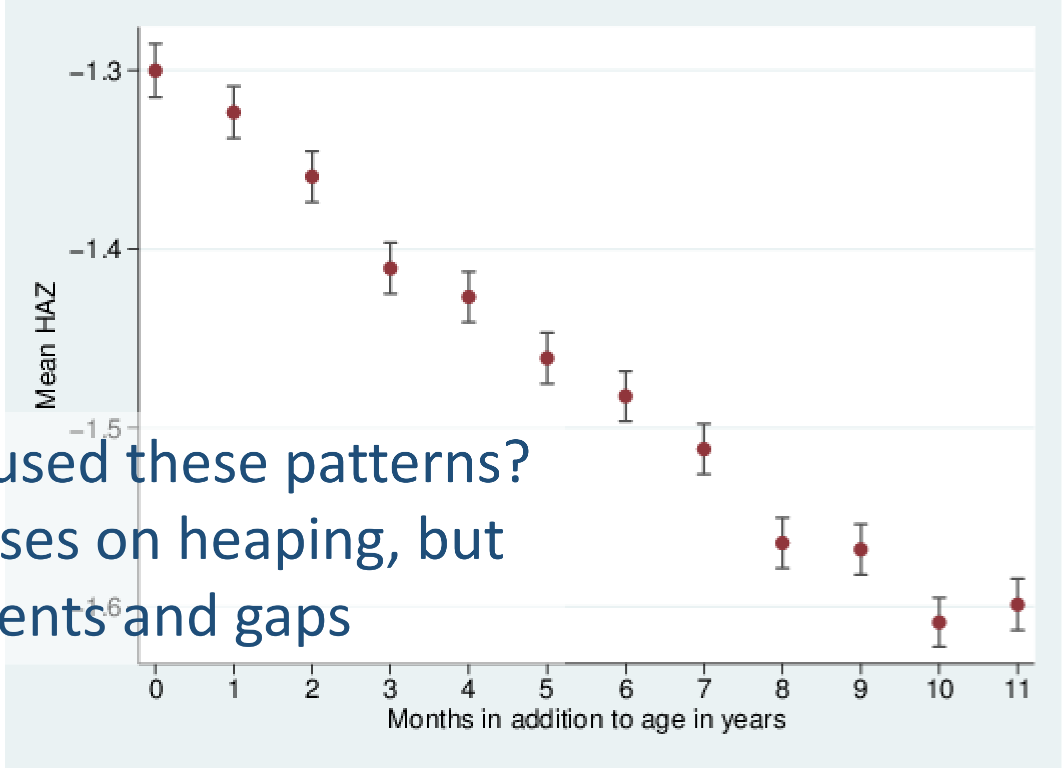
- Exposure to agro-environmental and other shocks in utero and infancy has lifelong consequences for health, human capital and productivity
- Health outcomes are often measured by attained height
  - Heights can be measured quickly and non-invasively, during an interview
  - Potential heights vary for individuals but not for (most) populations
  - Population heights are sensitive to shocks, especially if experienced before age 2
  - Population heights in childhood predict many later outcomes
- We find strong patterns of seasonality, with poor outcomes for children born at bad times (e.g. monsoons, droughts, Ramadan, lean seasons etc.)

The puzzle

We stumbled on this:



Also, this:



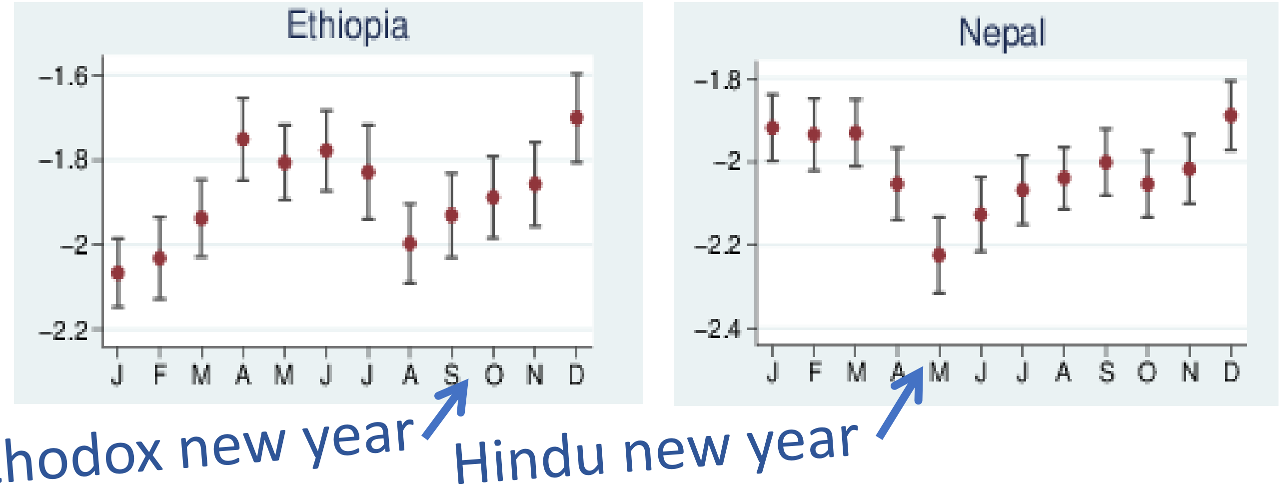
What could have caused these patterns?
Previous work focuses on heaping, but
we find gradients and gaps

Source: DHS data for 990,231 children from 62 countries, various years.
Note: Data shown are mean height-for-age z-scores (HAZ) by month of birth (MOB).
Vertical bars indicate standard errors of the mean HAZ.

The puzzle

The January-December gradient (and Dec-Jan gap) arises in diverse regions:

...and the pattern differs where calendars differ:

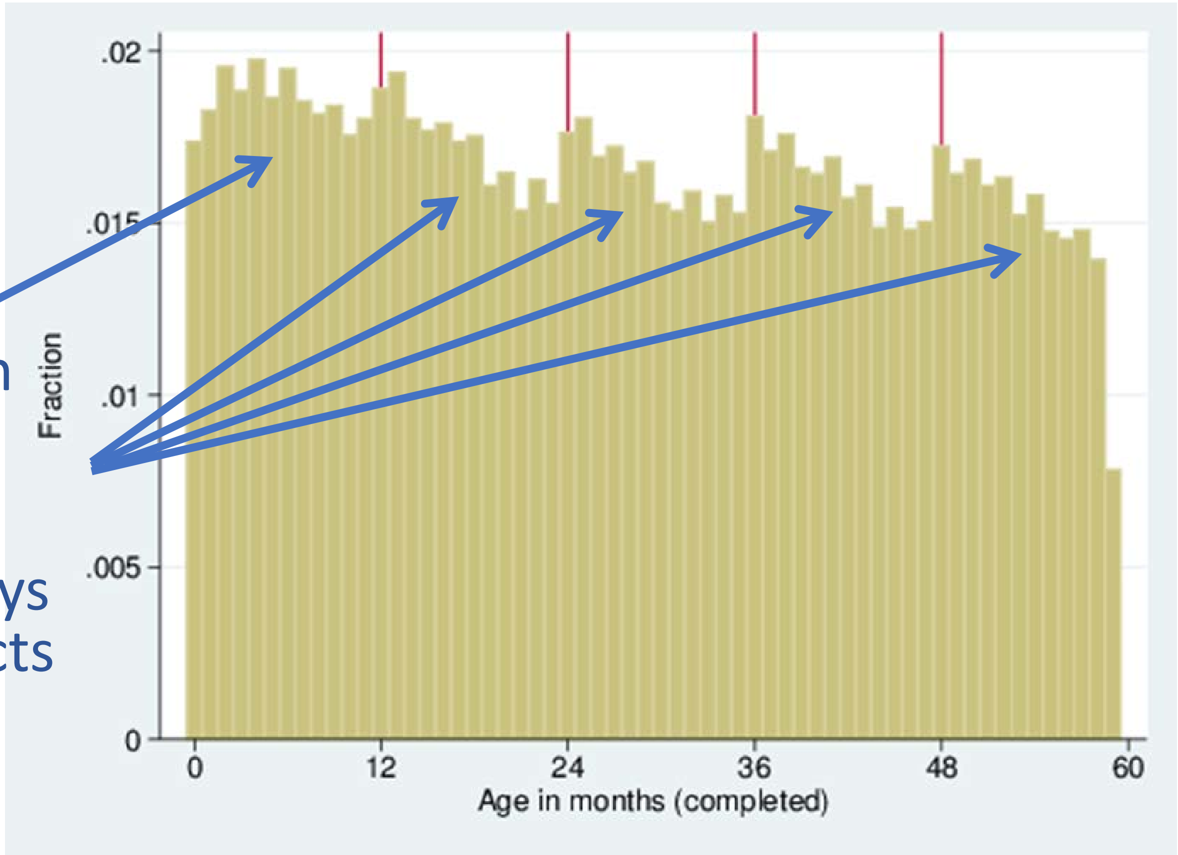


Data shown are mean height-for-age z-scores (HAZ) by month of birth (MOB). Vertical bars indicate standard errors of the mean HAZ.

Children's age is subject to systematic errors

Roughly same number of births each calendar month among infants

After first year, reported birthdays have clear artifacts



Source: DHS data for 990,231 children from 62 countries, various years.

Note: Data shown are fraction of children at each age, with birth date in months prior to the survey date.

Can we explain the anomaly, and adjust for it?

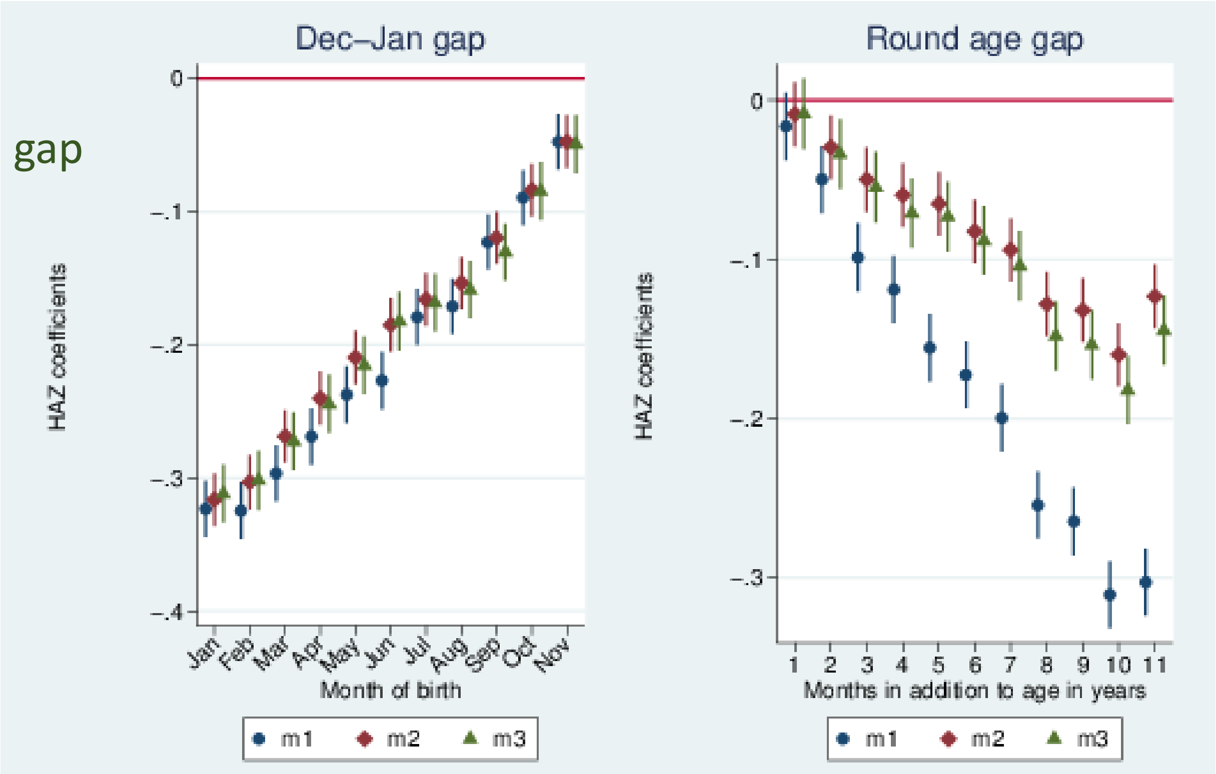
- We can use child growth as a biological clock, to detect systematic errors
  - MOB within calendar years from Jan to Dec, hence a Dec-Jan gap in heights
  - Months past completed years from 1 to 11, hence an end-start gap in heights
- To explain the anomaly, we use a novel model of MOB error
  - we simulate DHS data to replicate the anomaly as an artifact of these errors
  - we use actual DHS data to test where and when the anomaly is largest
  - we use the estimated extent of error to derive corrected stunting rates

Summary of results

- Calendar year anomalies can be replicated with random month of birth
  - The Dec-Jan gap is 0.32 HAZ points, over the entire DHS sample (990,231 children)
  - That could be explained by 11% of children having randomly assigned birth months
  - This kind of error expands the tails of HAZ distribution, causing:
    - 0.5 percentage points increase in stunting (HAZ < -2)
    - 0.7 percentage points increase in severe stunting (HAZ -3)
- The completed-year anomaly is harder to replicate and correct
  - The end-start gap is always confounded by actual aging, so cannot be estimated
  - But this kind of error would systematically understate age and overstates HAZ level, offsetting any effect of MOB error on stunting rates
- The Dec-Jan gap can be used to detect errors in age reporting
  - When using existing surveys in studies of seasonality or early life shocks
  - While conducting new surveys to improve data quality

Statistical controls can't solve the problem

None of Dec-Jan gap is explained by covariates

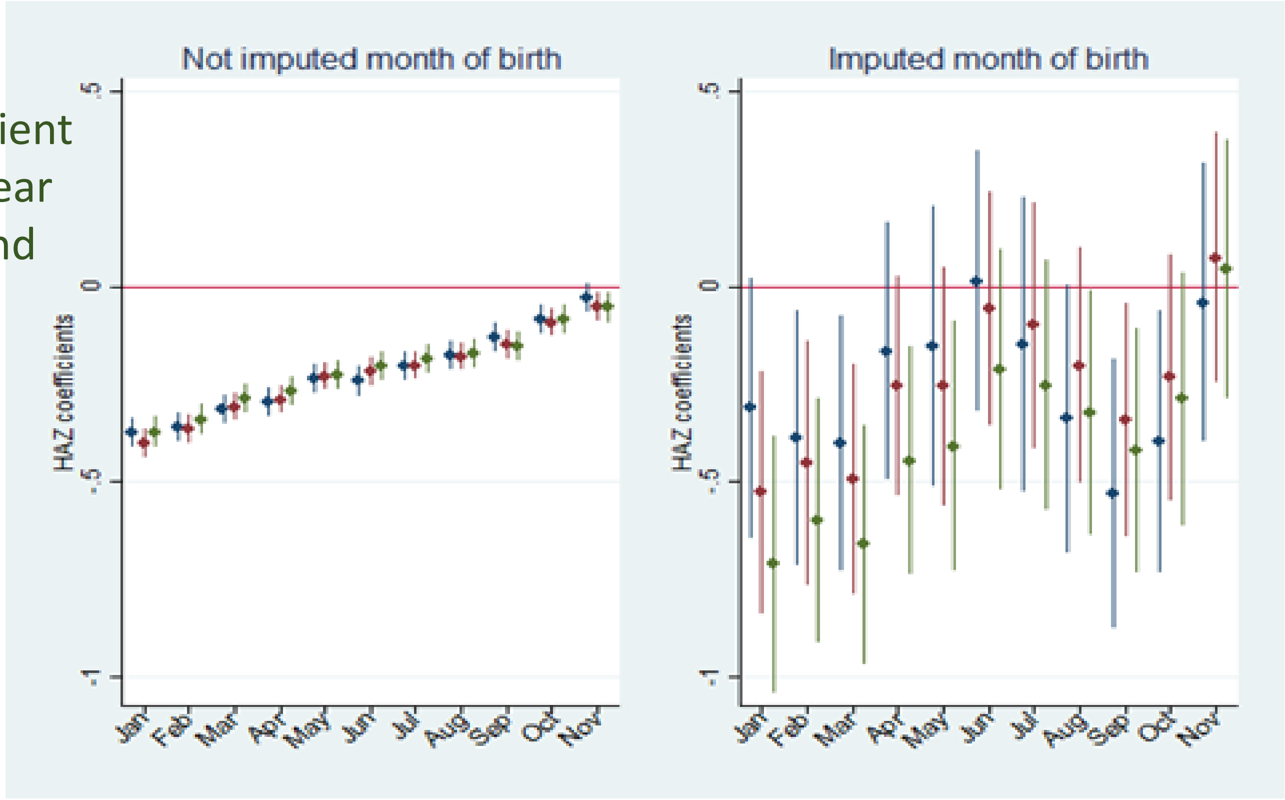


About half of round-age gap is explained by actual child aging

- m1: MOB anomalies without other controls
- m2: adds controls for child sex, age, age-squared and survey fixed effects
- m3: adds household assets, parental education, total number of children, total number of adults, toilet availability, water source and rural location

Actual data have a clearer problem than imputed data

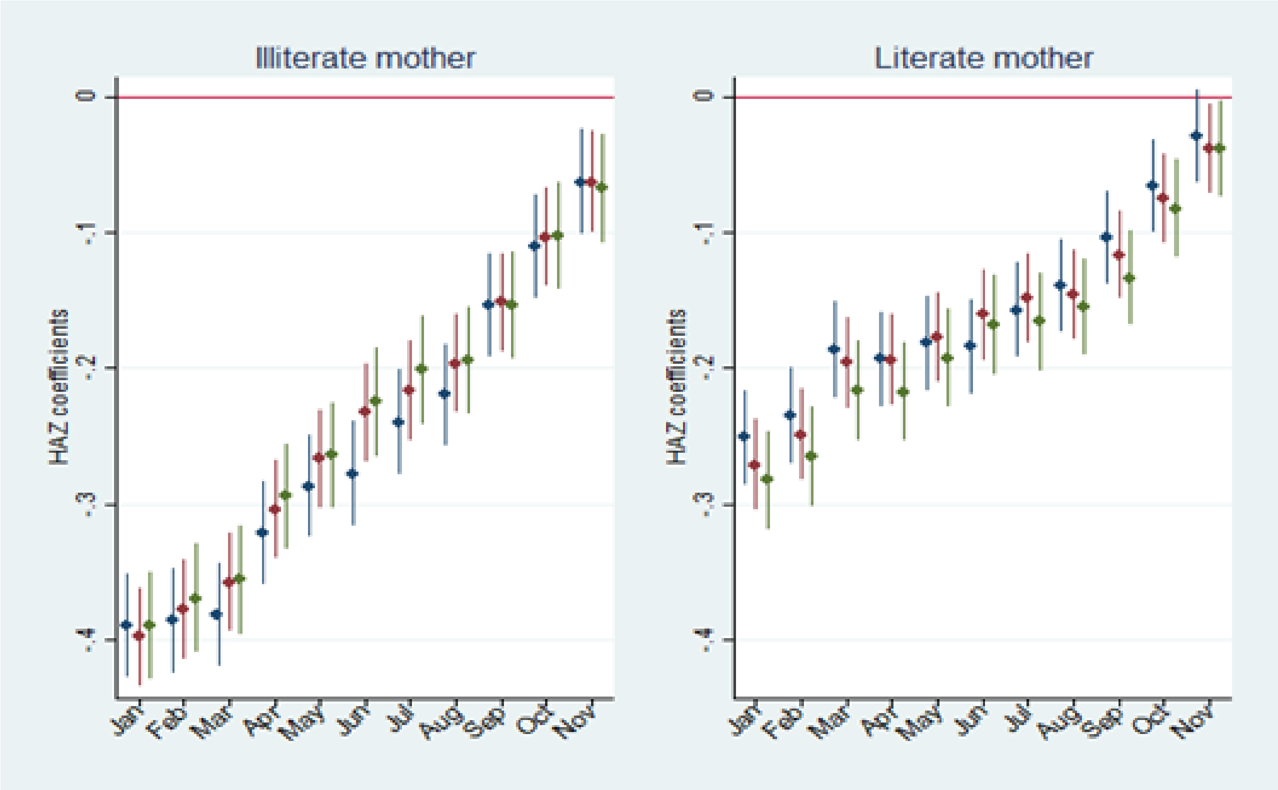
The anomalous gradient from Jan to Dec is clear in the actual data, and not caused by imputation



Imputed data is much more random

- m1: MOB anomalies without other controls
- m2: adds controls for child sex, age, age-squared and survey fixed effects
- m3: adds household assets, parental education, total number of children, total number of adults, toilet availability, water source and rural location

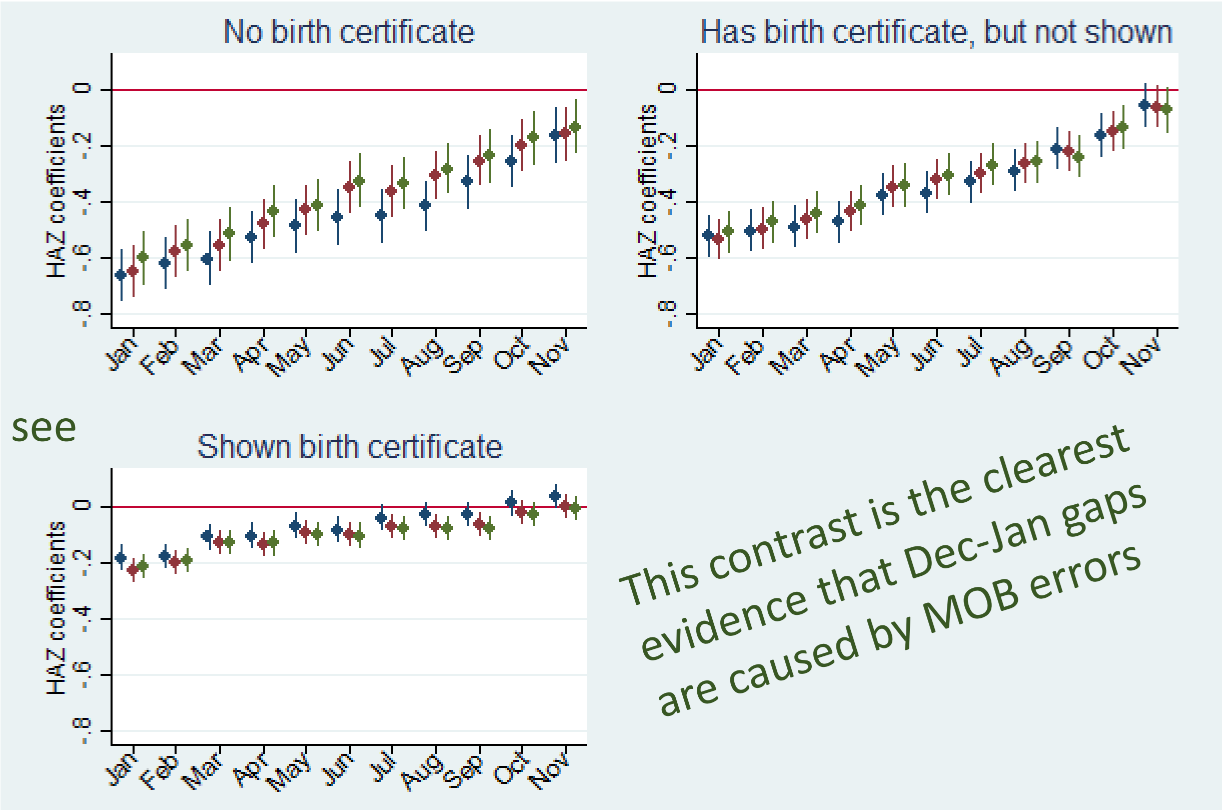
Literacy is associated with a smaller anomaly



Literate mothers have a smaller gradient

- m1: MOB anomalies without other controls
- m2: adds controls for child sex, age, age-squared and survey fixed effects
- m3: adds household assets, parental education, total number of children, total number of adults, toilet availability, water source and rural location

What about birth records?



Having a birth certificate helps…

When enumerators see the certificate, the anomaly almost disappears

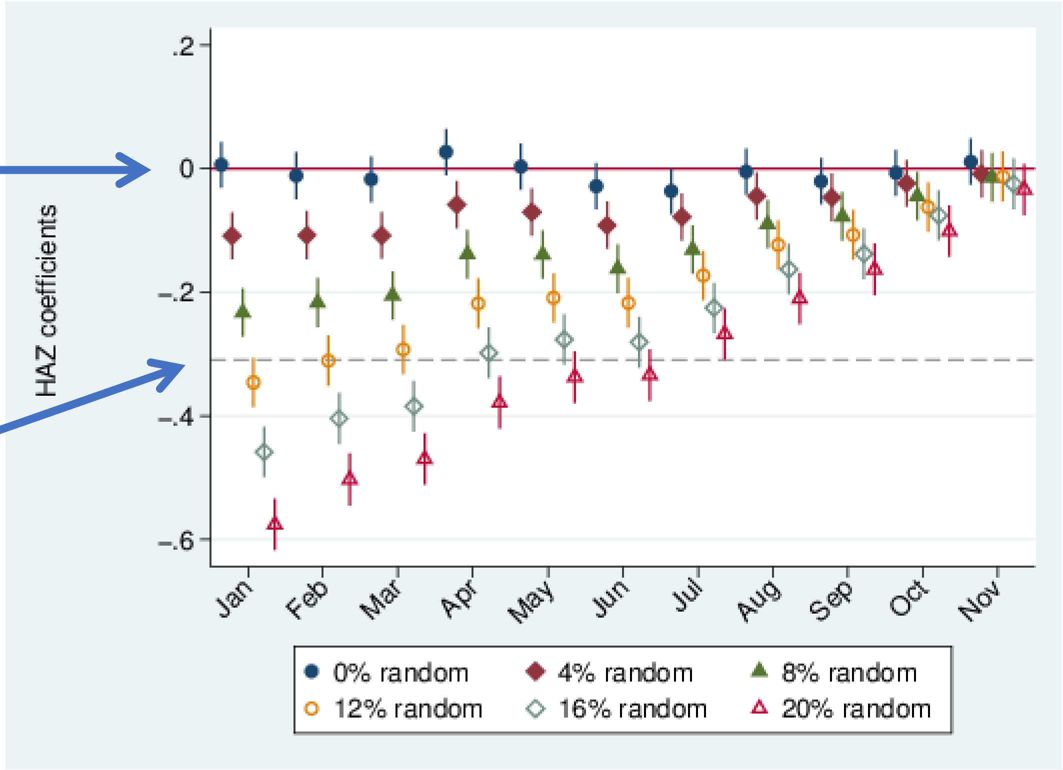
This contrast is the clearest evidence that Dec-Jan gaps are caused by MOB errors

- m1: MOB anomalies without other controls
- m2: adds controls for child sex, age, age-squared and survey fixed effects
- m3: adds household assets, parental education, total number of children, total number of adults, toilet availability, water source and rural location

We can replicate the anomaly by introducing purely random MOBs

With 0% random, there is no MOB effect

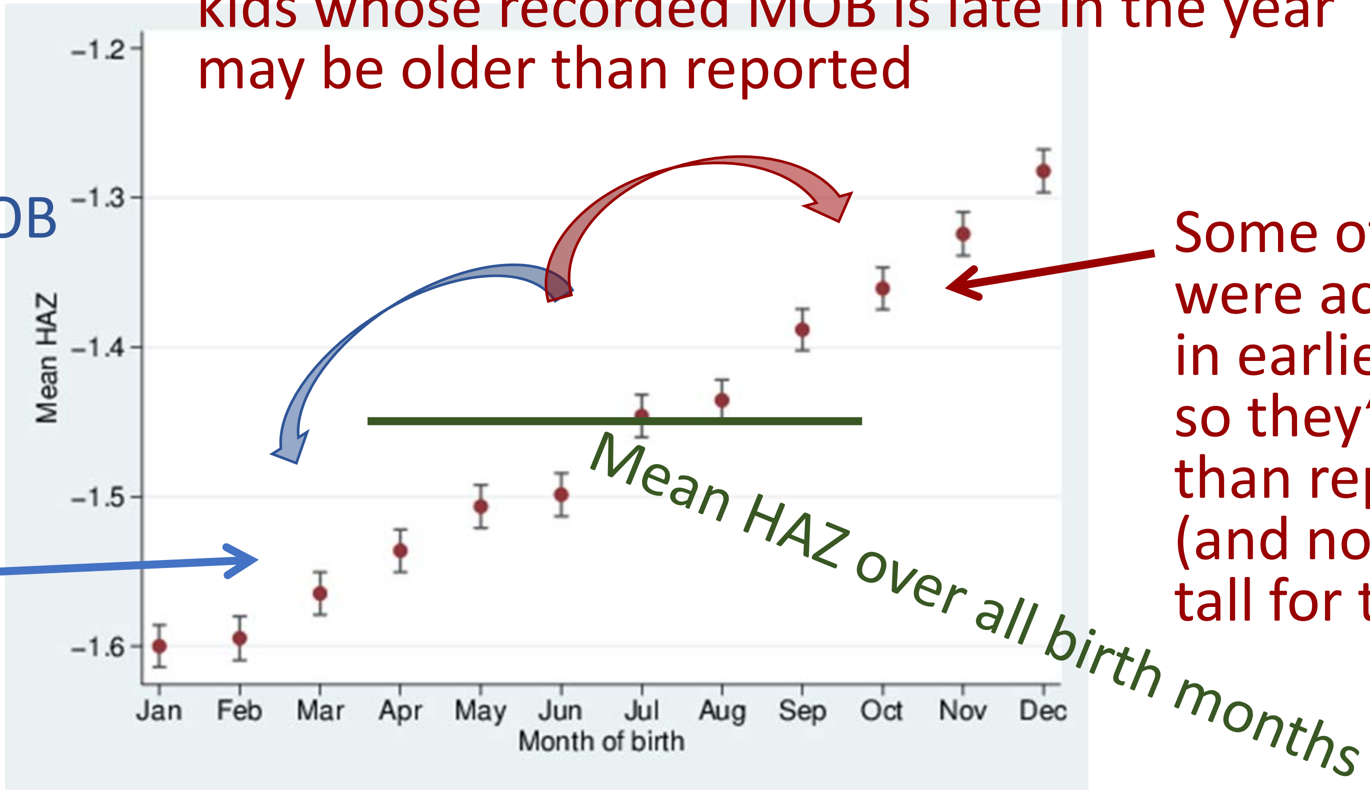
With 11% random, we replicate the actual gap of 0.3 HAZ points



How does *random* MOB leave a nonrandom trace?

When random scrambling occurs within years, kids whose recorded MOB is late in the year may be older than reported

Those with recorded MOB early in the year may be younger than reported



Some of these kids were actually born in earlier months, so they're older than reported (and not actually this tall for their age)

Some of these kids were actually born in later months, so they're younger than reported (and not actually so short for their age)

Data shown are mean height-for-age z-scores (HAZ) by month of birth (MOB), across all DHS surveys (990,231 children from 62 countries, various years). Vertical bars indicate standard errors of the mean HAZ.

In summary…

- Calendar year anomalies can be replicated with random month of birth
  - The Dec-Jan gap is 0.32 HAZ points, over the entire DHS sample (990,231 children)
  - That could be explained by 11% of children having randomly assigned birth months
  - This kind of error expands the tails of HAZ distribution, causing:
    - 0.5 percentage points increase in stunting (HAZ < -2)
    - 0.7 percentage points increase in severe stunting (HAZ -3)
- The completed-year anomaly is harder to replicate and correct
  - The end-start gap is always confounded by actual aging, so cannot be estimated
  - But this kind of error would systematically understate age and overstates HAZ level, offsetting any effect of MOB error on stunting rates
- The Dec-Jan gap can be used to detect errors in age reporting
  - Before using existing surveys in studies of seasonality or early life shocks
  - While conducting new surveys to improve data quality

Thank you!

Contact details:

- Anna Folke Larsen, Univ. of Copenhagen (afl@econ.ku.dk)
- Derek Headey, IFPRI (d.headey@cgiar.org)
- Will Masters, Tufts (william.masters@tufts.edu)

Funding:

- IFPRI: Bill & Melinda Gates Foundation, for Advancing Research on Nutrition and Agriculture (ARENA) at IFPRI
- Tufts: Feed the Future Innovation Lab for Nutrition (USAID grant AID-OAA-L-1-00005) and the Feed the Future Policy Impact Study Consortium (USDA cooperative agreement TA-CA-15-008).
- Copenhagen: Danish Council for Independent Research