Demographic Surveillance

in the contemporary developing world

Introduction and Examples of Analysis

Samuel Clark

Department of Sociology, University of Washington
Institute of Behavioral Science, University of Colorado at Boulder
Agincourt Health and Population Unit & Computational and Applied Maths, University of the Witwatersrand

Who I am

- *Demographer* trained in biology, computer science and demography
- Work in Africa using demographic surveillance system data:
  - Descriptive empirical demography
  - Evaluate population-level impacts of HIV
  - Model populations affected by HIV
  - Develop data management methods for longitudinal data
- All my work involves contemporary longitudinal data similar to traditional historical demography data sets

Today

- Informal discussion of the aspects of my work that share some similarity with historical demography
- Longitudinal data
  - Structure
  - Management
  - Analysis
- Please do interrupt and ask questions as we go along

Demographic Surveillance

- Continuous monitoring of geographically defined populations
  - ~ 10^3 people
  - ~ 10 years of observation
  - Monitor:
    - Vital events: birth, death
    - Movement: in/out/internal migration
    - Sometimes marriage
    - Variety of other attributes, many clinical in nature
- Result is complex longitudinal data sets that describe the histories of individuals and the populations they comprise, including social connections

INDEPTH Network

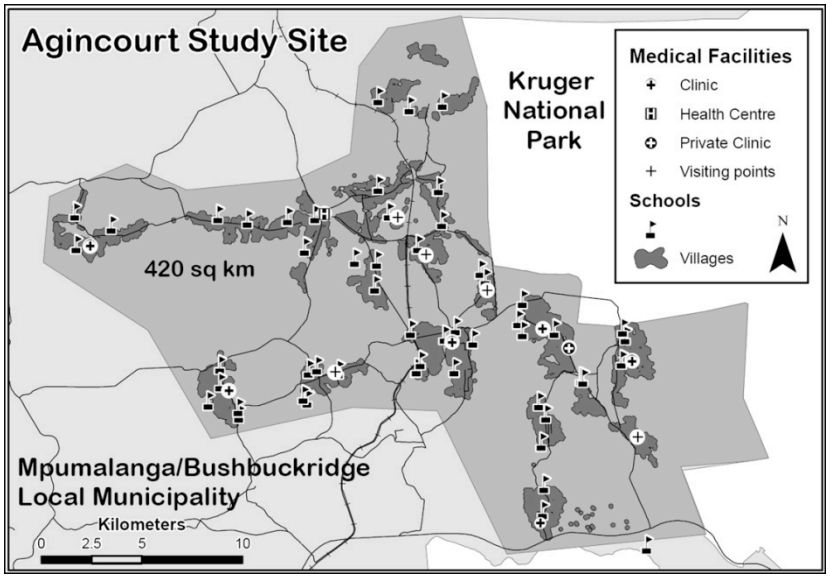
- About 40 DSS sites; mostly in developing world; mostly in Africa
- INDEPTH Network is a professional body to which DSS sites belong
  - Useful web site: **http://www.indepth-network.org**
  - Organizes intra/cross-site analytical work

    Example: INDEPTH Network. 2002. Population, Health, and Survival at INDEPTH Sites, vol. 1, Population and Health in Developing Countries. Edited by INDEPTH Network. Ottawa: IDRC Press.
  - Raises funds
  - Clearinghouse for ideas
  - Working on DSS data archive
  - Organizes informative annual meeting

Example Site: Agincourt, South Africa

- Demographic surveillance site in northeast South Africa
- Data collecting through annual census rounds starting in 1992
- ~70,000 people under observation
- Data describes vital events; births deaths, migrations, and more
- Additional population-wide data collected at individual and population levels

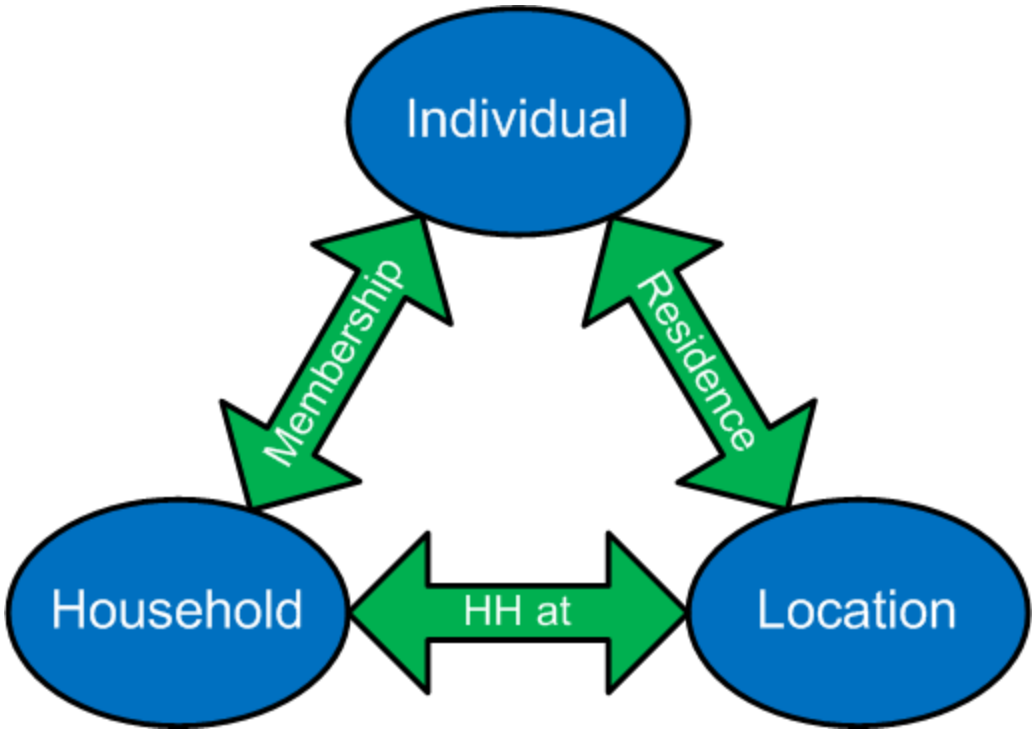
- Additional nested projects collect data using DSS population as a sampling frame for targeted studies



http://www.agincourt.co.za/DataSection/index.htm

DSS Data

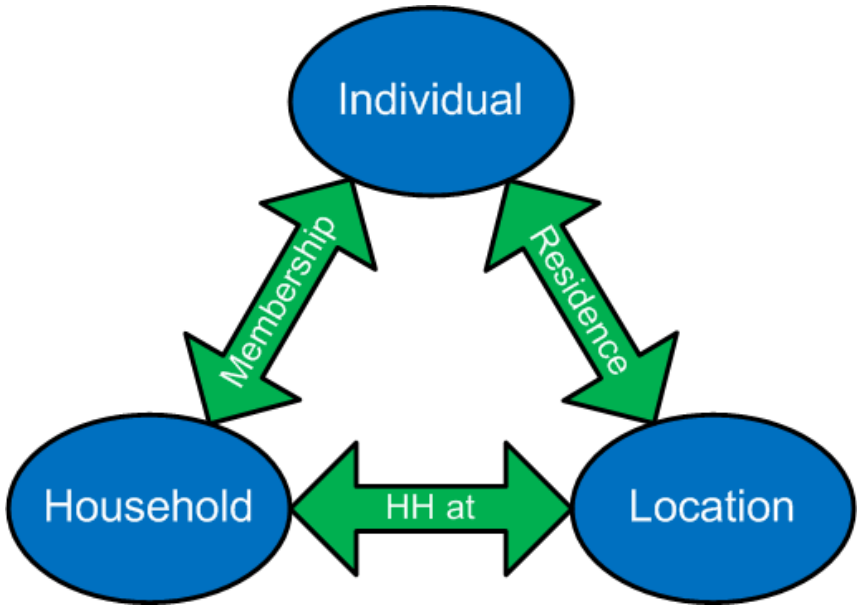
- DSS = **D**emographic **S**urveillance **S**ystem
- The core data at most DSS sites organized around the 'triangle schema':



DSS Data

- Entities (blue) have both constant and time-varying attributes and usually a beginning and end
- "Memberships" (green) are defined by the entities they link and their start and end times

➔ Both the entities and the memberships can be conceived as ***episodes***



Conceptual Generalization

- The episodes associated with entities and memberships can be collapsed into a single entity called a ***state***
- The ***events*** that define the start and end of the episodes can all be stored together
- The events can be linked to the states using an ***influence***
- The ***event-influence-state*** triad handles all the time-varying linking and attributing
- Relationships group states in non time-dependent ways

DSS Data Management

- DSS data typically managed in relational databases using a structure that reflects the ‘triangle schema’
  - The Reference Data Model

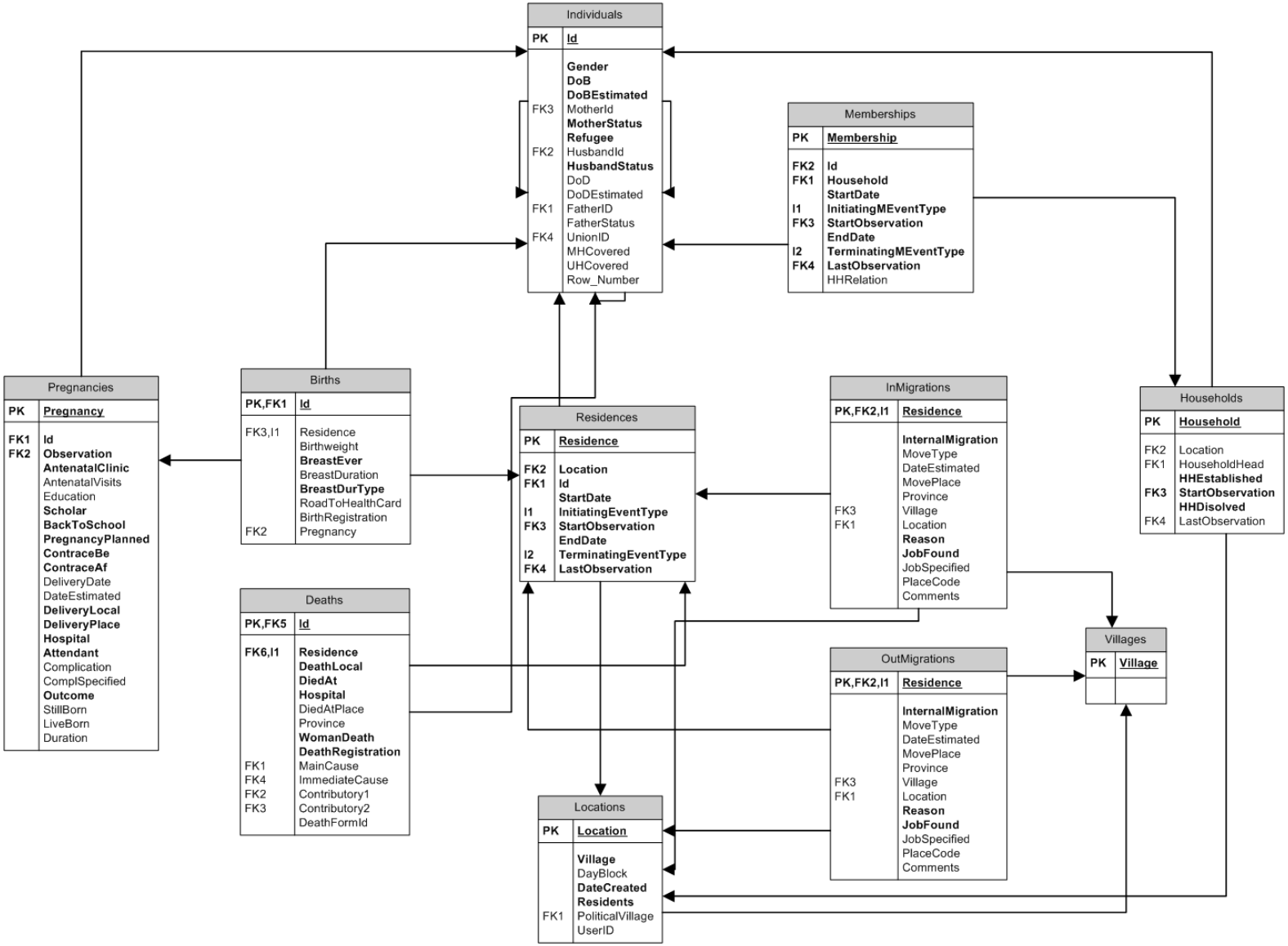
Benzler, Justus, Kobus Herbst and Bruce MacLeod. 1998. "A Data Model for Demographic Surveillance". http://www.indepth-network.org/publications/zindpubs/DM_for_Demographic.htm

  - Structured Population Event History Register

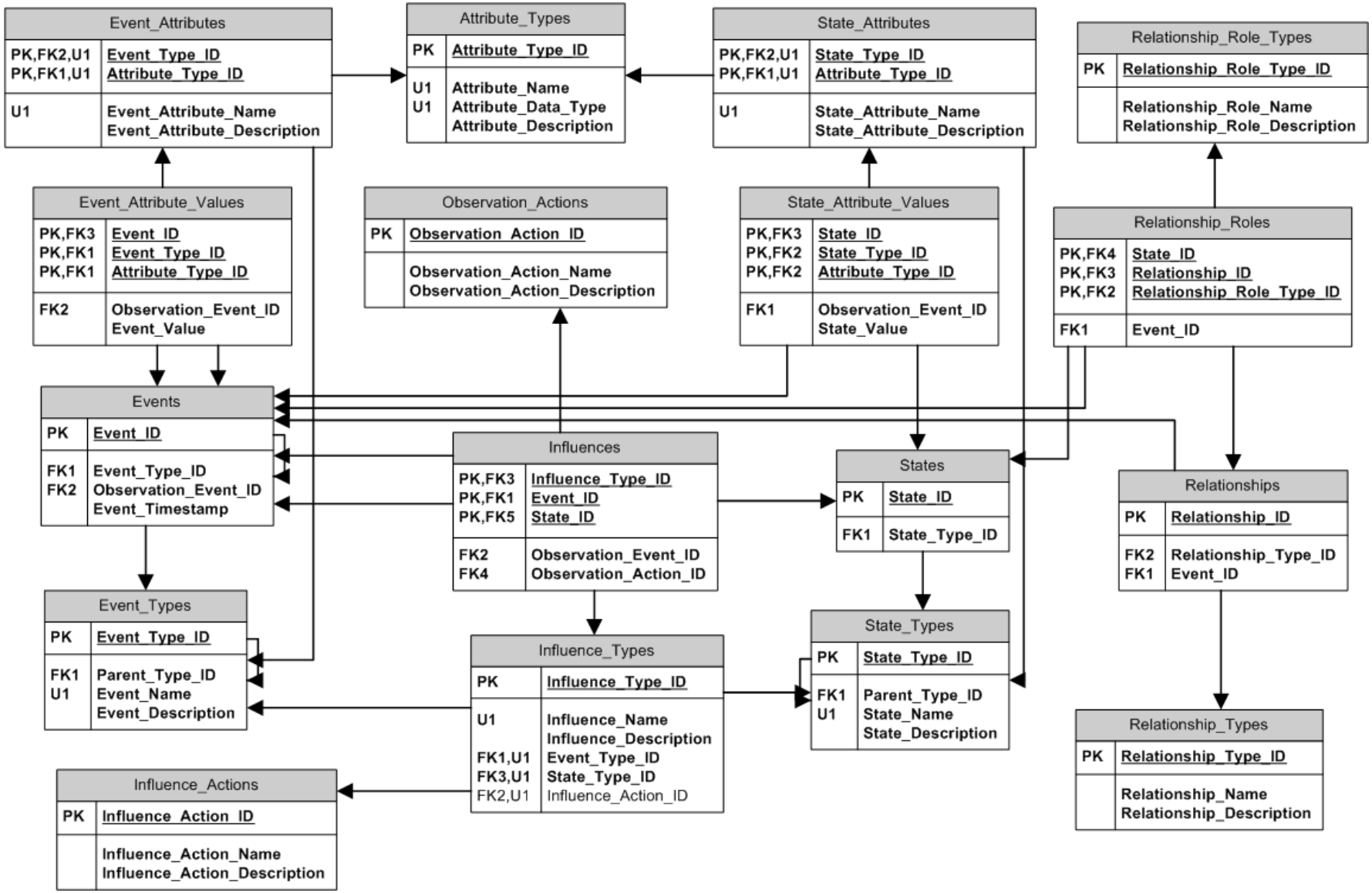
Clark, Samuel J. 2006. "A General Temporal Data Model and the Structured Population Event History Register." Demographic Research, 15(7):181-252. <http://www.demographic-research.org/Volumes/Vol15/7/default.htm>.

- Common relational database management systems include:
  - MS SQL Server (there’s a free Express edition)
  - Foxpro, PostgreSQL, MySQL, etc.
- Data manipulation/preparation done using ***SQL***

Typical RDM Schema (greatly simplified)



SPEHR Schema



Analyzing DSS Data

- Time is an inherent feature of DSS data

  ➔ **requires longitudinal data analysis techniques**
- Event history analysis
- Survival analysis
- Proportional hazards models (Cox regression, etc.)
- Discrete time event history analysis

My *Personal* Preference for Analysis

- I strongly prefer a discrete time event history analytical (DTEHA) approach for DSS data, in comparison to Cox regression:
  - Assumption of proportional hazards almost never holds over long periods of time in complex dynamic situations
  - Specification of complex multivariate models in Cox regression framework with time-varying explanatory variables can be very tricky … *IF* proportionality assumption is met
- DTEHA shifts the difficult part to the data management/ data preparation phase of the work (which is easier for me!) and makes the statistical models comparatively simple to specify, run and interpret

Discrete Time Event History Analysis 1

1. Prepare an entity-time data file (the hard part)
   - Unit of analysis is entity-time; e.g. person-year, household-month, etc.
   - Response variable indicator coded (usually binary) and reflects the occurrence of something during the time duration of the entity-time unit; e.g. death during a person-year
   - Explanatory variables given values at the beginning of the entity-time unit, e.g. age and sex at the beginning of a person-year
   - ***Censoring***: only include periods of time that are fully observed – this is critical and difficult
   - Accomplished by writing SQL to manipulate and extract data from relational database tables

Discrete Time Event History Analysis 2

2. Define and estimate a model
   - Hypotheses articulated as probability of response conditional on values of explanatory variables at the beginning of entity-time unit; e.g. probability of death conditioned on age, sex and calendar year during which person-year starts
   - Estimated using logistic or multinomial logistic regression using maximum likelihood
4. Interpret regression results and calculate predicted probabilities
5. Display predicted probabilities and standard errors for exposition

Examples of Analysis

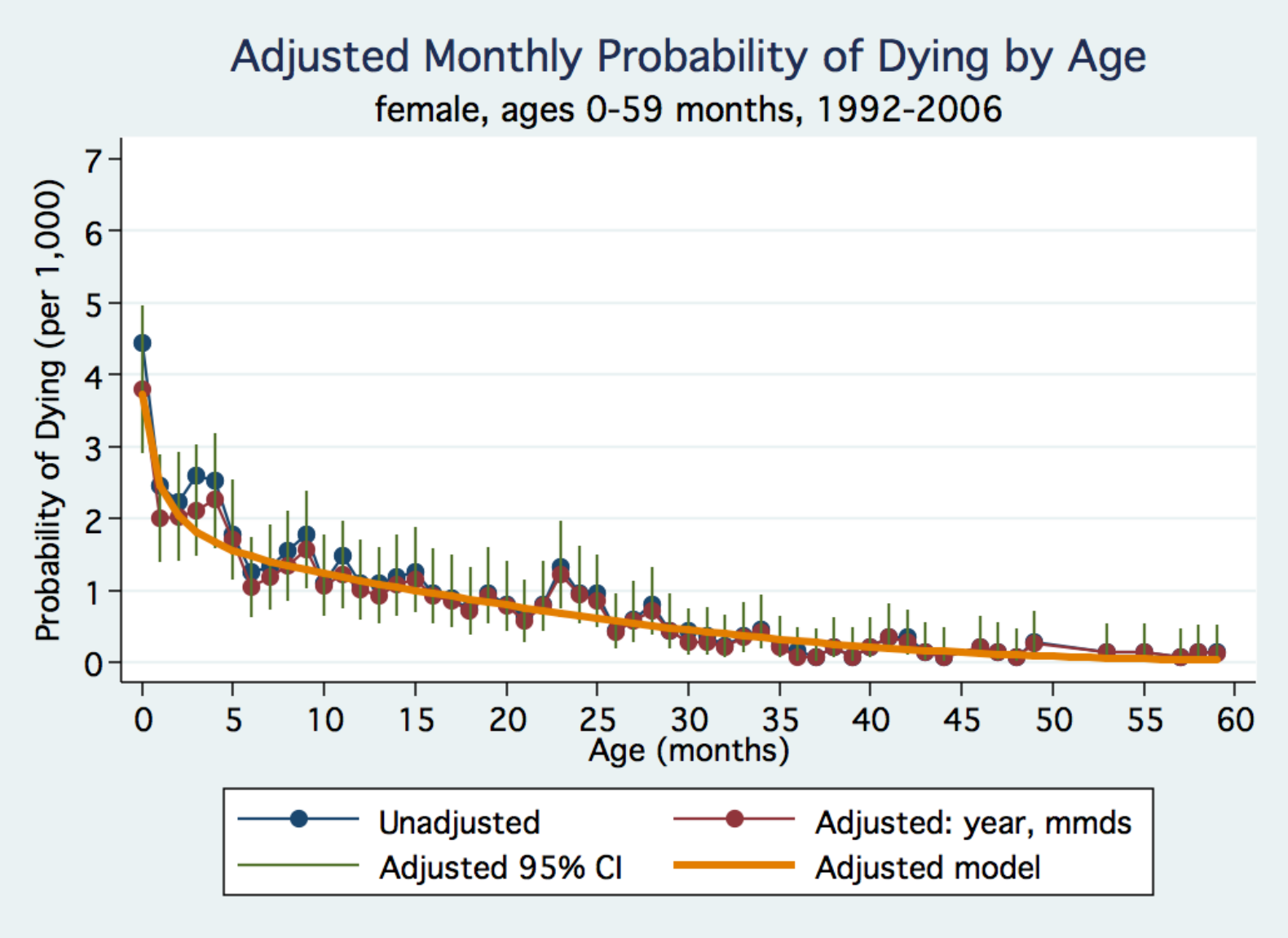
1. Returning Home to Die

2. Orphan Mortality

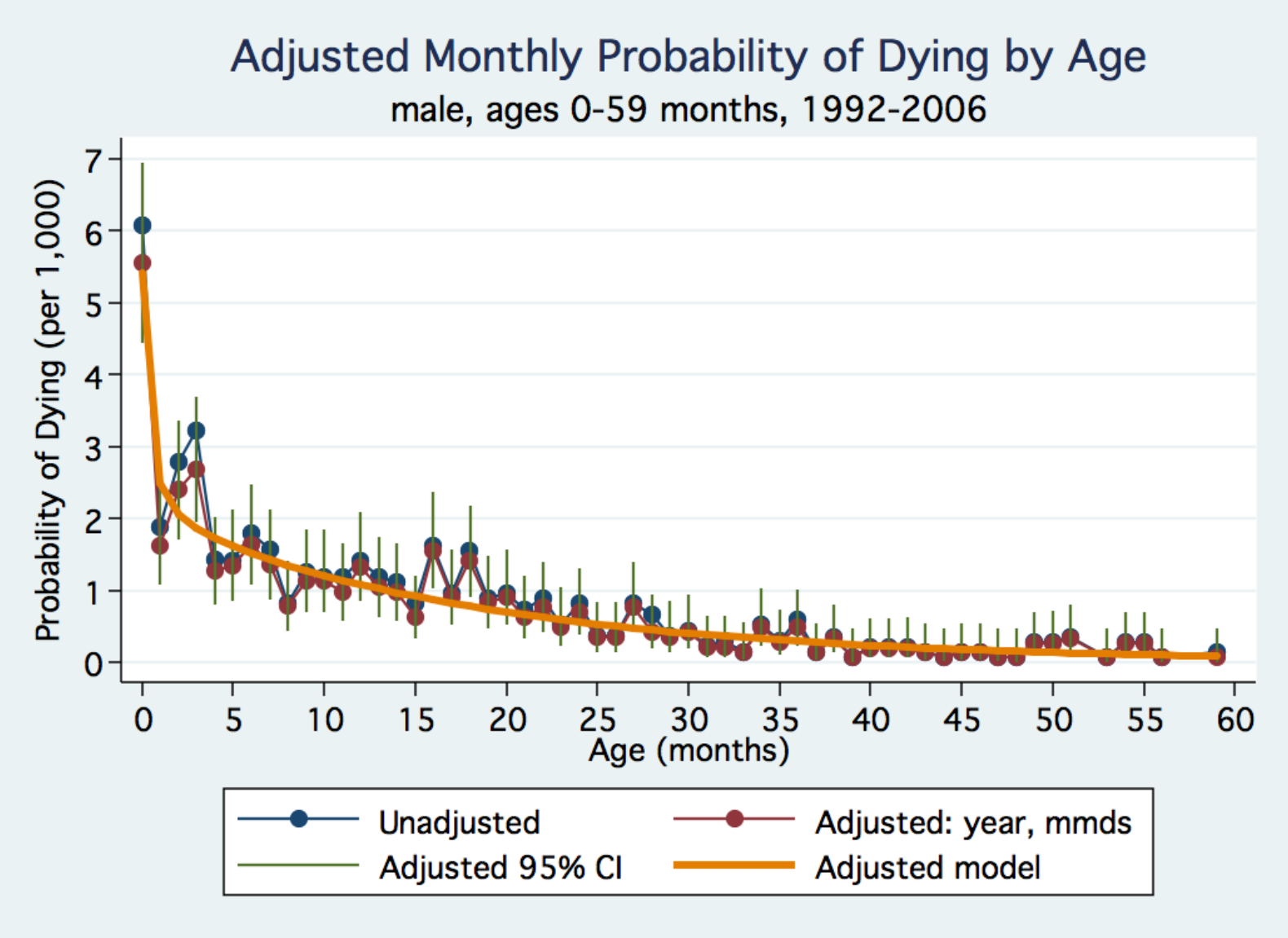
Orphan Mx

- Investigate the probability of dying conditional on age and months since or untiol mother's death controlling for SES of household
- Prepare child-month analysis file:
  - Granularity of time is 'month'
  - Entity is children (0-59 months of age)
  - Response is death
  - Explanatory variables include age, sex, household SES and months since or until mother's death
- Requires mother-child link, vital dates, sex and individual-household links and a little SQL

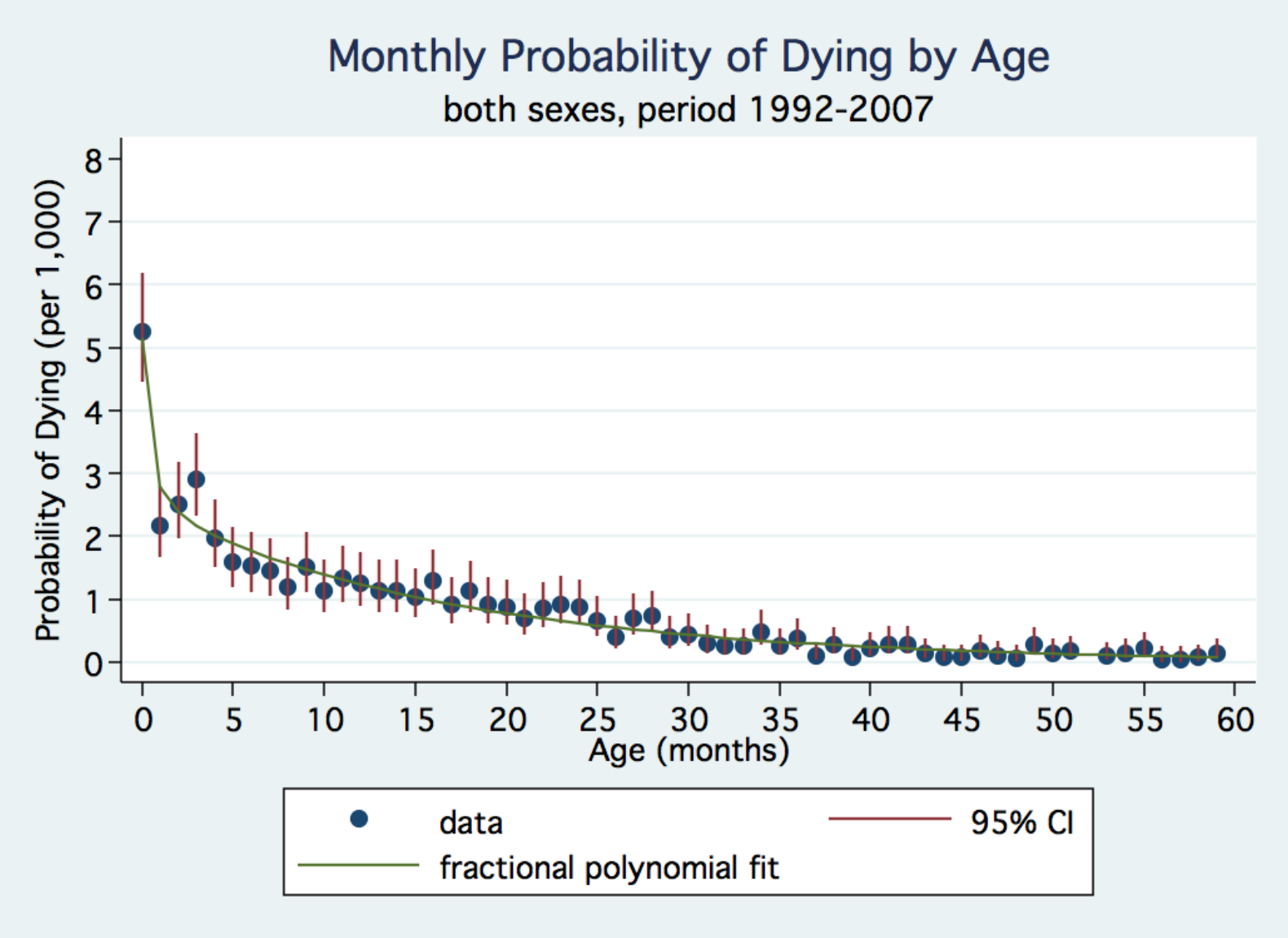
Orphan Mx


Adjusted Monthly Probability of Dying by Age
female, ages 0-59 months, 1992-2006
Probability of Dying (per 1,000)
0
1
2
3
4
5
6
7
0
5
10
15
20
25
30
35
40
45
50
55
60
Age (months)
Unadjusted
Adjusted: year, mmds
Adjusted 95% CI
Adjusted model

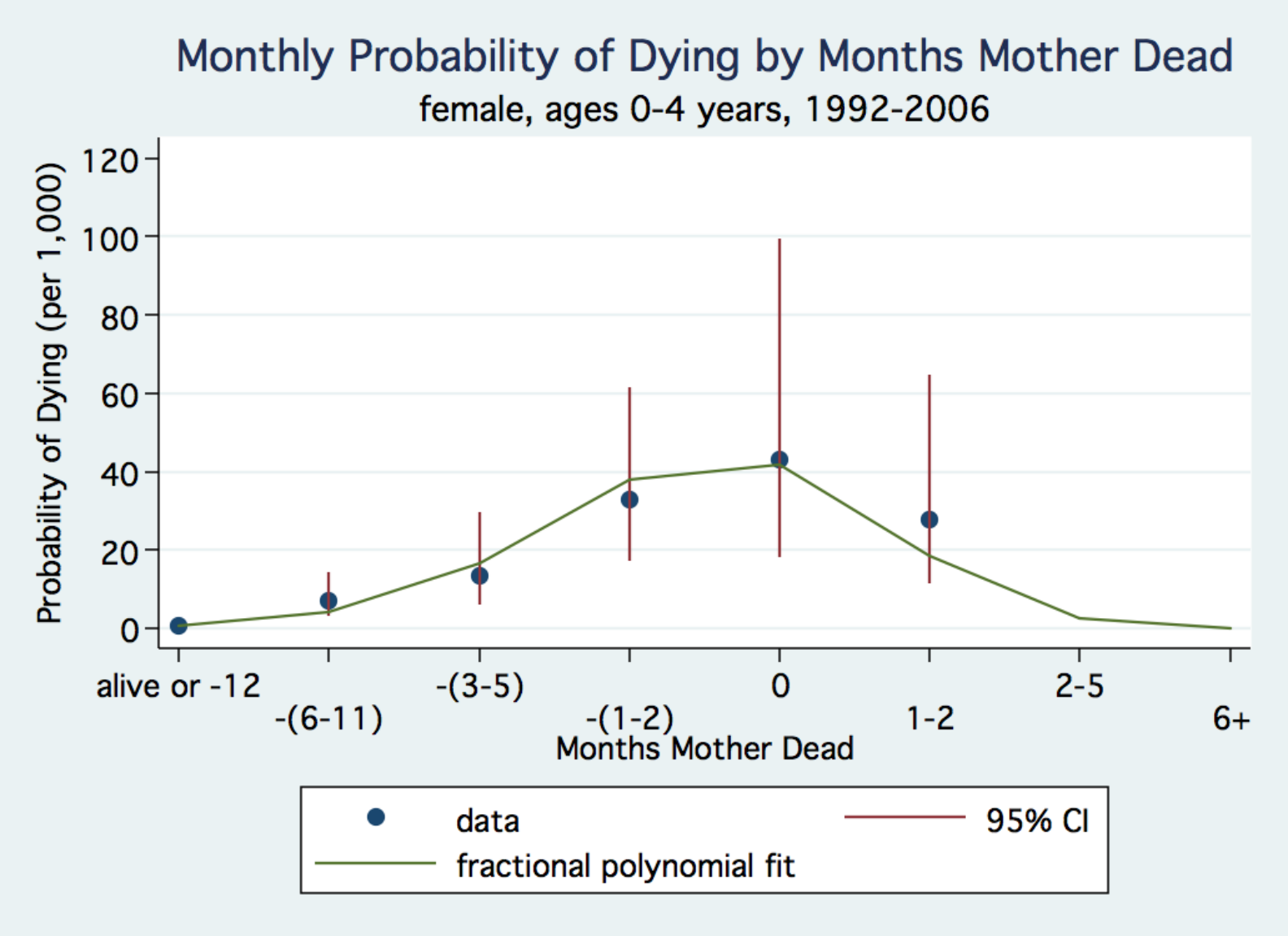
Orphan Mx


Adjusted Monthly Probability of Dying by Age
male, ages 0-59 months, 1992-2006
Probability of Dying (per 1,000)
0
1
2
3
4
5
6
7
0
5
10
15
20
25
30
35
40
45
50
55
60
Age (months)
Unadjusted
Adjusted: year, mmds
Adjusted 95% CI
Adjusted model

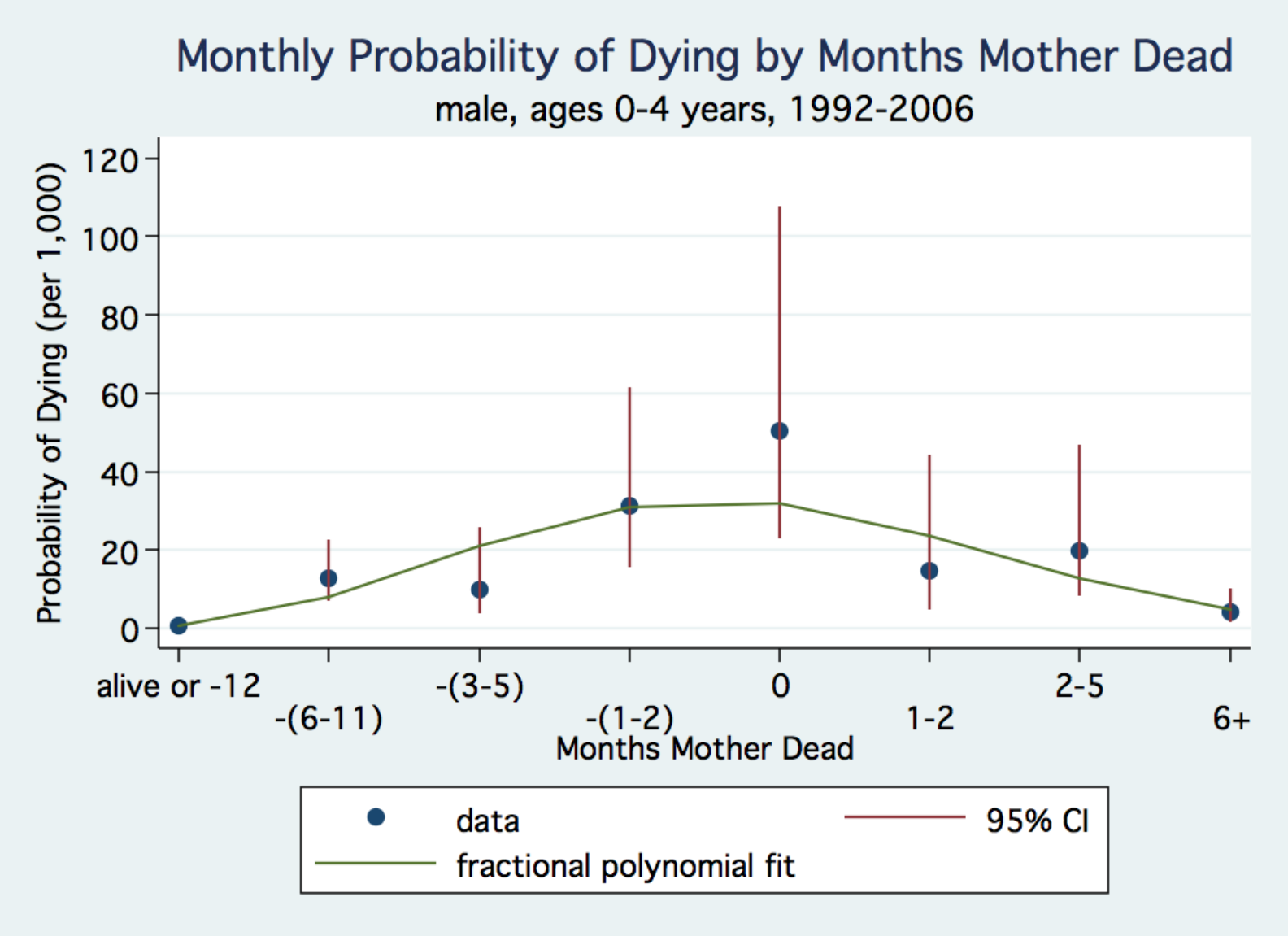
Orphan Mx


Monthly Probability of Dying by Age
both sexes, period 1992-2007
Probability of Dying (per 1,000)
0
1
2
3
4
5
6
7
8
0
5
10
15
20
25
30
35
40
45
50
55
60
Age (months)
data
95% CI
fractional polynomial fit

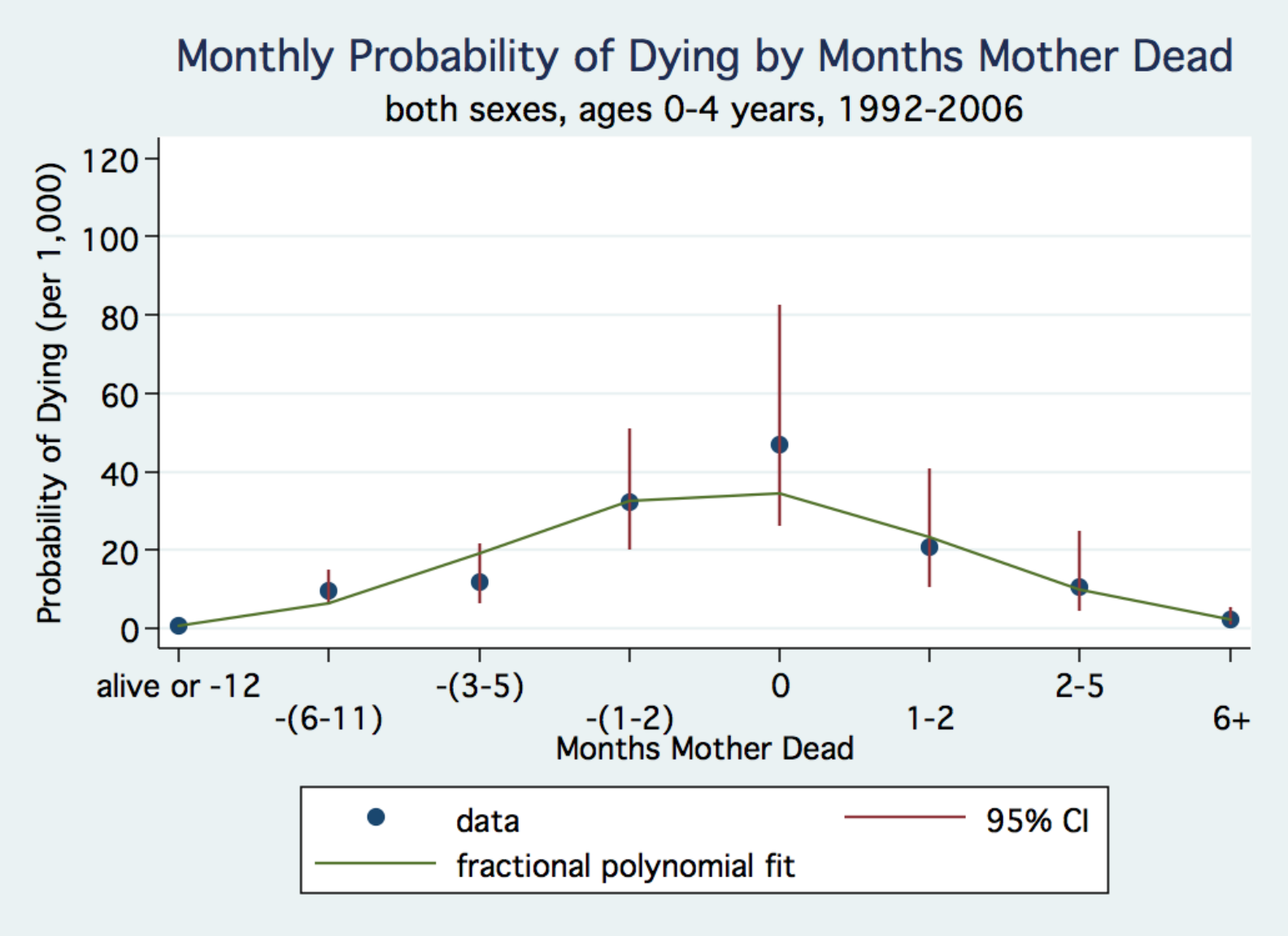
Orphan Mx


Monthly Probability of Dying by Months Mother Dead
female, ages 0-4 years, 1992-2006
Probability of Dying (per 1,000)
120
100
80
60
40
20
0
alive or -12
-(6-11)
-(3-5)
-(1-2)
0
1-2
2-5
6+
Months Mother Dead
data
95% CI
fractional polynomial fit

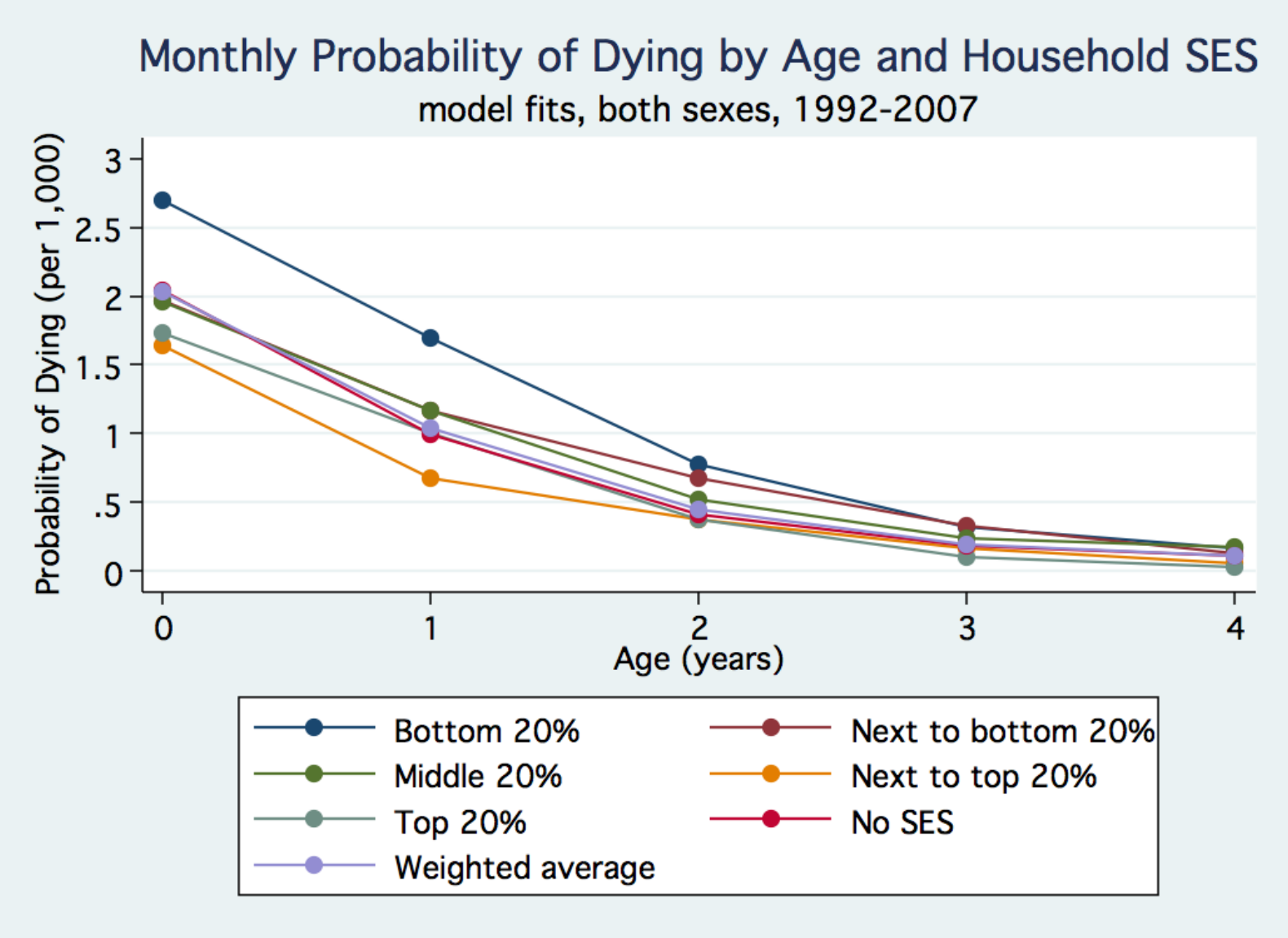
Orphan Mx


Monthly Probability of Dying by Months Mother Dead
male, ages 0-4 years, 1992-2006
Probability of Dying (per 1,000)
120
100
80
60
40
20
0
alive or -12
-(6-11)
-(3-5)
-(1-2)
0
1-2
2-5
6+
Months Mother Dead
data
95% CI
fractional polynomial fit

Orphan Mx


Monthly Probability of Dying by Months Mother Dead
both sexes, ages 0-4 years, 1992-2006
Probability of Dying (per 1,000)
120
100
80
60
40
20
0
alive or -12
-(6-11)
-(3-5)
-(1-2)
0
1-2
2-5
6+
Months Mother Dead
data
95% CI
fractional polynomial fit

Orphan Mx


Monthly Probability of Dying by Age and Household SES
model fits, both sexes, 1992-2007
Probability of Dying (per 1,000)
3
2.5
2
1.5
1
.5
0
0
1
2
3
4
Age (years)
Bottom 20%
Next to bottom 20%
Middle 20%
Next to top 20%
Top 20%
No SES
Weighted average