

Database Construction for Population-based Social Research

Hongwei Xu

Institute for Social Research

November 17, 2016

Outline

- Introduction
 - Who am I?
 - What is population-based social research?
- Case Study 1
 - “Is Lying Contagious?”
- Case Study 2
 - “Long-term Health Effects of China’s 1959-61 Famine”
- Concluding Remarks

Introduction

- Who am I?
 - PhD in sociology from Brown University (2012)
 - Research assistant professor at the Survey Research Center of ISR (2014-)
 - Faculty affiliate at the Population Studies Center, the Lieberthal-Rogel Center for Chinese Studies (LRCCS), and the China Data Center
 - Social demography, population health and aging, spatial analysis, survey research
 - xuhongw@umich.edu

Introduction

- What is population-based social research?
 - Well-defined population of interest and units of analysis
 - Example: what is China's fertility rate?
 - Families, women, married women, married women 15-49 years old, married women 15-49 years old currently live in China, married women 15-49 years old currently live in China with Chinese citizenship?
 - Document a pattern (i.e., descriptive) or test a theoretical hypothesis
 - Use empirical evidence from data to support your argument
 - Primarily quantitative data
 - Sometimes supplemented by qualitative data
 - Data often from a probabilistic sample that is representative of the population
 - Not sure about using qualitative data alone, at least not yet
 - Correlational or causal analysis

Study 1: Historical Background

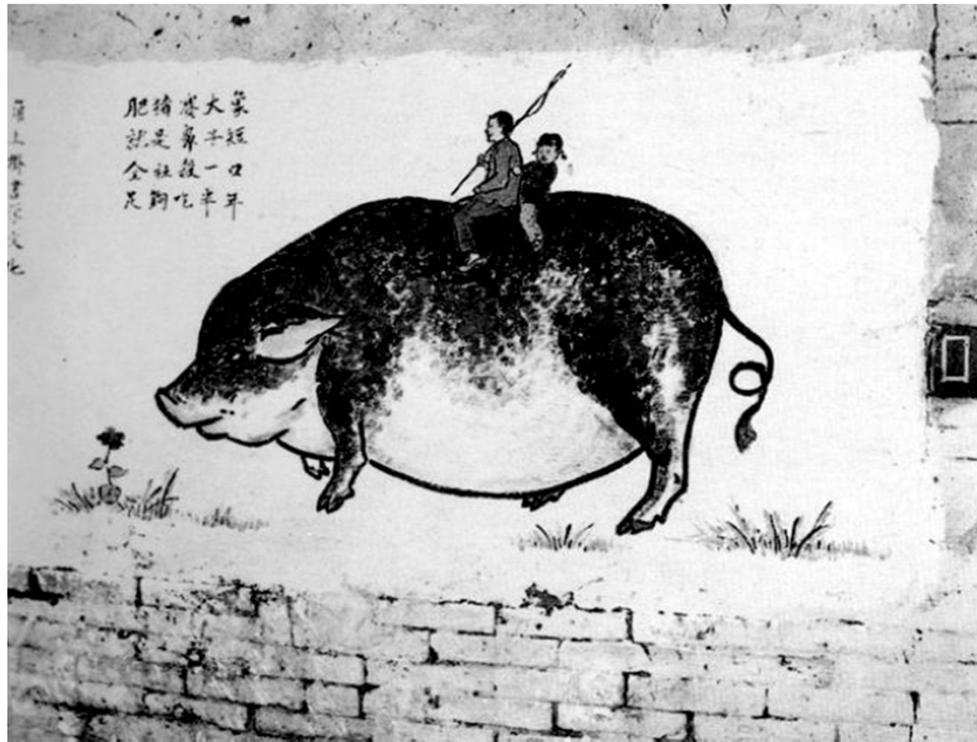
- Pre-1949: Severely traumatized economy and population growth after decades of political upheavals, civil wars, and foreign invasions
- 1953-1957: Gradual recovery during the First Five-Year Plan
- 1958-1962: Ambitious Second Five-Year Plan and aggressive measures

Study 1: Historical Background

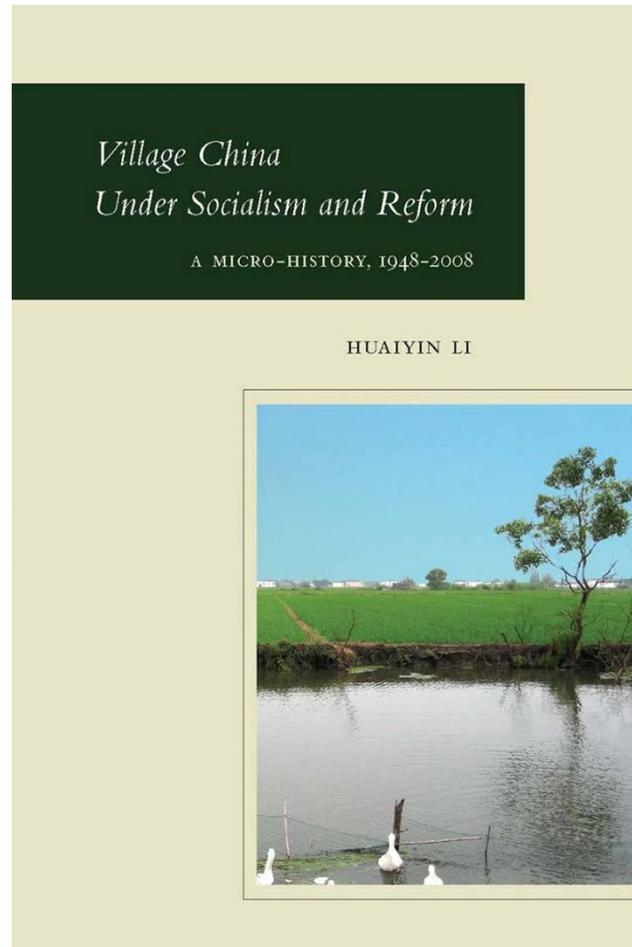
- The *Peoples' Daily* called for a “Great Leap Forward” in agricultural and industrial productions (November 1957)
 - The targeted goal for grain yield raised from 500 (more than doubled from 1957) to 700 MMT
 - Local officials exaggerated their targeted figures by 50% or more, or cut down the planned work period for the original target by half, or both
- The Great Leap Forward (GLF) movement: 1958-1961
 - For a rapid transition from an agricultural society to an industrialized socialist economy
 - Radical social and economic policies

Study 1: Historical Background

- Exaggerated agricultural plans turned into high-yield “satellites” when the 1958 summer harvest season began
 - **Exaggerating grain yields to astronomical levels**
 - Naming inspired by the Soviet Union’s Sputnik satellite launched in 1957
 - First nationally reported by **the *People’s Daily***, the Communist Party of China’s (CPC) official newspaper, on June 8, 1958
 - 2,105 catties of wheat per *mu*, or 6,315 kg per acre in the Chayashan People’s Commune (later known as the “Satellite” Commune), Suiping County of Henan Province.
 - Topped on the next day by another “satellite” in Hubei: 2,357 catties of wheat per *mu*
 - The biggest: 8,586 catties of wheat and 130,435 catties of rice per *mu*



A Case Study from Li's Book (p. 88-89)

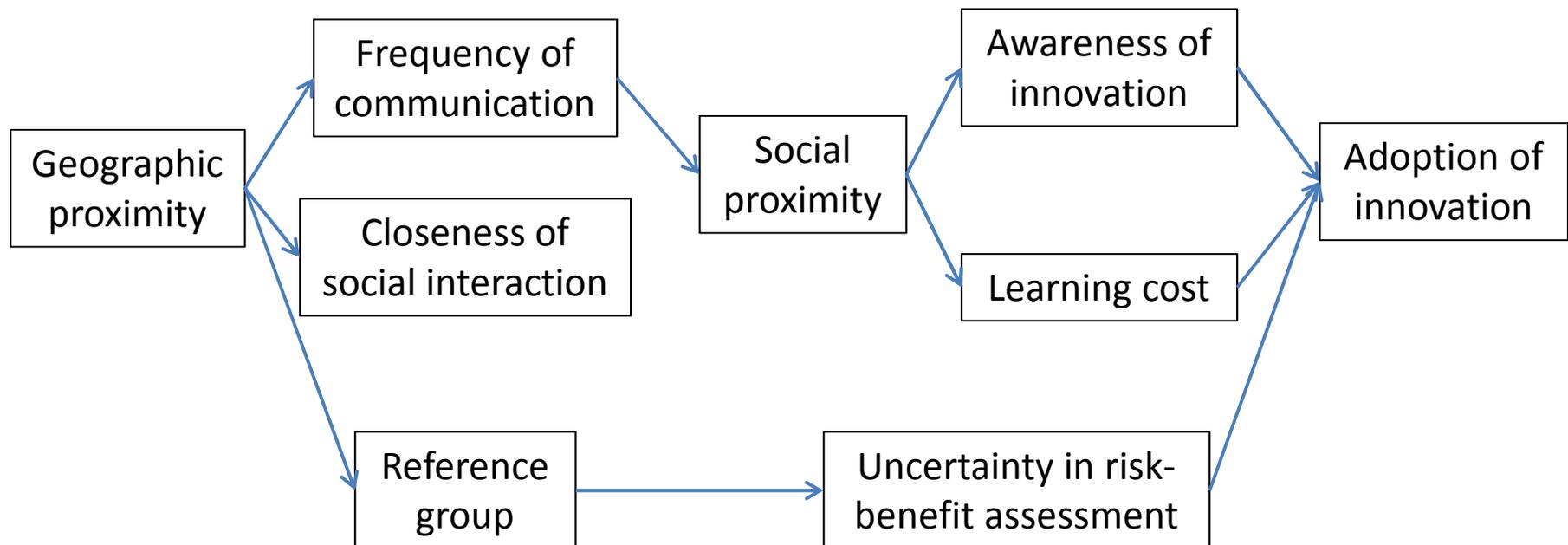


Study 1: Research Questions

- Was there a spatial diffusion process of launching agricultural satellites?
- If yes, what contextual factors contributed to the spatial diffusion?
 - Geographic proximity
 - Political proximity

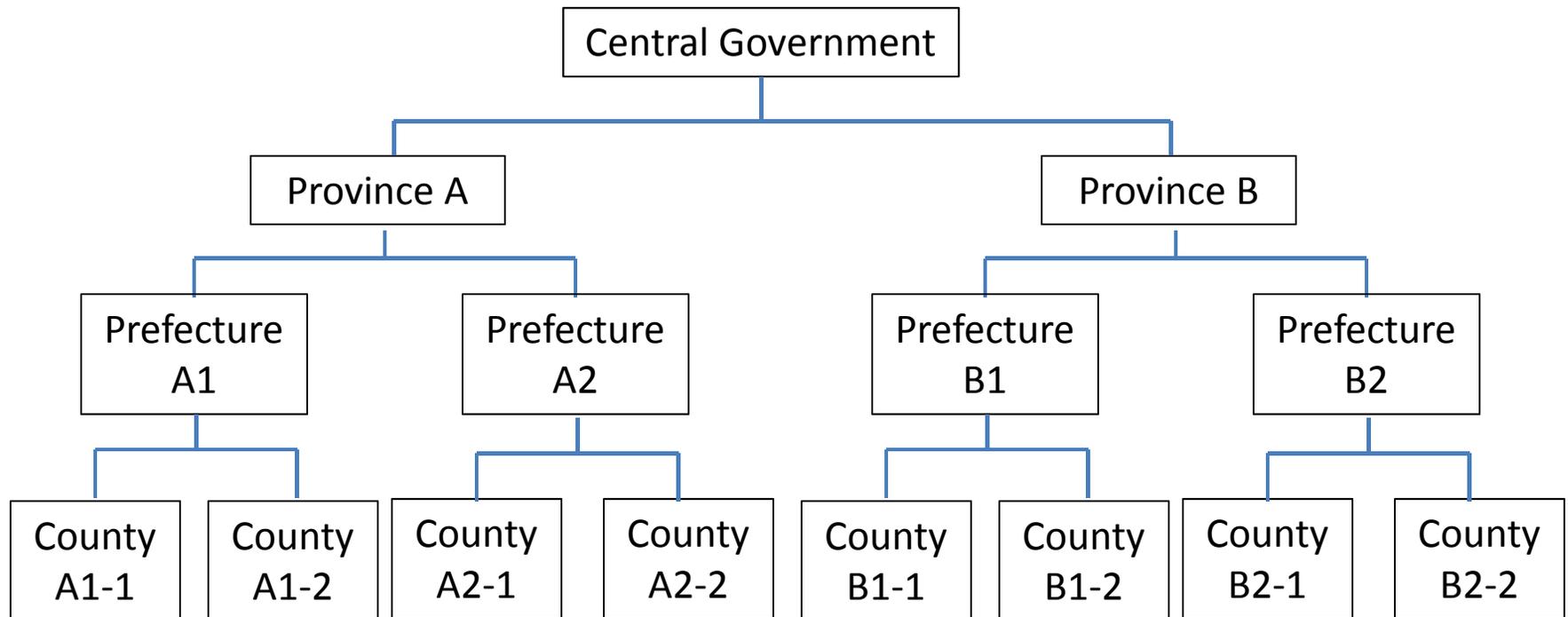
Geographic Proximity

- Launching high-yield satellites as a risky “innovation”



Structural Equivalence in the Form of Political Proximity

- China's multilevel political hierarchy
- Peer pressure and competition to demonstrate political loyalty among structurally equivalent cadres



Source: Modification of Figure 1 in Manion (1985)

Research Hypotheses

- **Hypothesis 1** (main effect of geographic proximity):
 - The probability of a county exaggerating its grain yield would increase if other counties in close geographic proximity had already exaggerated their grain yields.
- **Hypothesis 2** (main effect of political proximity):
 - The probability of a county exaggerating its grain yield would increase if other counties in close political proximity had already exaggerated their grain yields.
- **Hypothesis 3** (interaction effect):
 - The probability of a county exaggerating its grain yield would become higher still if similar exaggerations had been made by counties that were close in both the geographic and political spaces.

Typology of Geographic and Political Proximity

Share geographic boundary	Relative Locations of Two Given Counties		
	In the same province		In different provinces
	In the same prefecture	In different prefectures	
Yes	Type A1 <ul style="list-style-type: none"> • Close geographic proximity • 1st-order political proximity 	Type A2 <ul style="list-style-type: none"> • Close geographic proximity • 2nd-order political proximity 	Type A3 <ul style="list-style-type: none"> • Close geographic proximity
No	Type B1 <ul style="list-style-type: none"> • 1st-order political proximity 	Type B2 <ul style="list-style-type: none"> • 2nd-order political proximity 	Type B3 <ul style="list-style-type: none"> • Not close in geographic or political space

Typology of Geographic and Political Proximity

Hypothesis 1

Share geographic boundary	Relative Locations of Two Given Counties		
	In the same province		In different provinces
	In the same prefecture	In different prefectures	
Yes	Type A1 <ul style="list-style-type: none"> • Close geographic proximity • 1st-order political proximity 	Type A2 <ul style="list-style-type: none"> • Close geographic proximity • 2nd-order political proximity 	Type A3 <ul style="list-style-type: none"> • Close geographic proximity
No	Type B1 <ul style="list-style-type: none"> • 1st-order political proximity 	Type B2 <ul style="list-style-type: none"> • 2nd-order political proximity 	Type B3 <ul style="list-style-type: none"> • Not close in geographic or political space

Hypothesis 3

Hypothesis 2

Data Source: News Paper Archive

The screenshot shows the 'Research Guides' page for 'Chinese Studies: Databases'. It features a navigation menu with 'Databases' selected. The page is divided into three main sections: 'For Research on Pre-Modern China', 'For Research on Modern China', and 'Search Engines'. Each section contains a list of database resources with brief descriptions and search guides.

M LIBRARY | Research Guides

University of Michigan Library / Research Guides / Chinese Studies / Databases

Chinese Studies: Databases Search

A guide to selected digital resources for conducting research in Chinese Studies at the University of Michigan

Home Bibliographies **Databases** E-Journals E-Books Theses Reference Works Web Resources

For Research on Pre-Modern China

- **Scripta Sinica 漢籍電子文獻**
Full-text database of pre-modern Chinese written works. Developed by Academia Sinica in Taipei, Taiwan. [Search guide.](#)
- **CHANT: Chinese Ancient Texts**
Full-text database of early (pre-7th century) Chinese written works produced by the Chinese University of Hong Kong. [Search guide.](#)
Note: The interface has been redesigned. Now the log-in is automatic--no need to click anywhere.
- **Unihan Databases 書同文古籍數據庫**
Includes full-text databases of 中醫中藥古籍, 歷代書畫文集, 四部叢刊, 四部備要, 十通, 清代史料文獻, 明代史料文獻 and 歷代石刻史料彙編.
- **Siku quanshu 四庫全書**
Digital version of the *Wenyuan Ge* 文淵閣 copy of the *Siku quanshu*. IP log-in for on-campus use only. Off-campus log-in instruction please see [here](#). [Search guide.](#)
Please note:

For Research on Modern China

- **Late Qing Dynasty Periodical Full-text Database (1833-1911) [晚清期刊全文数据库]**
- **Chinese Periodical Full-text Database (1911-1949)**
These two databases are located in the same platform and can be cross-searched. 晚清 database includes 302 periodicals published between 1833 and 1911, and 民国 database includes approximately 25,000 periodicals published between 1912 and 1949. Under 晚清 database, there are two single periodical collections: 遐迩貫珍 and 點石齋畫報, both cross-searched with main databases.
- **The North-China Daily News & Herald Newspapers and Hong Lists (1850~1951) 字林洋行中英文報紙全文數據庫**
Covers the English and Chinese newspapers published by North-China Daily News & Herald, Limited, the largest British-founded press agency in Shanghai. Newspaper titles include The North-China Herald 北華捷報, The North-China Daily News 字林西報, The Chinese Shipping List 中國輪船公司, 上海新報, 北華商報, 北華新報, 北華西報, 北華南報, 北華東報, 北華西報, 北華南報, 北華東報, 北華西報, 北華南報, 北華東報.

Search Engines

- **Duxiu Knowledge Search 读秀中文学术搜索**
- **CNKI (China National Knowledge Infrastructure) 中国知网**
- **Wanfang Data 万方数据知识服务平台**
- **中国国家图书馆文津搜索**

Archival Sources

- **China: Trade, Politics and Culture 1793-1980**
It contains archival sources from the School of Oriental and African Studies (SOAS) and British Library.
- **Foreign Office Files for China, 1919-1980**
Sourced from the National Archives, Kew--the UK government's official archive, this database currently contains five sections:
1919-1929, Kuomintang, CCP and the Third

Data Source: News Paper Archive

- **China Core Newspapers 中国重要报纸全文数据库**
This full-text database is a part of the CNKI (China National Knowledge Infrastructure) e-resources package provided by EastView Information Services. [Search guide](#).
- **Renmin Ribao 人民日报 (People's Daily)**
Full-text database (with PDF images of the original print version) of Chinese Communist Party's most important newspaper. Cover 1946-2008.
- **Guangming Ribao 光明日报**
Full-text database (with PDF images of the original print version) of China's most important official newspaper. Cover 1949-2010.

Data Source: News Paper Archive

- The People's Daily (June 8 – September 30, 1958)
 - Grain (rice and wheat) yield $\geq 1,000$ catties/*mu* (or 3,000 kg/acre)
 - By 1980, the 99th percentile of county-level average grain yield was merely 740 catties per mu
 - More than 600 “satellites” June-Sep 1958 transcribed

Data Source: GIS Map

- Geocode each satellite to the county level according to the 1958 administrative boundary map (China Data Center at Michigan)

The screenshot displays the Mirlyn Catalog library interface. At the top, the header includes the 'Mirlyn Catalog' logo and navigation links for 'Search Tools', 'ArticlesPlus', 'MGet It', 'I.L.L.', and 'Ask a Librarian'. Below the header is a search bar with a 'Clear' button, a search input field, a dropdown menu set to 'All Fields', and a 'Find' button. Underneath the search bar are two checkboxes: 'Limit to Available online' and 'Include Search only (no full text)'. A row of action buttons includes 'Cite this', 'Text this', 'Email', 'Export to reworks', and 'Export to Endnote', followed by a 'For Selected items' link. The main content area shows a search result for 'The administrative boundary maps of China : (1949-2014)'. To the left of the title is a placeholder icon with the text 'NO IMAGE AVAILABLE' and a 'Save to Favorites' button. The title is underlined. Below the title is a table of metadata:

Contributors:	China. Guo jia tong ji ju. Beijing Hua tong ren shi chang xin xi you xian ze ren gong si. University of Michigan. China Data Center.
Format:	• CDROM Map
Language:	English, Chinese
Published:	2013 [Beijing, China] : All China Marketing Research Co., LTD.,
Distributed:	Ann Arbor, MI : University of Michigan China Data Center, March 27, 2013.

Below the metadata is a tabbed interface with 'Holdings' selected. Under the 'Holdings' tab, the following information is displayed:

Hatcher Graduate Asia Library Office - Rm. 418 N

CD-ROM AC 19

Data Source: Government Statistics

Mirlyn Catalog Search Tools • ArticlesPlus • MGet It • I.L.L. • Ask a Librarian

MLibrary

Clear All Fields

Limit to **Available online** ? Include **Search only** (no full text) ?

[Back to Results](#)
Showing record 1 of 1

For Selected items (0):

 **Save to Favorites**

Zhongguo nong cun jing ji tong ji da juan, 1949-1986 / Zhonghua Renn
中国农村经济统计大全, 1949-1986 / 中华人民共和国农业部计划司编.

Contributors:	China. Nong ye bu. Ji hua si. China. 农业部. 计划司.
Format:	 Book
Language:	Chinese
Published	1989 Beijing : Nong ye chu ban she : Xin hua shu dian Beijing fa xing suo fa xing, 北京 : 农业出版社 : 新华书店北京发行所发行,
Edition:	Di 1 ban. 第1版.

Data Source: Government Statistics

Mirlyn Catalog Search Tools • ArticlesPlus • MGet It • I.L.L. • Ask a Librarian

MLibrary

Clear All Fields

Limit to **Available online** Include **Search only** (no full text)

For Selected items (0): List |

Zhongguo fen xian nong cun jing ji tong ji gai yao : 1980-1987 / Guo jia tong ji
中国分县农村经济统计概要：1980-1987 / 国家统计局农村社会经济统计司编.

Contributors:	China. Nong cun she hui jing ji tong ji si. China. 农村社会经济统计司.
Format:	Book
Language:	Chinese
Published	1989 [Peking] : Zhongguo tong ji chu ban she : Xin hua shu dian Beijing fa xing suo fa xing, [Peking] : 中国统计出版社 : 新华书店北京发行所发行,
Edition:	Di 1 ban. 第1版.

Data Source: Government Statistics

- County-level control variables from government statistics
 - Terrain (1980): plain, hilly, pasture, mountain
 - Old revolutionary base area (1949)
 - Ethnic minority area (1980)
 - High per capita income in the rural population (1986)

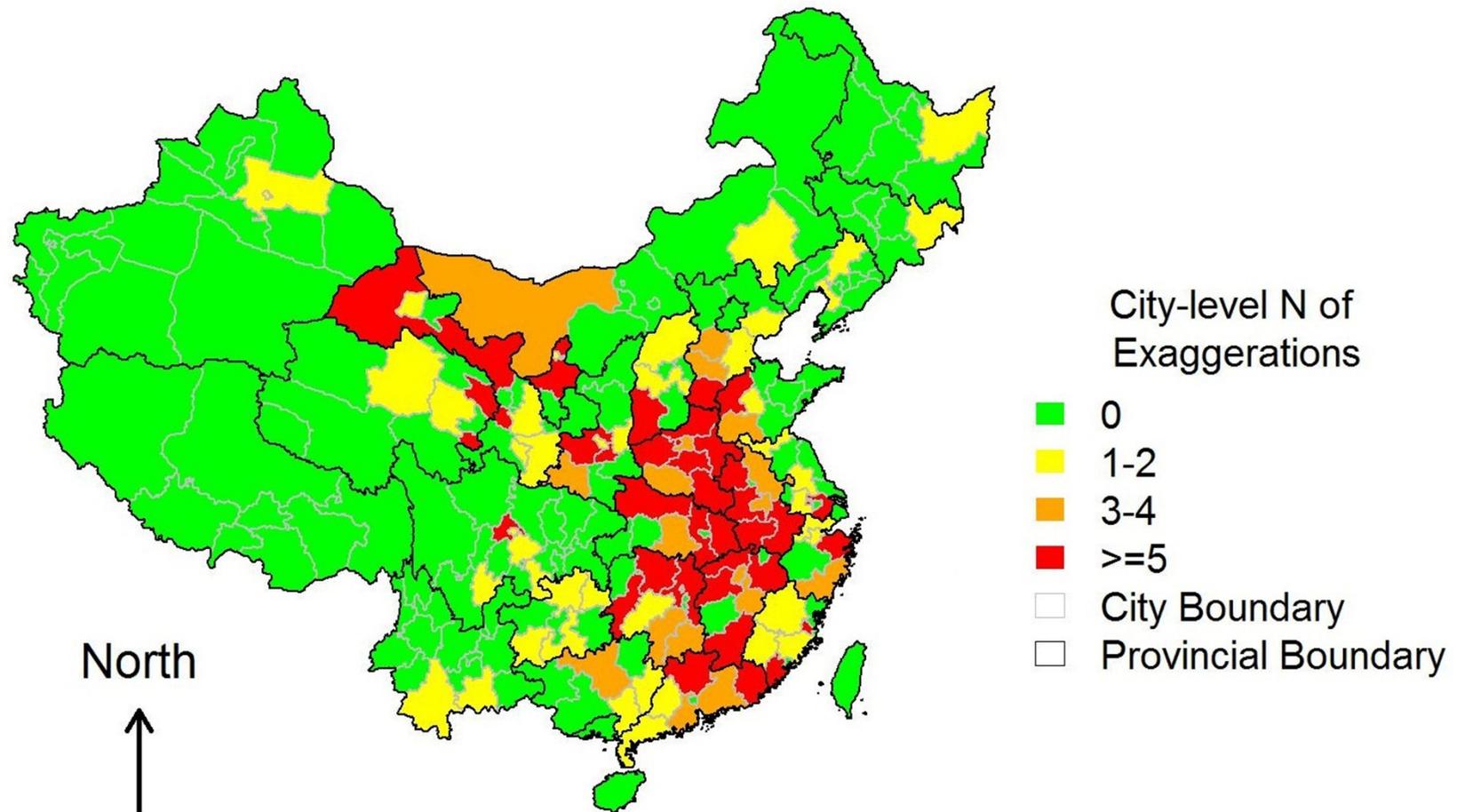
Event History Analysis

- Units of analysis: Counties in the mainland China, 1958
- Event of interest: Launching a high-yield satellite
- Exposure of risk began
 - On June 8, when the notion of high-yield satellites was invented, for event debut
 - Or the next day after an event occurrence for repeated events
- Right censoring if no satellite launched by Sep 30
- Sensitivity analysis
 - Include other food crops than rice/wheat
 - Subsample restriction: only include counties that have at least one neighbor of each type
- Model the logged hazard of event debut
 - $\log h_i(t) = h_0(t) + \sum_{j=A1}^{B2} \beta_j x_{i,j}(t) + \gamma Z_i$
- Model the logged hazard of k th repeated event
 - $\log h_{ik}(t) = h_0(t_k - t_{k-1}) + \sum_{j=A1}^{B2} \beta_j x_{i,j}(t) + \gamma Z_i$

Frequency Distribution of Exaggerations

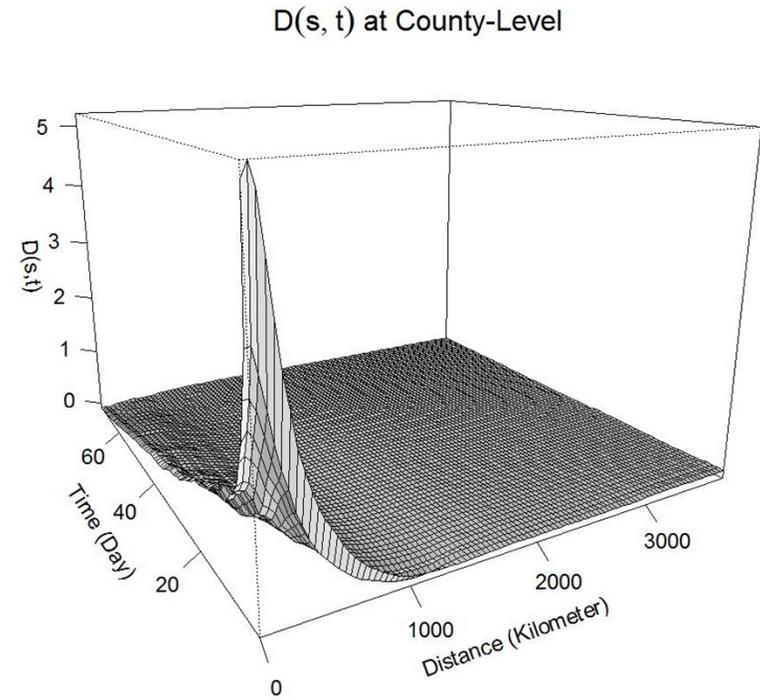
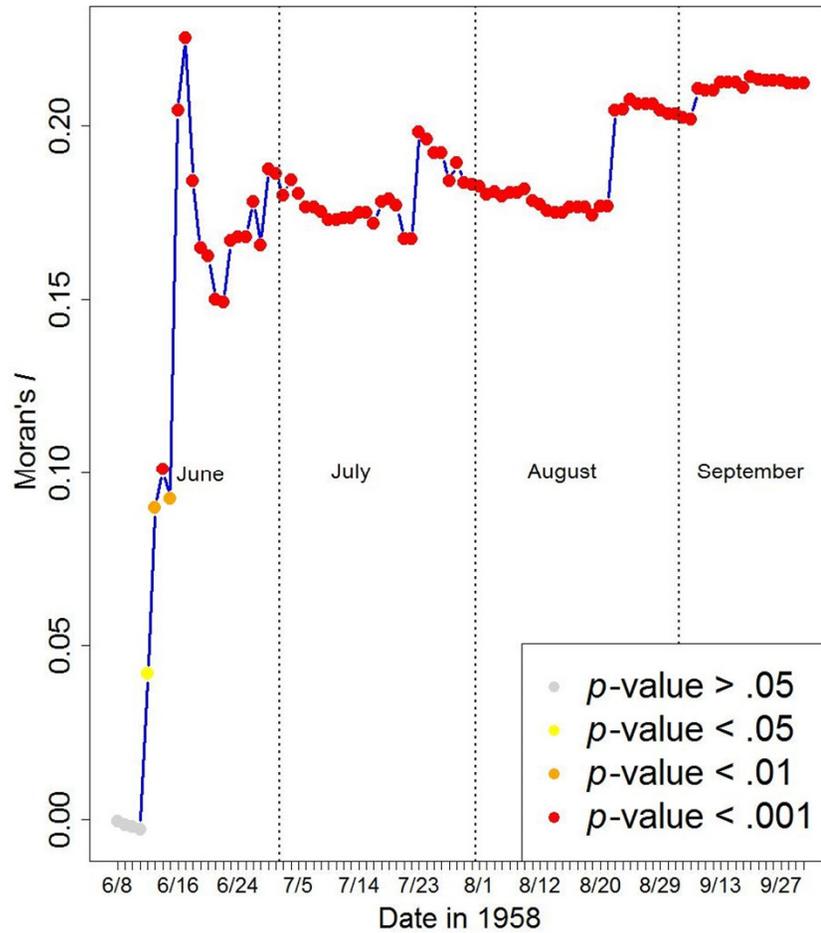
	Full Sample		Subsample	
N of exaggerations of rice/wheat yields	N of Counties	%	N of Counties	%
0	1,944	87.4	294	87.8
1	181	8.1	27	8.1
2	62	2.8	7	2.1
3	20	0.9	5	1.5
4+	18	0.8	2	0.6
Total	2,225	100.0	335	100.0

Spatial Clustering



Animation

Space-Time Clustering



Study 1: Conclusions

- There was a spatial diffusion process of exaggerating grain yields in the summer of 1958
- The diffusion process was mostly driven by the interaction between geographic proximity and political proximity

Consequences of Exaggerating Grain Yields

- A delusion of an unprecedented harvest to China's leaders
 - Excessive compulsory grain procurement
 - Shifted labor and economic resources away from agriculture to fruitless projects (e.g. backyard furnace movement)
 - Reduction in actual grain production
 - Misguided grain ration and redistribution despite the famine
- In combination with other natural and man-made factors
 - The GLF famine of 1959-1961 caused 16.5-30 million excess deaths

Study 2: Long-term Health Effects of the GLF Famine

- The fetal origin hypothesis by Barker (1990, 1995)
 - Prenatal exposure to an adverse environment, in particular to malnutrition, may “program” the fetus to develop particular metabolic characteristics, likely through environmental effects on the epigenome
 - Such developmental changes may persist over the life course and increase risks of cardiovascular and metabolic diseases in middle and later ages

Endogeneity

- Empirical test of the fetal origins hypothesis often relies on observational data
 - For example, use low birth weight as a proxy for fetal malnutrition (from survey data or medical record)
 - Not randomized experimental data
- Omitted variable bias
 - The observed association between birth weight and health outcomes in later life could reflect many unobserved joint determinants such as genetic traits, family socioeconomic status, and environmental factors
- Measurement error
 - Difficult to measure each individual famine survivor's degree of prenatal malnutrition suffered decades ago

GLF Famine as a Natural Experiment

- Exposure to famine is often beyond the control of most individuals, regardless of their genetic traits, personality, or socioeconomic status
- The process governing an individual's prenatal exposure to famine-induced malnutrition is arguably exogenous and resembles random assignment
- We can infer casual effects of prenatal malnutrition on adult health by comparing two similar subpopulations that differ in the famine exposure

Simple Cohort Comparison

- Exploit between-cohort temporal variation: famine vs. non-famine cohorts

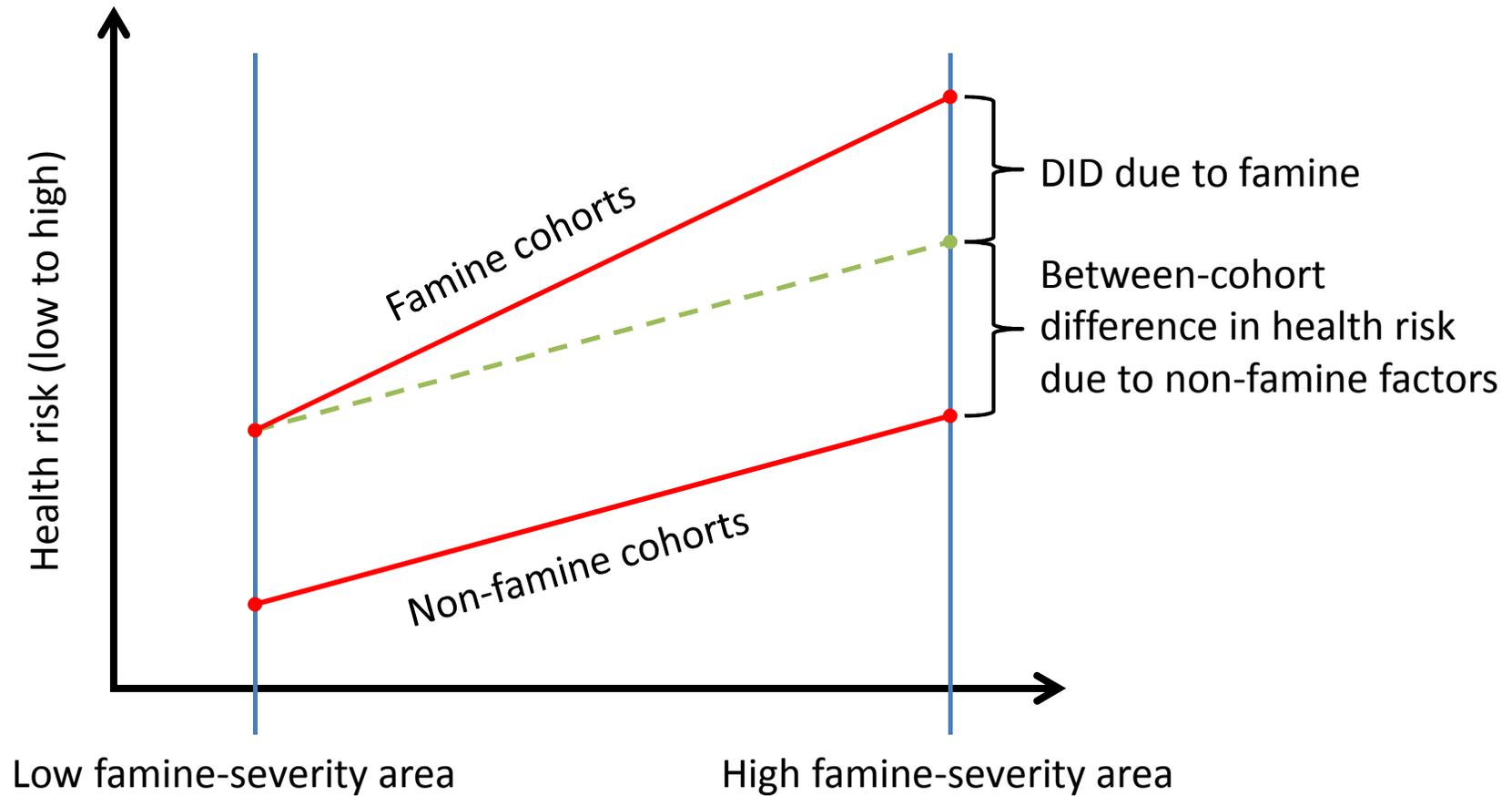
Famine exposure during	Pre-famine cohorts (1956-58)	Famine cohorts (1959-61)	Post-famine cohorts (1962-64)
Prenatal period	No	Yes	No (except those born in the early 1962)
Postnatal period	From ages 1-3 yrs to ages 4-6 yrs	From birth to ages 1-3 yrs	No

- A strong assumption of “constant cohort difference”: the only source of between-cohort difference in the outcome of interest is famine exposure (or not)

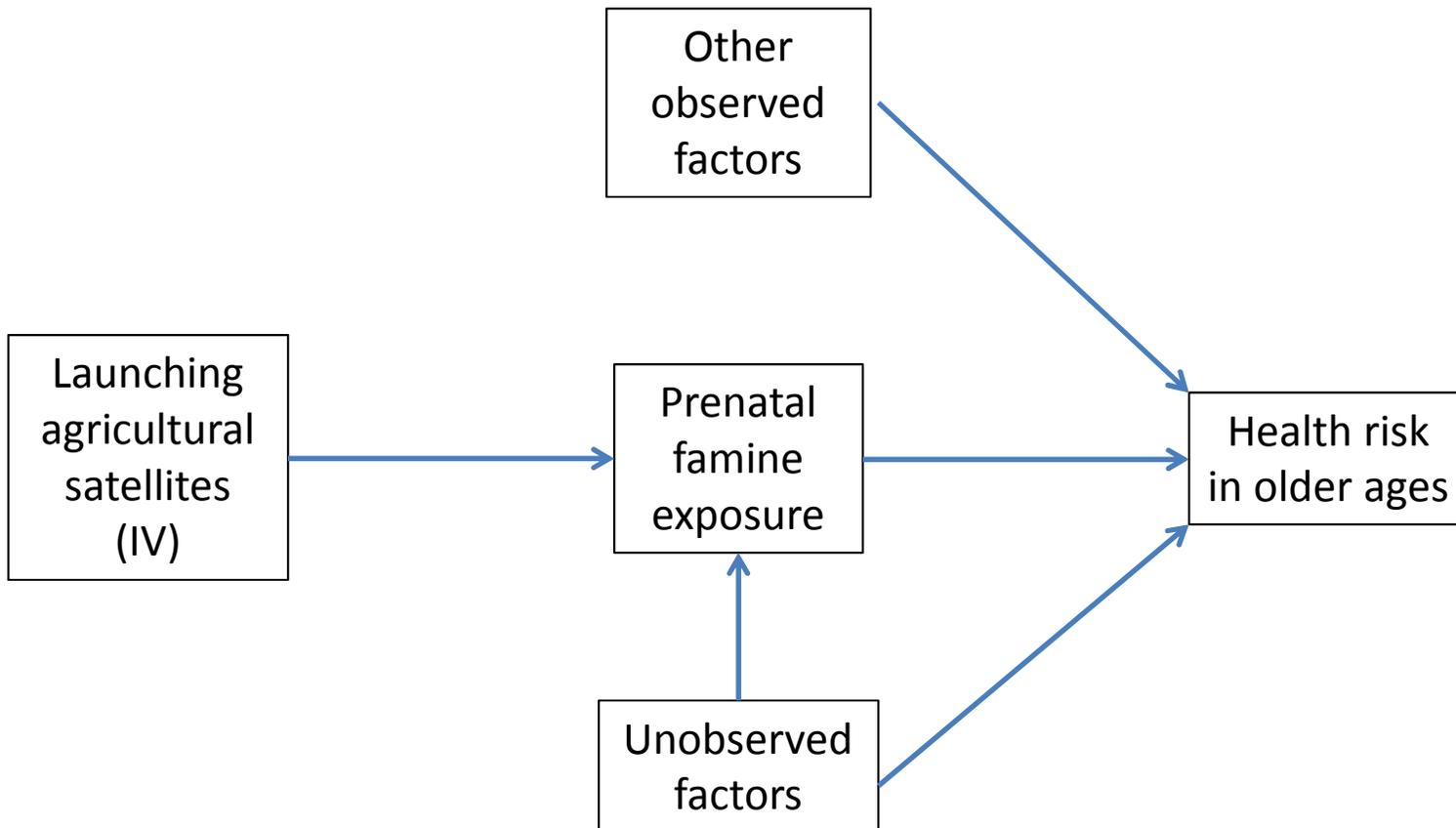
Difference-in-Differences (DID)

- Additionally exploit spatial variation: areas of greater famine severity vs. areas less affected
- A weak assumption
 - other sources of between-cohort difference are possible
 - as long as such difference is constant between areas that were heavily affected by the famine and those that were less affected

Difference-in-Differences (DID)

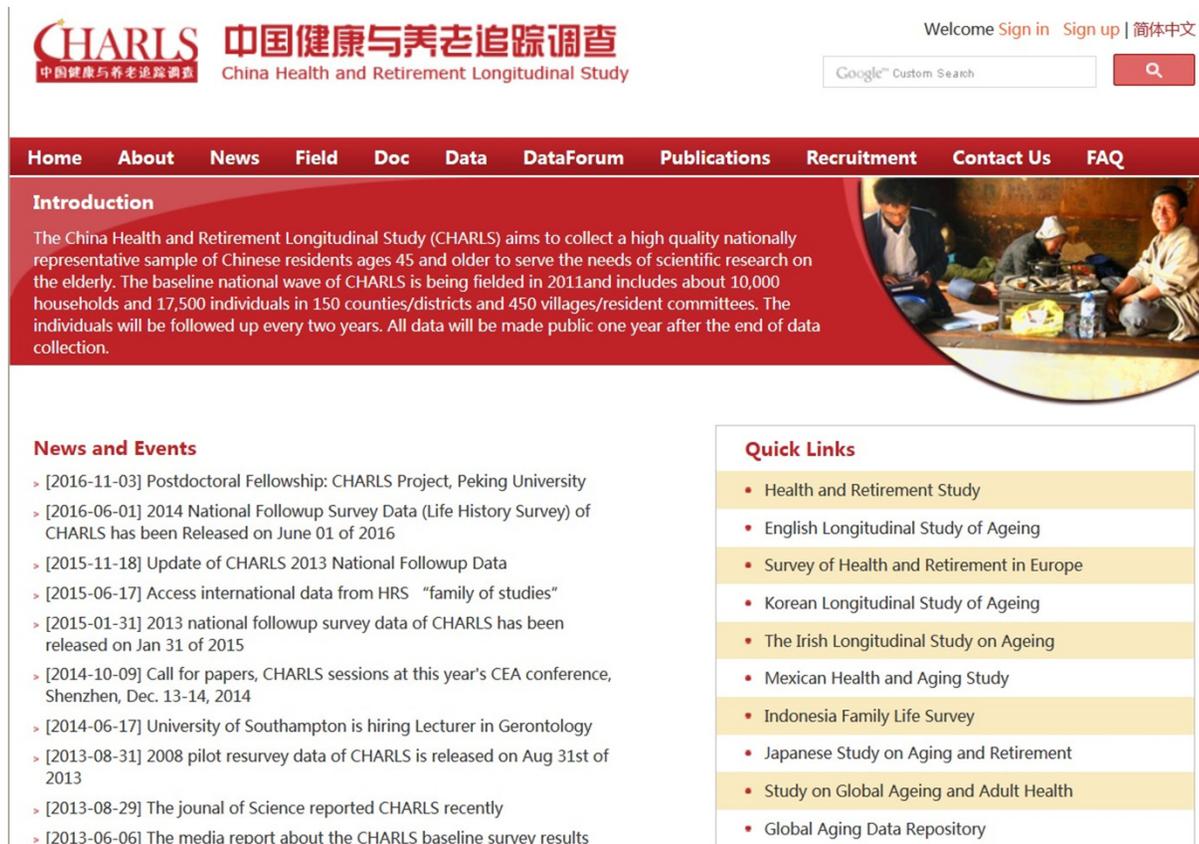


Instrumental Variable (IV)



Data Source: Survey

- Units of analysis: individuals born 1956-64



CHARLS 中国健康与养老追踪调查
China Health and Retirement Longitudinal Study

Welcome Sign in Sign up | 简体中文

Google™ Custom Search

Home About News Field Doc Data DataForum Publications Recruitment Contact Us FAQ

Introduction

The China Health and Retirement Longitudinal Study (CHARLS) aims to collect a high quality nationally representative sample of Chinese residents ages 45 and older to serve the needs of scientific research on the elderly. The baseline national wave of CHARLS is being fielded in 2011 and includes about 10,000 households and 17,500 individuals in 150 counties/districts and 450 villages/resident committees. The individuals will be followed up every two years. All data will be made public one year after the end of data collection.

News and Events

- › [2016-11-03] Postdoctoral Fellowship: CHARLS Project, Peking University
- › [2016-06-01] 2014 National Followup Survey Data (Life History Survey) of CHARLS has been Released on June 01 of 2016
- › [2015-11-18] Update of CHARLS 2013 National Followup Data
- › [2015-06-17] Access international data from HRS "family of studies"
- › [2015-01-31] 2013 national followup survey data of CHARLS has been released on Jan 31 of 2015
- › [2014-10-09] Call for papers, CHARLS sessions at this year's CEA conference, Shenzhen, Dec. 13-14, 2014
- › [2014-06-17] University of Southampton is hiring Lecturer in Gerontology
- › [2013-08-31] 2008 pilot resurvey data of CHARLS is released on Aug 31st of 2013
- › [2013-08-29] The journal of Science reported CHARLS recently
- › [2013-06-06] The media report about the CHARLS baseline survey results

Quick Links

- Health and Retirement Study
- English Longitudinal Study of Ageing
- Survey of Health and Retirement in Europe
- Korean Longitudinal Study of Ageing
- The Irish Longitudinal Study on Ageing
- Mexican Health and Aging Study
- Indonesia Family Life Survey
- Japanese Study on Aging and Retirement
- Study on Global Ageing and Adult Health
- Global Aging Data Repository

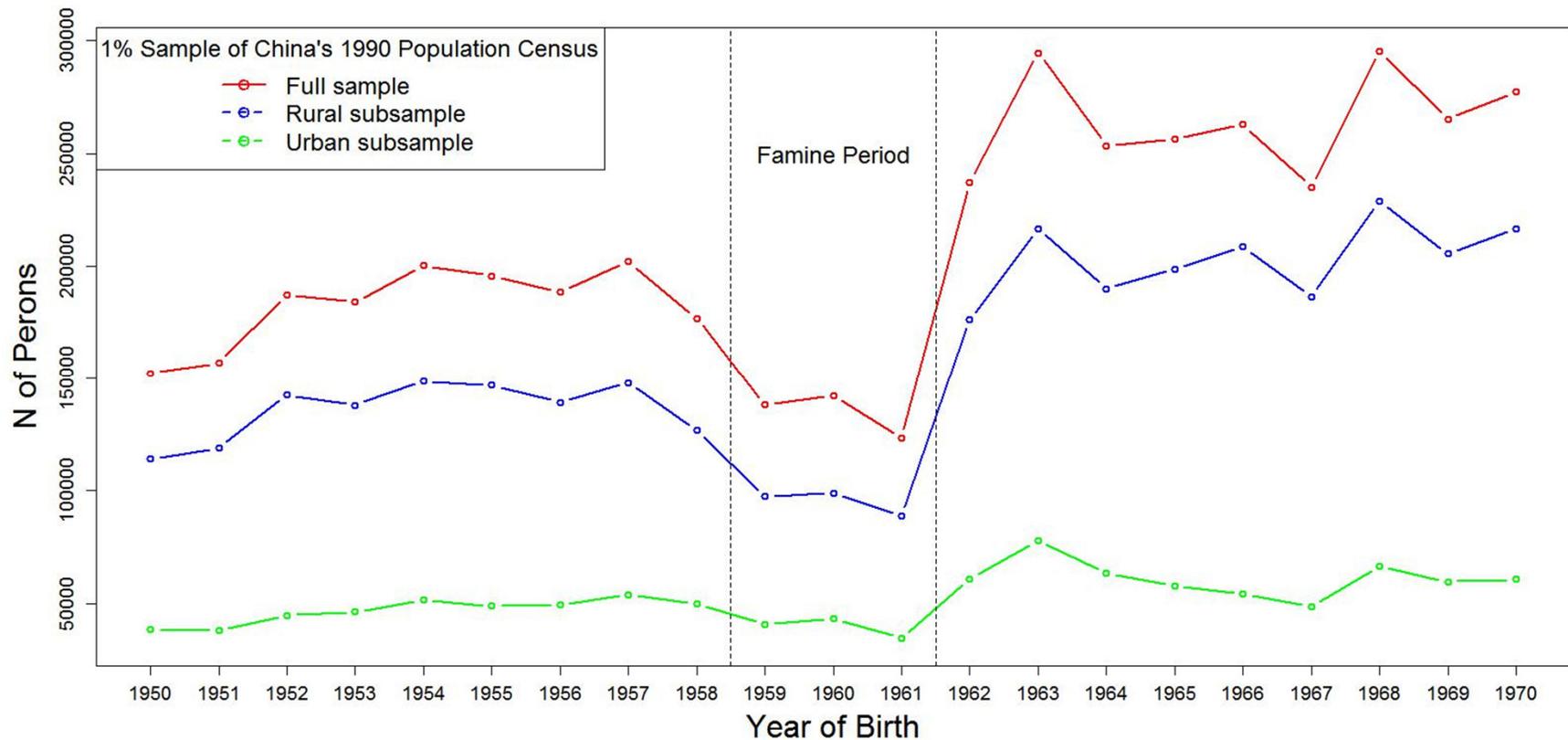
Data Source: Survey

- China Health and Retirement Longitudinal Study (CHARLS)
 - Nationally representative sample of Chinese adults aged 45 and older and their spouses, if available
 - 2011 baseline survey
 - Restricted to the rural subsample in this study
 - Anthropometric measures: weight, height, waist circumference
 - Biomarkers: blood sample
 - Self-reported birth prefecture

Table 1. Descriptive Statistics of Health Outcomes by Birth Cohorts (CHARLS 2011)

Dependent Variables	High-Risk Cut-Points	Pre-Famine		Famine		Post-Famine	
		%	(N)	%	(N)	%	(N)
Cardiovascular							
Diastolic BP	>=90 mmHg	15.9	(1,201)	17.5	(804)	14.3	(1,407)
Systolic BP	>=140 mmHg	25.0	(1,201)	21.6	(804)	17.3	(1,407)
Resting Pulse	>100 beats/min	11.8	(1,334)	11.7	(900)	11.5	(1,568)
Dyslipidemia							
HDL Cholesterol	<40 mg/dL in men; <50 mg/dL in women	40.0	(1,030)	42.0	(686)	41.8	(1,129)
LDL Cholesterol	>160 mg/dL	11.1	(1,028)	10.7	(684)	6.8	(1,129)
Total Cholesterol	>=240 mg/dL	12.1	(1,028)	9.8	(686)	9.1	(1,129)
Triglyceride	>=150 mg/dL	15.4	(1,030)	17.2	(686)	15.7	(1,129)
Diabetes							
Glucose	>=126 mg/dL	12.8	(1,027)	13.0	(684)	9.8	(1,127)
HbA1c	>=6.5%	4.3	(1,039)	4.1	(689)	3.5	(1,139)
Anthropometric							
Height (cm; mean)		159.6	(1,188)	159.6	(789)	160.0	(1,400)
Body Mass Index (mean)		23.5	(1,184)	24.0	(788)	24.1	(1,396)
Overweight	BMI >=25	31.5	(1,184)	35.3	(788)	37.0	(1,396)
Abdominal Obesity	Waist circumference >90 cm in men; >80 cm in women	44.8	(1,195)	48.6	(796)	50.4	(1,404)

Birth Cohort Size Counted in 1990



Measure Famine Severity

- Prefecture-level cohort size shrinkage index (CSSI)
- $$CSSI = \frac{N_{nonfamine} - N_{famine}}{N_{nonfamine}} \times 100\%$$
 - $N_{nonfamine}$ is the average cohort size of those born during the three years preceding the famine (1956-58) and the three years after the famine (1962-64)
 - N_{famine} is the average cohort size of those born during the three famine years (1959-61)
 - Calculated for each prefecture
- A larger CSSI value indicates a greater reduction in cohort size due to
 - reduced fertility
 - increased infant mortality
 - both presumably caused by the GLF famine

Measure Famine Severity

- Calculate prefecture-level CSSI from the 1% sample of the 1990 China Population Census
- Matched to the CHARLS respondents based on their self-reported birth prefecture

MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA



Home | Select Data | FAQ | Help | Login

PROJECT About IPUMS-I How to Cite IPUMS-I User Registration and Login	<h2>Integrated Public Use Microdata Series, International</h2> <p>census microdata for social and economic research</p> <hr/> <p>IPUMS-International is a project dedicated to collecting and distributing census data from around the world. Its goals are to:</p> <ul style="list-style-type: none">• Collect and preserve data and documentation• Harmonize data• Disseminate the data absolutely free! <hr/> <p>82 countries - 277 censuses - 614 million person records</p> <hr/> <p>Source data for IPUMS-International are generously provided by participating National Statistical Offices</p>	IPUMSI News NEW! 19 samples added NSF extends IPUMS project thru 2019 Geography variables revised Online tabulator Class accounts Latin America project extended Integrated DHS project funded ... All news items
DATA Browse and Select Data Download Your Data Extract Geography and GIS Supplemental Data Files Analyze Data Online		MPC Data Projects IPUMS-USA and others
SAMPLES Sample Descriptions Source Documents Variance Estimation		Sign up to receive News of interest to IPUMS-International data users and producers.
RESOURCES Help User Forum Bibliography International Partners Teaching with IPUMS		

Funding provided by: [National Science Foundation](#), [National Institutes of Health](#), and [Stat/Transfer](#).
Copyright © Regents of the [University of Minnesota](#). All rights reserved.

Spatial Variations in Famine Severity (CSSI) and IV

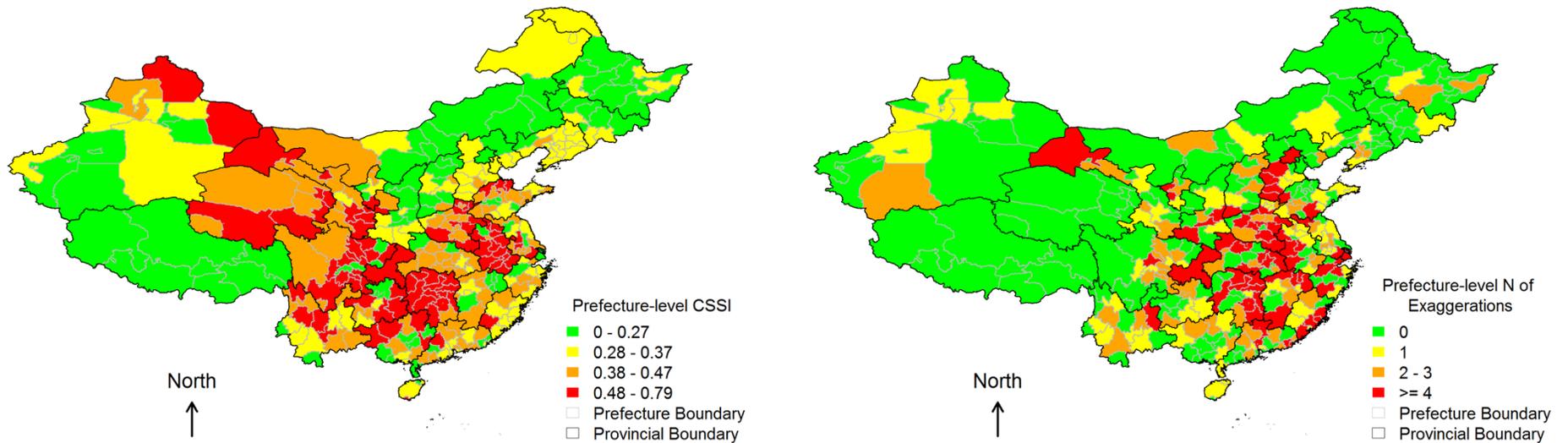


Table 2. Selected Regression Estimates of the Causal Health Effects of Famine Exposure

Dependent Variables	SCD		DID		2SLS
	Pre-Famine	Famine	CSSI x Pre-Famine	CSSI x Famine	Famine Severity
Cardiovascular					
Diastolic BP	0.074	0.119	-0.005	-0.004	0.001
Systolic BP	0.271***	0.152*	-0.003	-0.004	-0.004
Resting Pulse	0.041	0.013	-0.005	-0.001	0.002
Dyslipidemia					
HDL Cholesterol	-0.020	-0.037	-0.003	0.000	0.004
LDL Cholesterol	0.270***	0.272***	-0.007	-0.022***	-0.004
Total Cholesterol	0.166**	0.044	-0.010†	-0.019**	-0.005†
Triglyceride	-0.003	0.062	-0.002	-0.001	0.000
Diabetes					
Glucose	0.207**	0.221*	0.007	-0.001	-0.001
HbA1c	0.139	0.097	0.006	0.001	-0.001
Anthropometric					
Height (cm)	-1.032***	-0.640†	-0.021	0.015	-0.167†
Body Mass Index	-0.411**	-0.091	-0.019†	-0.019†	0.007
Overweight	-0.093†	-0.047	-0.002	-0.008†	-0.004
Abdominal Obesity	-0.018	-0.025	-0.004	0.005	0.010

Note: SCD = simple cohort difference; DID = difference-in-differences; 2SLS = two-stage least squares; CSSI = cohort size shrinkage index. All the models control for gender, marital status, educational attainment, household asset, smoking, and drinking. The SCD and DID models additionally control for provincial fixed effects.

† p < 0.1; * p < 0.05; ** p < 0.01; *** p < 0.001 based on city-level cluster standard errors.

Study 2: Conclusions

- Simple cohort comparisons reveal increased later-life health risks due to famine exposure
- DID and IV estimates suggest that famine either has no effect or reduces later-life health risks
 - likely due to mortality selection

Overall Concluding Remarks

- Collaboration with people
 - From the same discipline but with different skills
 - From different disciplines
- “Old” (secondary) data but new ideas
- Sometimes (or very often?) one data source is not enough
- Ask your colleagues and librarians!