The Geographical and Biophysical Correlates of Hunger and Infant Mortality: Lessons from CIESIN's Poverty Mapping Activities

Alex de Sherbinin

AGU Annual Meeting

San Francisco, CA 14 December 2004

Earth Science, Human Well-Being, and the Alleviation of Global Poverty

MillenniumProject
Commissioned by the UN Secretary General and supported by the UN Development Group

CIESIN
Columbia University

The Earth Institute
AT COLUMBIA UNIVERSITY

Coauthors
Bob Chen
Deborah Balk
Marc Levy
Adam Storeygard

Acknowledgments: Rafael Flor, Joanne Gaskell, Fern Greenwell, Jukay Hsu, Malanding Jaiteh, Peter Matlon, Melissa Neuman, Tom Pullum, Jeff Sachs, Pedro Sanchez, Manohar Sharma

Key messages

- There are significant relationships between our poverty proxies (infant mortality and child malnutrition) and biophysical variables such as:
  - Soil quality
  - Drought prevalence
  - Malaria prevalence (which is closely tied to climate)
  - Elevation & slope
  - Water availability
- Geographical factors – proximity to urban areas, distance to ports, roads and railroads – are also significantly related
- The scale of analysis affects our ability to quantify the relationship

CIESIN poverty mapping analyses

Variable	Scope	Unit of Analysis
Infant mortality	Global	Quarter-degree grid cell
Infant and child mortality	10 West African countries	Child
Analysis of hunger hotspots	382 Subnational Units in Africa	Subnational survey region
Underweight status	19 African countries	Child
Underweight status	44 countries in Africa, Asia, and Latin America	Subnational survey region

Motivation

Objectives

- Test proposition that the world's poor live under geographic and biophysical conditions that are significantly different than other populations.
- Characterize the nature of those conditions.
- Experiment with alternative methods for analyzing global spatial poverty data and evaluate their strengths and weaknesses

Context

- Part of multi-pronged approach aimed at understanding global poverty from spatial perspective, within Millennium Project (http://www.unmillenniumproject.org/)
  - ***D***escribe conditions of absolute poverty
  - ***D***iagnose drivers, constraints, opportunities
  - ***D***esign interventions
- Global data cataloging, collection, creation, integration
- Higher-resolution data integration and analysis within hotspots and within select countries
- Partnerships critical
  - World Bank, IFPRI, WRI, Macro, others
- Initial data sets beginning to be released over next few months; others to be released over next 12 months.
  - National Human Development Report Catalog at http://sedac.ciesin.columbia.edu/hdr/

Data and Methods

- Outcome measures
  - Infant Mortality (IMR)
  - Malnutrition
  - [Literacy]
  - [Household Assets]
- Driver and constraint measures
  - Access (roads, rails, ports, distance to coast)
  - Climate (precipitation, growing season, drought, climate zones)
  - Demography (density, urbanization)
  - Topography (elevation, slope)
  - Soil quality
  - Malaria
  - [water availability]
- Convert all data to common grid (.25 degree)
- Produce integrated database with grid cell as unit of analysis
- Bivariate analysis
  - Correlations
  - Pseudo dose-response and exposure estimates
- Multiple regression (OLS)
  - Which variables are significant?
  - What combination generates best fit?
  - How much of the poverty variance is accounted for in equation that uses only geographic variables?
- Spatial regression
  - Control for spatial autocorrelation
  - Experiment with spatial lag and spatial error models
  - Compare results to OLS regression

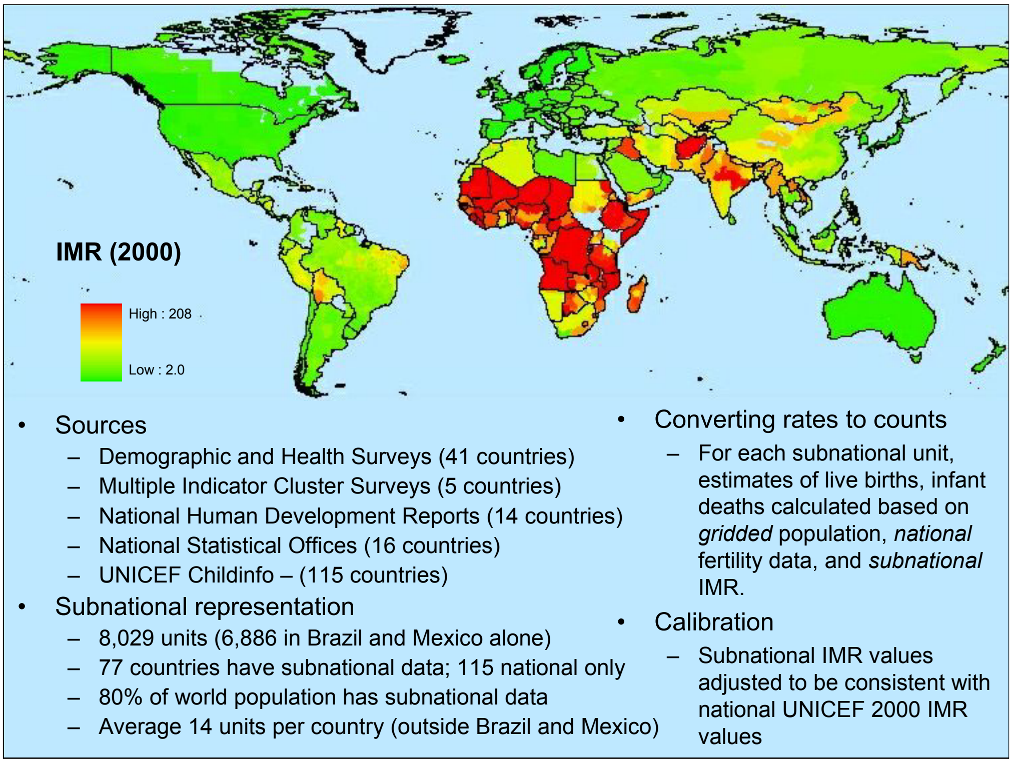
Infant Mortality and child malnutrition were the outcome variables here. We had hopes to include literacy and assets, to give a broader take on poverty, but the data integration issues across countries have been insurmountable so far.

The drivers and constraints we looked at include access to infrastructure like roads, rails and ports; climate data; demographic data, both simple population density as well as access to urban areas; topography; soil constraints; and the stability of malaria transmission.

We have recently been given access to a rich water availability dataset that we hope to incorporate a new round of the analysis that we are beginning now.

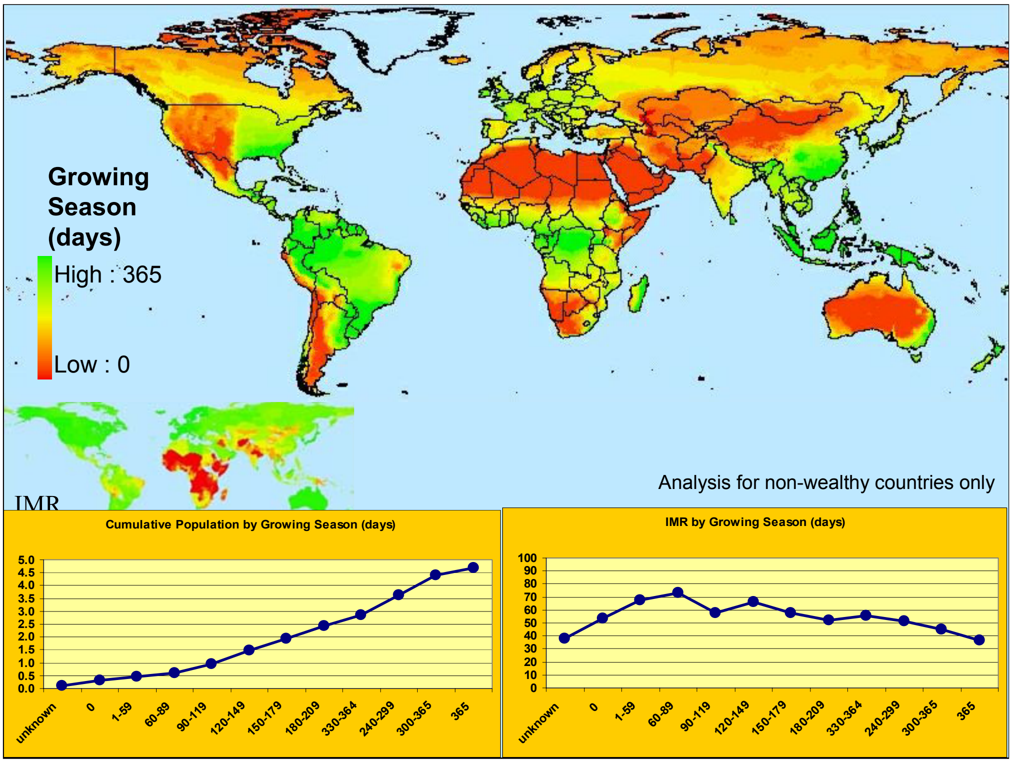
We converted all these data to a common quarter-degree grid and combined them into an integrated database with each grid cell as the unit of analysis.

We looked at correlations on a bivariate basis, before running an OLS regression to determine best fit. We were not able to to control for spatial autocorrelation in a sophisticated manner, but again, that is something we hope to correct in this new round we are starting.



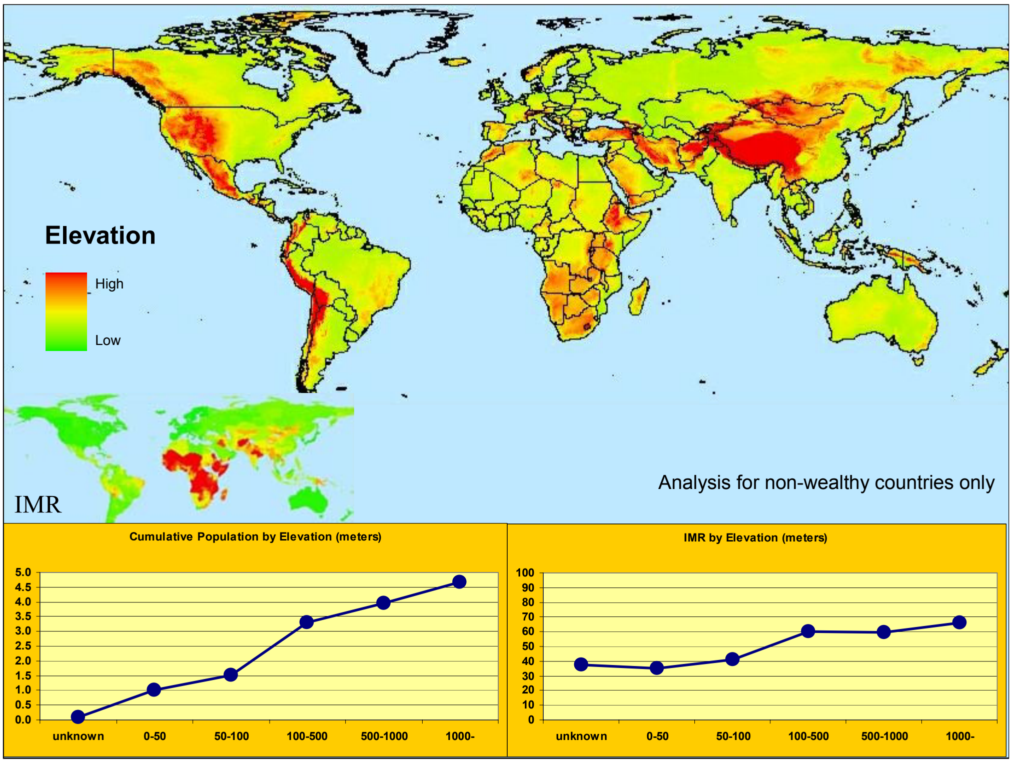
Our main left hand side variable was infant mortality. It had the best data availability, with easily accessible data covering about 99% of the world's population. For 77 countries, we have subnational data, averaging 14 units per country, excluding Brazil and Mexico, for which we have thousands each.

These rates were converted to counts of infant deaths by multiplying by national fertility rates and subnational populations. They were scaled to a common national source and year, namely UNICEF's ChildInfo database for 2000.

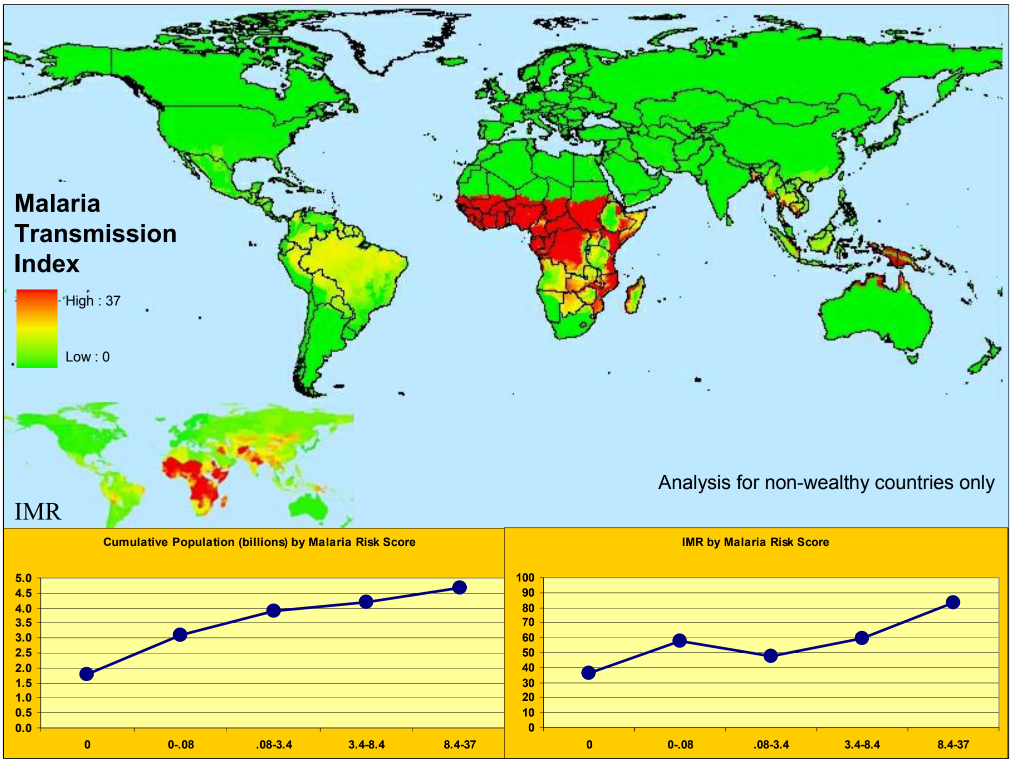


I'd like to go through data on each of the drivers very quickly. Below each on the right is a graph of infant mortality by the mapped parameter.

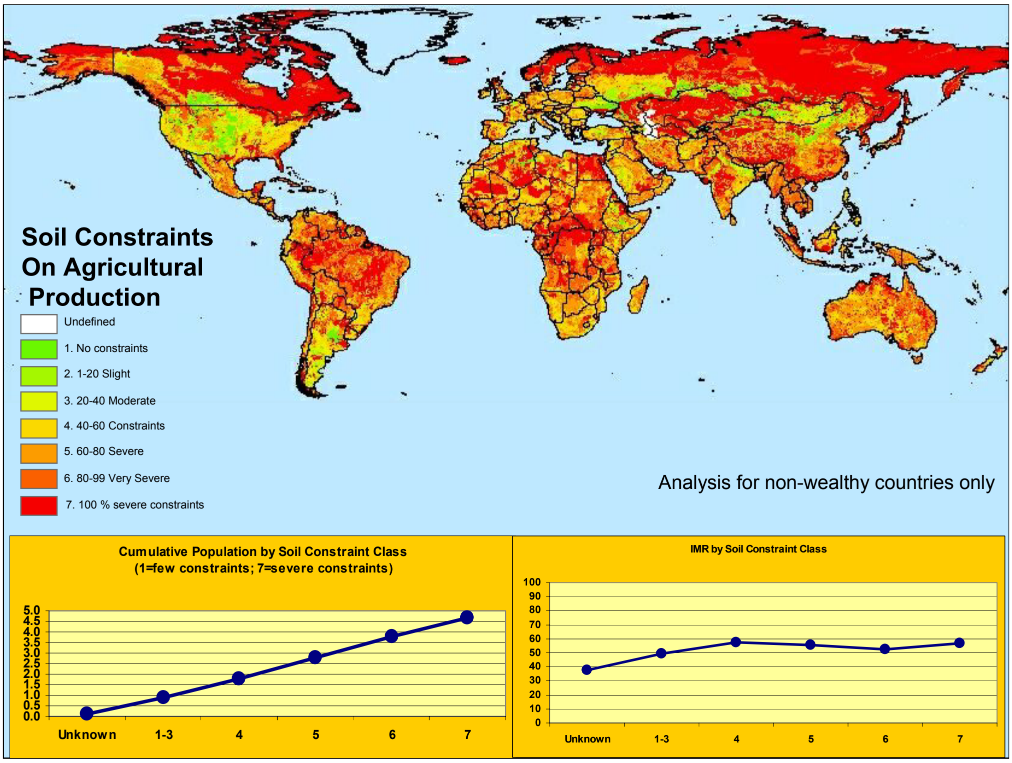
Here is growing season.


Elevation
High
Low
IMR
Analysis for non-wealthy countries only
Cumulative Population by Elevation (meters)
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
unknown
0-50
50-100
100-500
500-1000
1000-
IMR by Elevation (meters)
100
90
80
70
60
50
40
30
20
10
0
unknown
0-50
50-100
100-500
500-1000
1000-

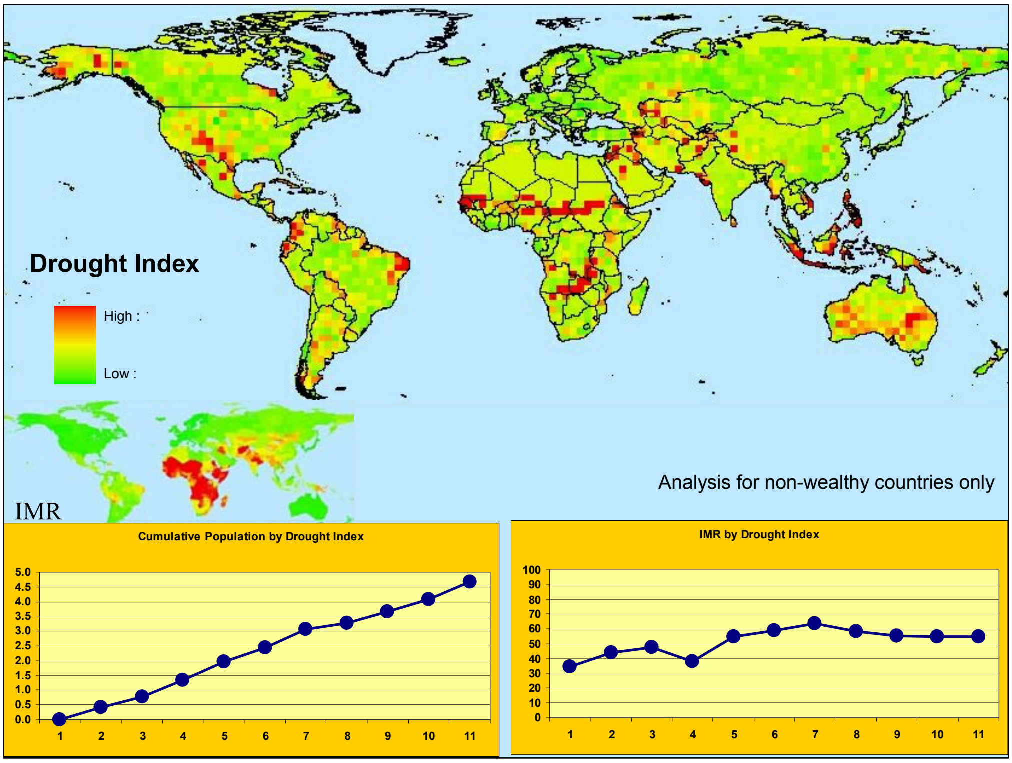
Elevation



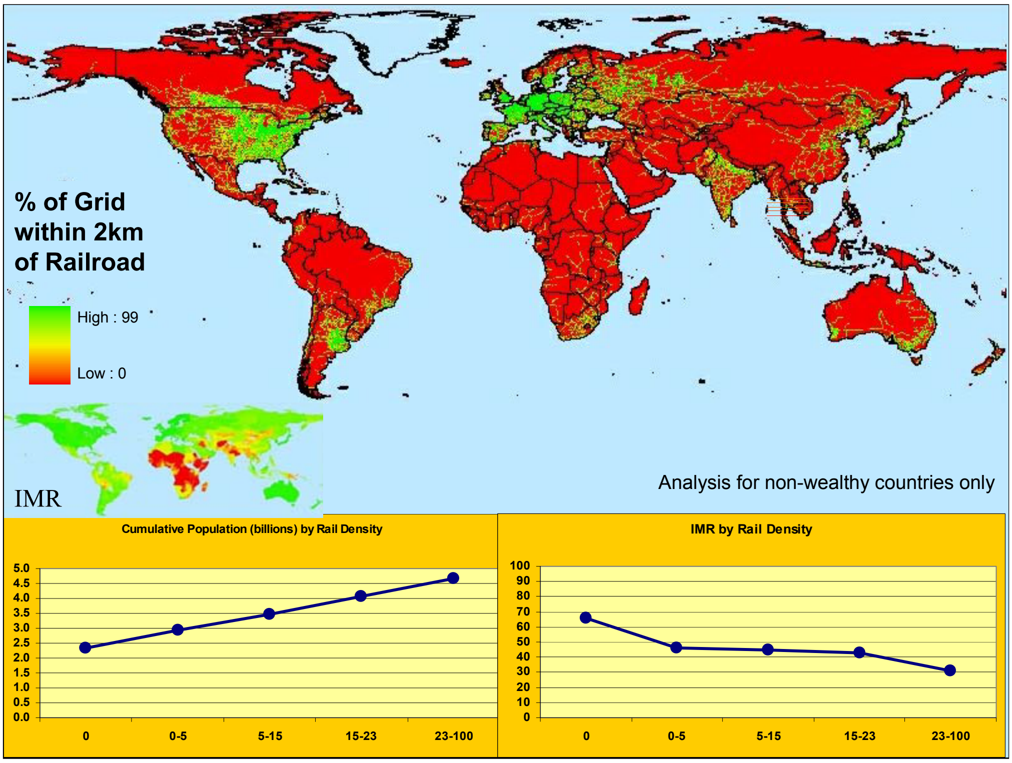
Kiszewski's Malaria Transmission index



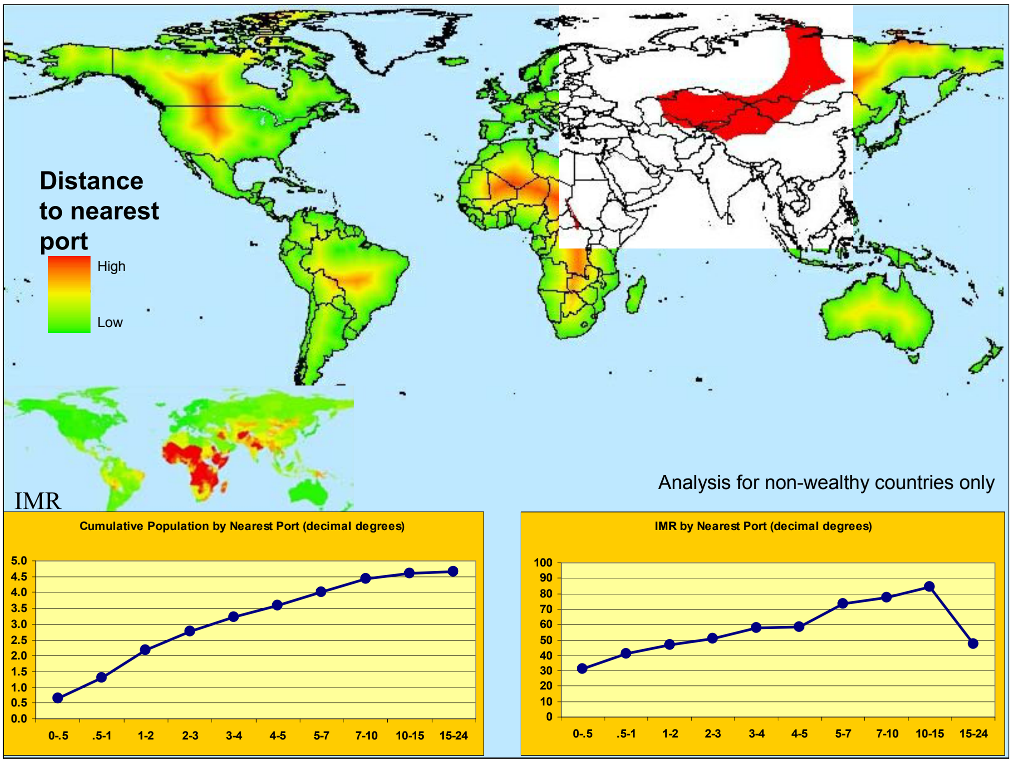
Soil constraints


Drought Index
High :
Low :
Analysis for non-wealthy countries only
IMR
Cumulative Population by Drought Index
5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
1 2 3 4 5 6 7 8 9 10 11
IMR by Drought Index
100 90 80 70 60 50 40 30 20 10 0
1 2 3 4 5 6 7 8 9 10 11

Drought


% of Grid within 2km of Railroad
High : 99
Low : 0
IMR
Analysis for non-wealthy countries only
Cumulative Population (billions) by Rail Density
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
0-5
5-15
15-23
23-100
IMR by Rail Density
100
90
80
70
60
50
40
30
20
10
0
0
0-5
5-15
15-23
23-100

Railroads



And distance to port. This was measured very crudely at the global scale. The largest distances are greatly affected by spurious latitude effects.

Multiple Regression Results (OLS)

- Tested all variables
- Reduced some variable clusters with high correlation
  - 6 different drought measures reduced via factor analysis
  - 4 port measures reduced by identifying nearest port (dropped distance to coast)
  - Dropped precipitation in favor of growing season
- Logged highly skewed variables; tested logged and unlogged

We tested all these variables, but ultimately we dropped some and combined others into principal components. Several drought measures were combined into one. We dropped precipitation due to high correlation with growing season, and distance to coast, reducing the four measures of distance to varying-sized ports into one instead.

	Coefficients				
	B	Std. Error	t	Sig.	
(Constant)	2.423	2.063	1.174	0.240	
Rail	-6.988	0.205	-34.009	0.000	**Access**
Road 2km	-12.049	0.604	-19.946	0.000	
Road 15km	-1.645	0.279	-5.890	0.000	
Nearest Port	0.578	0.016	37.182	0.000	
Elevation	14.914	0.224	66.536	0.000	**Topography**
Slope	-1.844	0.038	-48.366	0.000	
Malaria	2.307	0.014	164.159	0.000	**Disease vectors**
Soil Constraints	0.882	0.050	17.758	0.000	**Agriculture**
Growing Season	-1.170	0.022	-53.883	0.000	
Drought	3.219	0.168	19.145	0.000	
Density	1.262	0.044	28.445	0.000	**Settlement**
Small City	0.872	0.078	11.117	0.000	
Medium City	0.464	0.088	5.281	0.000	
Large City	-0.456	0.127	-3.582	0.000	
National Population-Weighted Density	-0.007	0.000	-101.864	0.000	**National Geographic Summaries**
National Area	-12.561	0.147	-85.736	0.000	
Dependent Variable: IMR	**R^2 = .477 n=156,750 F=8930**				

Here are the results.

Infrastructure variables behaved as expected, with the exception of ports, which behaved in an unstable fashion anyway.

Increased elevation corresponds to increased mortality, but increased slope does not.

Simply increased density, and small and medium sized cities do not show lower mortality (actually higher), but large cities do.

All told, these variables generate an r-squared of .47, and most of the remaining models in the global work are comparable in that respect.

		Coefficients				
		B	Std. Error	t	Sig.	
(Constant)		2.423	2.063	1.174	0.240	
Rail	**					Access
Road 2km						
Road 15km						
Nearest Port	*					
Elevation	*					Topography
Slope	**					
Malaria	*					Disease vectors
Soil Constraints	*-					Agriculture
Growing Season	*					
Drought	*					
Density						Settlement
Small City						
Medium City						
Large City		-0.456	0.127	-3.582	0.000	
National Population Weighted Density	*	-0.007	0.000	-101.864	0.000	National Geographic Summaries
National Area	*	-12.561	0.147	-85.736	0.000	
Dependent Variable: IMR		**R^2 = .477 n=156,750 F=8930**				

For the starred variables, square of the variable is also significant. Adding squared term changes R^2 to .536 and F to 7237.

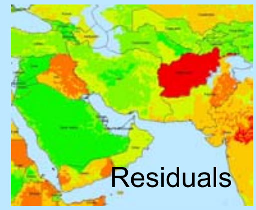
* = variable and its square take opposite signs

** = variable and its square take same sign

*- = original variable no longer significant when square entered

Controlling for Spatial Autocorrelation

- OLS regression assumes errors are distributed randomly through space
- We know this is not an accurate assumption



- Techniques exist to treat spatial autocorrelation explicitly
  - These impose high computing requirements for large grids such as these.
  - Results still pending
- In a separate analysis we report results from a second-best proxy which assumes that errors are correlated within countries (Stata’s robust-error / cluster regression).

Conclusions

- It is possible to characterize the geographic and biophysical conditions under which the world's poor live
  - Poor transportation access
  - Poor agricultural conditions
  - High exposure to malaria
  - Inefficient national geographies (territories smaller, people farther apart)

Read, then:

Caveats: this is not a predictive model; we don't know the causal impact of these factors on IMR; malaria has a very strong correlation with SSA – may be functioning as a SSA dummy.

CIESIN poverty mapping analyses

Variable	**Scope**	**Unit of Analysis**
Underweight status and Infant mortality	Global	Quarter-degree grid cell
Infant and child mortality	10 West African countries	Child
Analysis of hunger hotspots	382 Subnational Units in Africa	Subnational survey region
Underweight status	19 African countries	Child
Underweight status	44 countries in Africa, Asia, and Latin America	Subnational survey region

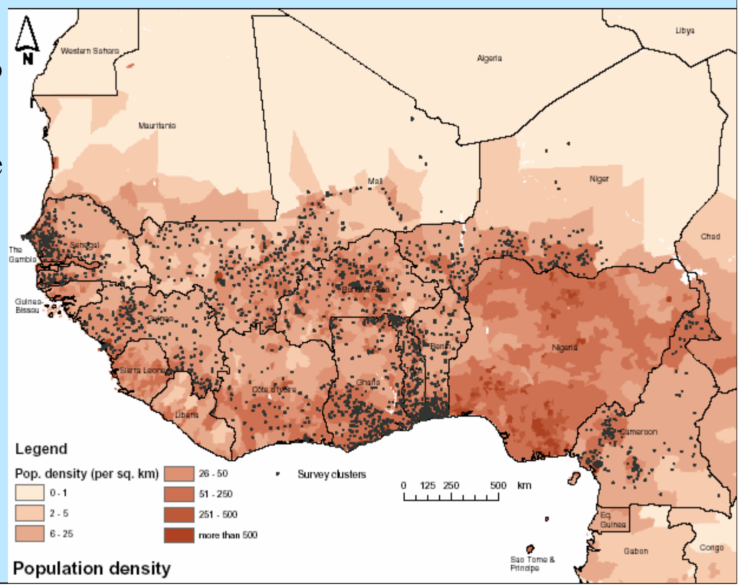
Study Design (1): Unit of Analysis

- Unit of analysis: Individual births (122,389)
  - Occurring in the past 10 years
  - Outcome variables: 1q0 (infant) and 4q1 (child) death rates
    - About half the observations are right-censored (i.e., they had not yet reached aged five so their exposure is incomplete).
- Primary data: Demographic Health Surveys (DHS)
  - Undertaken between 1997-2001
  - In countries where the DHS also had recorded the latitude-longitude location of the sampling cluster
    - GPS devices used to collect cluster location at time of survey

We looked at 120 thousand births from Demographic and Health surveys in ten West African countries. In each of these countries, GPS coordinates were recorded for sample clusters, groups of typically 20 to 30 sampled households in a census enumeration area.

Study Design (2): Areal extent

- **10 West African countries**
  - Benin
  - Burkina Faso
  - Cameroon
  - Cote d’Ivoire
  - Ghana
  - Guinea
  - Mali
  - Niger
  - Senegal
  - Togo



Here are the over 3000 cluster locations from the ten countries.

Study Design (3): Data integration

- Characteristics of
  - Births
  - Mothers
  - Household wealth
- Geographic and environmental variables
  - Climate (aridity, rainfall)
  - Agriculture (farming system, length of growing season)
  - Vector suitability (Malaria)
  - Urban (distance to nearest city of varying sizes)
  - Population Density
  - Access to markets (distance to coast)

The GPS-point locations of the sampling clusters allowed for data integration across varying data formats: points, polygons, lines, or grids. This produced a cluster-level values for all spatially-explicit variables.

Having these georeferenced locations allowed us to add spatial data to the standard set of survey predictors of child mortality, like mother's education and household sanitary facilities.

Study Design (4): Clustering

- **In part because of the complex-sampling frame, clustering is an issue:**
  - For example, all households within a cluster share a single value of each spatial or environmental variable.
  - This is one form of spatial autocorrelation, whereby spatial error is introduced.
  - We adopted an imperfect method (stata robust GLM with cluster) which uniquely specified cluster to accommodate complex sampling frames.
- Nevertheless, other types of clusters occur
  - Maternal clustering
    - Not accounted for here

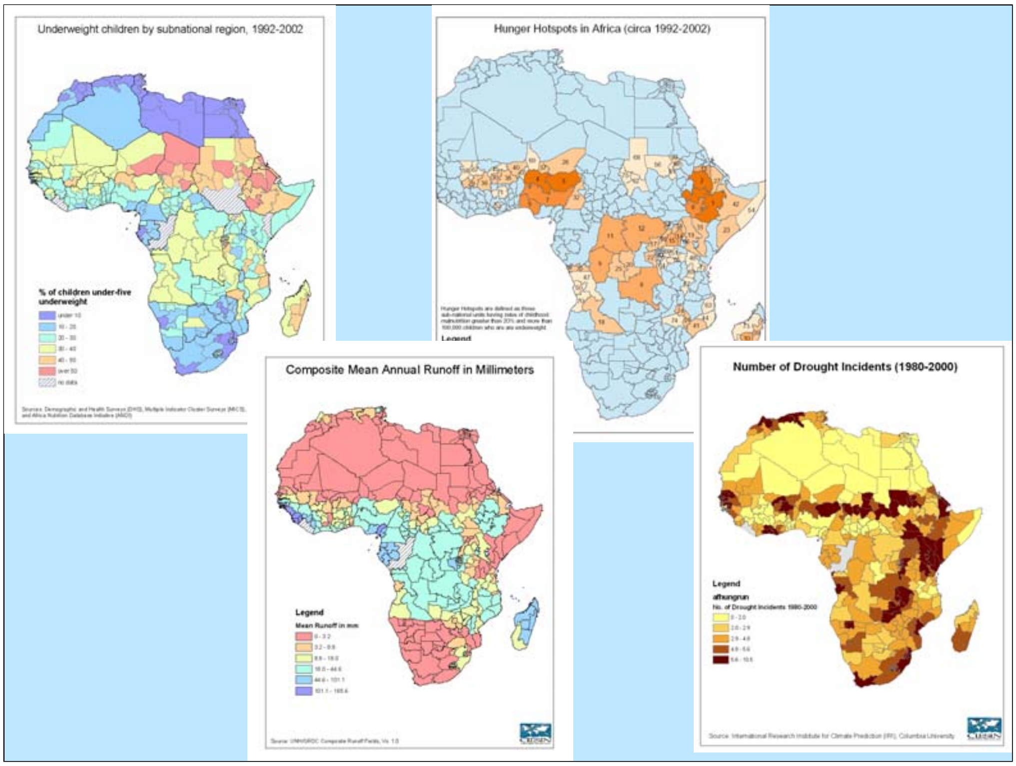
We needed to account for clustering here as well, because all birth in a sample cluster will identical values for all spatial variables; there is no intra-cluster georeferencing.

Findings

- Spatially-explicit variables account for nearly all variation between countries
- High population density, greater urban proximity, and coastal proximity correspond to lower mortality
- Remoteness leads to higher child mortality
- For more information: Balk et al., “A Spatial Analysis of Child Mortality in West Africa”, *Population, Space and Place** 10(3): 175 – 216.

CIESIN poverty mapping analyses

Variable	Scope	Unit of Analysis
Underweight status and Infant mortality	Global	Quarter-degree grid cell
Infant and child mortality	10 West African countries	Child
Analysis of hunger hotspots	382 Subnational Units in Africa	Subnational survey region
Underweight status	19 African countries	Child
Underweight status	44 countries in Africa, Asia, and Latin America	Subnational survey region



The map on the upper left shows the percentage underweight by sub-national survey unit. The map on the upper right shows the hunger hotspots, which were defined as those sub-national units with rates of childhood malnutrition >20% and more than 100,000 children who are underweight. There were 72 sub-national units that met this criteria in Africa.

The map at bottom left simply aggregates the runoff data (from UNH/GRDC's Runoff data set) to the same geospatial units as the underweight (hunger) data so that statistical analyses can be performed. Runoff is the proportion of precipitation that is left after evapotranspiration and the soil moisture deficit are satisfied, and like precipitation it is measured in millimeters.

The map at bottom right represents the average number of drought incidents from 1980-2000 for the same sub-national units. Drought is defined as precipitation less than 75% of the median for 3 months or more, and the values range from 0-14.

Water Runoff and % Underweight

	Unstandardized Coefficients	Sig.
(Constant)	33.781	.000
GDP per Capita (PPP, 1998)	-.003	.000
Average Runoff (in MM)	-.052	.008
North Africa Dummy Variable (N. Africa x GDP p.c.)	-.003	.000
Dependent Variable: Percent of Children Underweight Adjusted R^2 = .433 (P < .000) N = 349		

We hypothesized a negative relationship between runoff levels and malnutrition, and this relationship was confirmed. In Sub-Saharan Africa there appears to be a negative relationship between annual water availability and child malnutrition. Because per capita income also reduces child malnutrition, in North Africa where development levels are higher per capita income strongly mediates the negative effect of low water availability.

Best Predictors of % Underweight

	Unstandardized Coefficients	Sig.
(Constant)	23.063	.000
GDP per Capita (PPP, 1998)	-.002	.000
Average Number of Drought Incidents (1980-2000)	.604	.006
Proportion of the Subnational Unit that is Within 2km of a Road	-18.345	.000
Average Agricultural Constraints	1.747	.002
North Africa Dummy Variable (N. Africa x GDP p.c.)	-.002	.000
Hotspot (1 = yes, 0 = no)	8.59	.000
Dependent Variable: Percent of Children Underweight Adjusted R^2 = .574 (P < .000) N = 349		

The best predictors of underweight status appear to be GDP per capita, the likelihood of drought, accessibility to roads, agricultural constraints, and North African residence.

Hotspots vs. non-Hotspots

Hotspot		Percent of Children Underweight	GDP per Capita (PPP, 1998)	Average Runoff (in MM)	Average Number of Drought Incidents (1980-2000)	Mean Elevation (meters)	Proportion of the Subnational Unit that is Within 2km of a Road	Average Agricultural Constraints
No (N = 260-307)	**Mean**	**21.9**	**$2441**	**17.8**	**3.7**	**641**	**0.30**	**5.7**
	Std. Deviation	12.6	2203	30.3	2.3	588	0.18	0.9
Yes (N = 75)	**Mean**	**36.4**	**$952**	**16.8**	**4.2**	**747**	**0.24**	**5.6**
	Std. Deviation	8.4	509	18.6	1.8	456	0.12	0.7
Diff. In Means	**Significance**	**.000**	**.000**	**.775**	**.095**	**.145**	**.007**	**.336**

- Hunger hotspots have significantly lower GDP per capita and road access, and significantly higher probability of droughts
- No significant difference between hunger hotspots and non-hotspots in runoff, elevation, or agricultural constraints

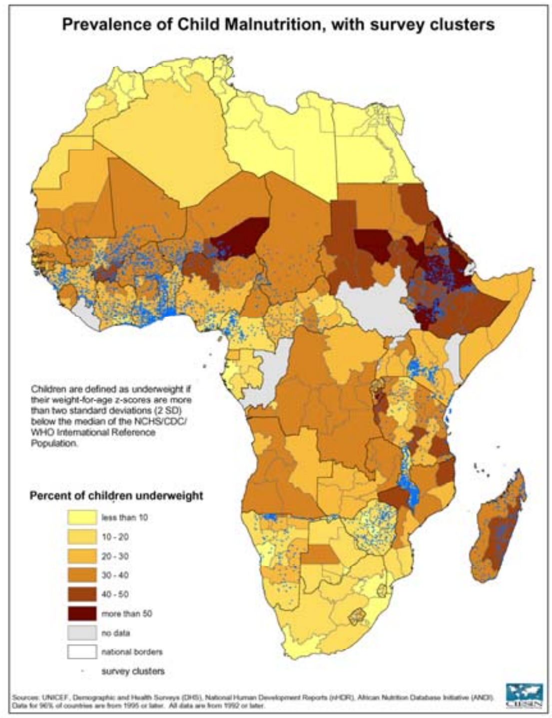
A difference of means analysis suggests that hotspots are indeed different from non-hotspots in a number characteristics such as national income, access to roads, and the probability of droughts. The fact that there was no significant difference between hotspots and non-hotpsots in agricultural constraints suggests to me that there may be problems with that data set.

CIESIN poverty mapping analyses

Variable	Scope	Unit of Analysis
Underweight status and Infant mortality	Global	Quarter-degree grid cell
Infant and child mortality	10 West African countries	Child
Analysis of hunger hotspots	382 Subnational Units in Africa	Subnational survey region
Underweight status	19 African countries	Child
Underweight status	44 countries in Africa, Asia, and Latin America	Subnational survey region

Child-level analysis

19 African countries with data



We did child-level analysis on 19 African countries with georeferenced survey data.

Child-level results

- The following geographic variables were significant, in the expected direction:
  - Length of growing season (too short or too long) (+)
  - Sandy soils (+)
  - Tree cover (-)
  - Distance to port (+)
  - Urban residence (-)
- High fractions of pasture lands unexpectedly corresponded to more hunger – perhaps related to drought prevalence?

These results are for a stable model that consisted of only variables that were significant in the full model. These variables were in turn

Region-level analysis

- 44 countries in Africa, Latin America and Asia
  - Region-level analysis does not require as precise geographic data
  - Regional averages or other summary statistics calculated for individual and household-level variables
  - Principal components created for:
    - simple fertility measures
    - Household assets
    - Household size and composition
    - Density and urban residence

Same regions as in the global work

Region-level results

- In the absence of country dummies, the malaria index, population density and small urban areas were significant
- When country dummies are entered, only the aggregates of household survey data were significant – no geographic variables

Conclusions (1)

- Global scale – we can paint with broad strokes
- Regional scale – a more complex picture emerges
- Finest scales (individual level analysis) – the analysis can be fully integrated with household variables but the strength of the analysis depends on underlying data inputs. Many biophysical data sets are too coarse for local-level analysis.
- Household survey community has moved towards geocoding; now the opportunity is ripe for the biophysical community to meet the needs of the human development/dimensions user needs.

There are problems with each of the levels of analysis. At the meso-scale of sub-national units, it is possible that an average of biophysical or geographic characteristics over an entire sub-national unit does not really best represent the characteristics of the physical environment or the geographic access of households where the survey data were collected. One can easily imagine a case where a large sub-national unit has very low rainfall in the north and more humid climatic conditions in the south. If most of the household sample clusters are in the south then the average of runoff over the entire area may mis-represent the conditions faced by those households. Although sub-national analysis is obviously much better than national-level analysis, the same basic problems apply.

For individual-level analyses using survey clusters, the resolution of the biophysical data can represent an impediment. If biophysical data are on a quarter-degree grid cell, this tells us little about the variability of environmental phenomenon at a highly localized sub-district level, where soil quality and even rainfall can vary dramatically (especially in areas of high topographic relief).

Conclusions (2)

- Multiple levels of analysis inform each other
- Understanding the biophysical constraints that countries face in poverty alleviation (for ag productivity, health constraints) is vital for developing strategies to address poverty
- But physical endowments do not tell the whole story

Finally, it is important to remember that physical or geographic endowments do not tell the whole story – there are obviously many factors affecting infant mortality and hunger, including government policies (and good governance more generally), levels of education, etc. Some of these we took into account through the use of household survey data, but some of them were beyond the scope of the current analyses. Future analyses may introduce other variables related to conflict or governance for a fuller analysis of the many causal mechanisms behind poor child health and nutrition.