Understanding and Measuring the Impact of Distance on Health
Evidence from Two Studies

Mahesh Karra
Pardee School of Global Studies
Boston University
October 3, 2017
Background and Motivation
Background

• Despite progress to reducing child mortality, nearly 18,000 children under 5 die every day
• Many of these deaths could be avoidable with increased utilization of health services
• But health service utilization by women around the world remains low
Motivation

• A large theoretical and empirical literature on geographical determinants for health care seeking and MCH outcomes
• Role of physical access (travel distance) to services
• Evidence of association between distance to facility and utilization has been generally consistent
• Empirical evidence on association between distance to facility and health outcomes (e.g. child mortality) is limited and mixed
• Methodological concerns around how distance is measured
  • Travel distance (Euclidean, road), travel time
  • Issues around measurement error and bias in distance
Objectives

• To understand how distance is related to utilization and health
• To explore measurement problems with distance data
• To propose a methodological solution to these problems
Objectives

Study 1 Objectives

• To empirically examine the relationships between
  • Travel distance to facility and health care utilization
    • Receipt of antenatal care
    • Delivery in a health facility
  • Travel distance to facility and health
    • Child mortality
Objectives

Study 2 Objectives

• To develop a theory that allows for unbiased and consistent estimation when we have deliberately induced measurement error in our distance data
  • And mismeasured explanatory variables, more generally
Facility Distance and Child Mortality: A Study of Health Facility Access, Service Utilization, and Child Health

M. Karra, G. Fink, and D. Canning
Objectives

• To examine the relationships between
  • Travel distance to facility and maternal health care utilization
    • Receipt of antenatal care (WHO-recommended 4 visits)
    • Delivery in a health facility
  • Travel distance to facility and child mortality
    • Disaggregated into neonatal, post-neonatal infant, and post-infant child
Data and Methods

- Pool data from Demographic and Health Surveys
  - 126,835 births to 124,719 mothers across 7,901 DHS clusters in 21 countries across 29 DHS surveys between 1990 and 2011
- Travel distance from DHS Service Availability Questionnaire (SAQ)
  - Administered at DHS cluster level
  - Group interview with 3-4 key informants in cluster
  - Informants identify nearest facility of each type from cluster
    - Hospital, health center, clinic, pharmacy, others
Countries
Distance Data – The SAQ

For each facility type:

1. How far in miles/km is the facility located from the cluster center?
2. Most common mode of transportation that is used to go to this facility?
3. How long (minutes/hours) does it take to go to the facility using the most common type of transportation?

• Following interview, facilities that were mentioned are visited by enumerator
• Advantages over using DHS GPS locations to match clusters to facilities
  • Avoids the bias induced by spatial displacement of clusters
  • Arguably more meaningful than straight-line distances
Distance Variable

- We consider reported distances to one of 4 facility types:
  - Nearest hospital
  - Nearest doctor or low-tiered clinic
  - Nearest mid-level health center
  - Nearest MCH center
- Calculate minimum distance to any of these 4 facility types
- Divide the distance variable into interval categorical variable
  - < 1 km to nearest facility, 1-2 km, 2-3 km, 3-5 km, 5-10 km, > 10 km
Distances to the Nearest Facility
Main Analysis

- Dependent variables for health care utilization:
  - Receipt of WHO-recommended 4 or more ANC visits
  - Whether or not the birth was delivered in a health facility
- Dependent variables for child mortality:
  - Child mortality (neonatal, post-neonatal infant, post-infant child)
- Main independent variable:
  - Interval categorical distance to nearest facility
- Analysis:
  - Multivariate logistic regression, reported odds ratios
Main Results: Utilization

Distance is strongly, inversely associated with service utilization

- Compared to living < 1 km from a facility, living > 10 km from a facility:
  - 38.8 percent lower odds of receiving 4 ANC visits
  - 55.3 percent lower odds of delivering in a facility

- Very similar findings when using time to facility
- Robust to alternative specifications
  - In-patient facilities only, non-migrating mothers, urban/rural, controlling for distance to other locations (school, market)
Main Results: Mortality

Distance is positively associated with child mortality (specifically in young children)

• Compared to living < 1 km from a facility, living > 10 km from a facility:
  • 17.9 percent higher odds of dying before 5th birthday
• Disaggregation suggests that the results driven by neonatal mortality
  • 26.6 percent higher odds of dying within the first 28 days

Distance not significantly associated with mortality in older age groups (post-neonatal infants and post-infant children)
Main Travel Distance Results

![Graph showing odds ratio with confidence interval for facility delivery, neonatal death, and ANC received based on distance traveled.](Image)

- **Facility Delivery**
  - Dashed line
  - Confidence intervals shown

- **Neonatal Death**
  - Solid red squares
  - Confidence intervals shown

- **ANC Received**
  - Yellow triangles
  - Confidence intervals shown

**Distance (km):**
- < 1 km
- 1 km – 1.9 km
- 2 km – 2.9 km
- 3 km – 4.9 km
- 5 km – 9.9 km
- > 10 km

**Odds Ratio with Confidence Interval:**
- The graph illustrates the relationship between the distance traveled and the odds ratio of facility delivery, neonatal death, and ANC received.
Neonatal Death by Survey

Odds Ratio with Confidence Interval

DHS Survey Country and Year
Conclusions

• People live relatively close to facilities
  • Literature is focused on the most remote areas (> 5 km or > 10 km), but such distances are rare
  • 50-60 percent of households are within 3 km
• Distance to facilities does not only matter when facilities are far, but also within relatively narrow radiiuses
  • Suggests that relatively minor factors are likely to have substantial effects on health behaviors
• Reducing distance to facilities may increase health care utilization and, more importantly, improve neonatal survival
Estimation with Induced Measurement Error in Explanatory Variables: A Numerical Integration Approach

M. Karra and D. Canning
The Measurement Error Problem

- Measurement error in an explanatory variable in a regression yields biased (attenuated) and inconsistent estimates
- Typically, structure of measurement error is unknown
- Sometimes, however, measurement error is often added to data to protect respondent confidentiality
- The structure of this induced measurement error may be known
The Measurement Error Problem

• Examples include:
  • Coarsening of the variable into bands (age, income, location)
  • Building error into the data collection (randomized response)
  • Deliberately adding noise / scrambling data (geographic locations)

• Naïve regressions with perturbed data can seriously bias results
• Previous methods to adjust for the error (e.g. regression calibration) assume normality in the variable and in the error
The Measurement Error Problem

• Want to estimate:

\[ y_i = \alpha + \beta g(x_i) + \gamma z_i + \varepsilon_i \]

• In the data, \( x_i \) not observed but we do get \( m_i \), which is \( x_i \) measured with error

• Running the regression with \( m_i \), i.e.

\[ y_i = \alpha + \beta g(m_i) + \gamma z_i + \varepsilon_i \]

will yield biased estimates of \( \beta \)
Objective

- To develop a theory that allows for unbiased and consistent estimation of a linear regression where measurement error in the explanatory variable is known
Approach

• Calculate the expected value of the true explanatory variable, given mismeasured variable and error generating process
  • Integrate over all possible actual values of the true data, weighted by conditional probability of data values given the observed perturbed data
• Replace the perturbed variable with this expectation
• This approach is related to regression calibration
  • Regression calibration is a special case where the true variable and error are independent and normally distributed
Data Requirement

- Our approach typically will require an independent source of the underlying true distribution of data, $p(x)$
  - To link individuals to exposures at the zip code level when the data reports only at the state level, we need independent information on the population distribution in each zip code
- One possible exception: if the distribution of the perturbed data can be inverted (see Appendix for technical explanation)
Applications of the Method

• Special cases include:
  • Normally distributed additive error (regression calibration)

• Applications include:
  • Coarsened location variables (state-county-zip, etc.)
  • Continuous variables in intervals (income levels, age bands)
  • Randomized responses in data (throwing a die to tell the truth)
  • Perturbed spatial data (geoscrizzling)
Application to Perturbed Spatial Data: A Simulation Exercise
Geoscrambling in the DHS

- In the Demographic and Health Surveys (DHS), GPS coordinates of surveyed household (HH) clusters are collected.
- These coordinates are then scrambled using a random angle, random radius displacement algorithm.
  - Urban HH clusters: displaced up to 2 km.
  - Rural HH clusters: displaced up to 5 km, with every 100th cluster displaced up to 10 km.
Geoscrumbling in the DHS

• A graphic example of having one facility (orange dot) and one HH cluster (blue dot)
• HH cluster is displaced by a distance at a random radius
• Calculating distance measures to this facility will be measured with error, and this error will bias estimates
Example: One Facility, One Cluster
One Facility, One Cluster

• Start with simple example of having one facility (orange dot) and one cluster (blue dot)
• Blue dot is displaced by various distances
One Facility, One Cluster

• Measurement of distance more likely to be biased upwards
• Displaced distances are more likely to be larger than original distances
Two Facilities, One Cluster
Two Facilities, One Cluster

- Extend the example of one facility-one cluster that is displaced to two facilities-one cluster.
- This implies that the cluster can potentially be mismeasured (distance is wrong) and mismatched (facility is wrong).
Simulation Setup

• Generate a 100 x 100 grid space
• Place 100 facilities uniformly across this grid at locations $r = (r_{z_1}, r_{z_2})$ for $z_1, z_2 = 1, ..., 100$
• Place 1,000 HH clusters uniformly across this grid at locations $x = (x_1, x_2)$. Cluster $i$ is denoted $x_i = (x_{i1}, x_{i2})$
• Since the placement of clusters is uniform, we know that $p(x) = p(x_1, x_2)$ is uniform
Simulation Setup

- We want to run the regression of the association between distance from the cluster to the nearest facility, $g(x_i)$ on an outcome of interest, $y_i$.
- In the equation $y_i = \alpha + \beta g(x_i) + \gamma z_i + \varepsilon_i$, the component $g(x_i)$ is the function that specifies the facility that is nearest to a household cluster, i.e.

$$g(x_i) = \min_{z_1,z_2} \sqrt{(x_{i1} - r_{z_1})^2 + (x_{i2} - r_{z_2})^2}$$

- We calculate the distance to the nearest facility $g(x_i)$ for each cluster $x_i$.
Simulation Setup

• For simulation purposes, we generate the outcome of interest $y_i$ in accordance to relationship:
  \[ y_i = 1 + 1 \cdot g(x_i) + \varepsilon_i \]
  where $\varepsilon_i \sim \mathcal{N}(0,1)$
• Here, the true parameter values are $\alpha, \beta = 1$ and $\gamma = 0$
• To validate, we can estimate this equation
  \[ y_i = \alpha x + \beta_x g(x_i) + \varepsilon_i \]
  and show that $\hat{\beta}_x$ is unbiased.
Simulation Setup

• We now assume that we are given displaced cluster coordinates \( m = (m_1, m_2) \) instead of \( (x_1, x_2) \)

• The displacement of the cluster is given by:
  • Random angle uniformly selected between \([0, 2\pi]\)
  • Random distance uniformly selected between \([0, 5]\)

• We run the regression
  \[
  y_i = \alpha_m + \beta_m g(m_i) + \epsilon_i
  \]

to show the bias in the \( \hat{\beta}_m \) estimate
Simulation Setup

\cdot Under these conditions, we know that the mechanism to induce the displacement error is:

\[ p((m_1, m_2)|(x_1, x_2)) = \begin{cases} 
0, & \sqrt{(m_1 - x_1)^2 + (m_2 - x_2)^2} > 5 \\
1 & \sqrt{(m_1 - x_1)^2 + (m_2 - x_2)^2} \leq 5 \\
5 \cdot 2\pi \sqrt{(m_1 - x_1)^2 + (m_2 - x_2)^2} & \text{if } \sqrt{(m_1 - x_1)^2 + (m_2 - x_2)^2} > 5
\end{cases} \]

\cdot We now have all of the components to do our simulation
Simulation Setup

- Run numerical integration over entire grid to get expectation
- Run the regression
  \[ y_i = \alpha_C + \beta_C E[g(x_i)|m_i] + \epsilon_i \]
- Compare estimated \( \hat{\beta}_C \) with \( \hat{\beta}_m \) and true value of \( \beta = 1 \), and show that \( \hat{\beta}_C \) is unbiased
Simulation Steps

1. Generate fixed set of 100 facilities and 1,000 clusters
2. Calculate real minimum distances for each cluster

Iterate over following 4 steps:

3. Draw random error $\varepsilon_i$ and generate outcome $y_i$
4. Run the true regression and get $\hat{\beta}_x$ estimate (unbiased)
5. Perturb each cluster $x_i$ to $m_i$, run naïve regression with $m_i$ and get $\hat{\beta}_m$ (biased)
6. Estimate expectation of the true distance by numerical integration, run adjusted regression, and get $\hat{\beta}_C$ (unbiased)

Iterate 1,000 times to get empirical distributions of $\hat{\beta}_x, \hat{\beta}_m, \hat{\beta}_C$
Simulation Results

Empirical Distributions of $\widehat{\beta}_x$, $\widehat{\beta}_m$, $\widehat{\beta}_c$ under 1,000 iterations, mesh length $h = 1$ (100 x 100 mesh)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\widehat{\beta}_x$</td>
<td>0.9997</td>
<td>0.0094</td>
<td>0.9703</td>
<td>1.0301</td>
</tr>
<tr>
<td>$\widehat{\alpha}_x$</td>
<td>1.0004</td>
<td>0.0587</td>
<td>0.8193</td>
<td>1.1965</td>
</tr>
<tr>
<td>$\widehat{\beta}_m$</td>
<td>0.8604</td>
<td>0.0151</td>
<td>0.8112</td>
<td>0.9085</td>
</tr>
<tr>
<td>$\widehat{\alpha}_m$</td>
<td>1.7238</td>
<td>0.0951</td>
<td>1.4458</td>
<td>2.0546</td>
</tr>
<tr>
<td>$\widehat{\beta}_c$</td>
<td>0.9920</td>
<td>0.0170</td>
<td>0.9427</td>
<td>1.0460</td>
</tr>
<tr>
<td>$\widehat{\alpha}_c$</td>
<td>1.0524</td>
<td>0.0945</td>
<td>0.7785</td>
<td>1.3634</td>
</tr>
<tr>
<td>$N$</td>
<td>1,000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Simulation Results

Kernel Density Plots for Beta: 1000 Repetitions, 100 x 100 Grid Size

- Black line: Distribution of beta based on true explanatory variable
- Red line: Distribution of beta based on perturbed explanatory data
- Green line: Distribution of beta based on expected value of the explanatory data

Note: The vertical purple line indicates the true parameter value of 1.
Discussion and Conclusions
Conclusions

This Study:

• Proposes a general method for consistent inference when an independent variable is deliberately measured with error
• Shows how we can use numerical integration to calculate the expected value of the true variable
• Shows an example of how the method can be used through a simulation exercise

Future Work:

• Apply this method to real datasets (e.g. DHS)
Thank You!

For additional information: mvkarra@bu.edu
Appendices
Previous Work

• Association between distance and MCH service utilization: well-established
  • Literature review by Gabrysch and Campbell (2009)
    • Found overall negative relationship between distance and utilization
  • Subsequent studies in Zambia, Bangladesh, Malawi have confirmed this inverse relationship
Previous Work

- Association between distance and child mortality remains unclear
  - Literature review by Rutherford, Mulholland, and Hill (2010)
    - Inconclusive evidence to demonstrate an association
  - Some studies found positive effects (Vietnam, Burkina Faso, Ethiopia)
  - Some studies found no effects (Malawi, Zambia, Kenya)
  - Literature review by Okwaraji and Edmond (2012)
    - Selection bias towards significant results, cannot pool results well
    - Issues around how distance is measured
Measures of Distance

- Key measure for analysis: travel distance to the nearest facility
  - Generate four distance indicators
    - Distance to the nearest hospital
    - Distance to the nearest low-tiered clinic (HC3)
    - Distance to the nearest mid-level health center (HC2)
    - Distance to the nearest MCH center or PHC (HC1)
  - Take the minimum of the four distance indicators
  - For main analysis, divide into interval categories:
    - < 1 km (ref.), 1 km – 1.9 km, 2 km – 2.9 km, 3 km – 4.9 km, 5 km – 9.9 km, > 10 km
  - Similar measure created for time to nearest facility
    - < 10 min (ref.), 10 – 19.9 min, 20 – 29.9 min, 30 – 59.9 min, > 60 min
Specification

\[
\ln \left( \frac{Pr[Y_{ihc} = 1|X_{ih}, Z_C, \zeta_j]}{1 - Pr[Y_{ihc} = 1|X_{ih}, Z_C, \zeta_j]} \right) = \beta_0 + \beta_D D_c + X_{ih} \gamma + Z_C \delta + \zeta_j + \varepsilon_{ihc}
\]

- \(Y_{ih}\) is the binary dependent variable for birth \(i\) in household \(h\) in cluster \(c\) in survey \(j\)
- \(D_c\) is the travel distance to nearest facility variable for cluster \(c\)
- \(X_{ih}\) is the vector of individual-level and HH-level controls
- \(Z_C\) is the vector of cluster-level controls
- \(\zeta_j\) are survey-level fixed effects

- Regression standard errors are clustered at the DHS cluster level
## DHS Countries, Years

<table>
<thead>
<tr>
<th>Country</th>
<th>Year</th>
<th>Country</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bangladesh</td>
<td>2004</td>
<td>Haiti</td>
<td>1994-95</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>2007</td>
<td>Haiti</td>
<td>2000</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>2011</td>
<td>Jordan</td>
<td>1990</td>
</tr>
<tr>
<td>Benin</td>
<td>1996</td>
<td>Kenya</td>
<td>1993</td>
</tr>
<tr>
<td>Benin</td>
<td>2001</td>
<td>Malawi</td>
<td>1992</td>
</tr>
<tr>
<td>Benin</td>
<td>2006</td>
<td>Mali</td>
<td>1995-96</td>
</tr>
<tr>
<td>Bolivia</td>
<td>1994</td>
<td>Mali</td>
<td>2001</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>1993</td>
<td>Morocco</td>
<td>1992</td>
</tr>
<tr>
<td>Cameroon</td>
<td>1991</td>
<td>Niger</td>
<td>1998</td>
</tr>
<tr>
<td>CAR</td>
<td>1994-95</td>
<td>Nigeria</td>
<td>1990</td>
</tr>
<tr>
<td>Chad</td>
<td>1996-97</td>
<td>Uganda</td>
<td>1995</td>
</tr>
<tr>
<td>Chad</td>
<td>2004</td>
<td>Vietnam</td>
<td>1997</td>
</tr>
<tr>
<td>Cote d’Ivoire</td>
<td>1994</td>
<td>Vietnam</td>
<td>2002</td>
</tr>
<tr>
<td>Gabon</td>
<td>2000</td>
<td>Zimbabwe</td>
<td>1994</td>
</tr>
<tr>
<td>Guinea</td>
<td>1999</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Control Variables

- Birth- and HH-level controls:
  - Birth order, mother’s education (categorical), HH wealth (quintiles), age of mother (categorical), place of residence (urban/rural)
  - For mortality regressions, hypothetical age of the child and the age of the child squared are added

- Cluster-level controls
  - Average wealth (quintiles), average schooling for mothers
## Descriptive Statistics: Distances

<table>
<thead>
<tr>
<th>Minimum Travel Distance, categorical</th>
<th>Mean</th>
<th>No.</th>
<th>Urban Mean</th>
<th>Rural Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum distance to facility, &lt; 1 km</td>
<td>0.279</td>
<td>35,387</td>
<td>0.534</td>
<td>0.177</td>
</tr>
<tr>
<td>Minimum distance to facility, 1 – 1.9 km</td>
<td>0.091</td>
<td>11,542</td>
<td>0.160</td>
<td>0.064</td>
</tr>
<tr>
<td>Minimum distance to facility, 2 – 2.9 km</td>
<td>0.152</td>
<td>19,279</td>
<td>0.158</td>
<td>0.150</td>
</tr>
<tr>
<td>Minimum distance to facility, 3 – 4.9 km</td>
<td>0.121</td>
<td>15,347</td>
<td>0.066</td>
<td>0.143</td>
</tr>
<tr>
<td>Minimum distance to facility, 5 – 9.9 km</td>
<td>0.153</td>
<td>19,406</td>
<td>0.050</td>
<td>0.194</td>
</tr>
<tr>
<td>Minimum distance to facility, &gt; 10 km</td>
<td>0.204</td>
<td>25,874</td>
<td>0.031</td>
<td>0.272</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td></td>
<td>126,835</td>
<td>42,746</td>
<td>84,089</td>
</tr>
</tbody>
</table>
### Descriptive Statistics: Outcomes

<table>
<thead>
<tr>
<th>Outcome Variables</th>
<th>Mean</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHO Recommended ANC Visits (1 = yes)</td>
<td>0.394</td>
<td>49,186</td>
</tr>
<tr>
<td>Delivery in a health facility (1 = yes)</td>
<td>0.426</td>
<td>53,152</td>
</tr>
<tr>
<td>Child death</td>
<td>0.082</td>
<td>10,427</td>
</tr>
<tr>
<td>Neonatal death</td>
<td>0.030</td>
<td>3,806</td>
</tr>
<tr>
<td>Post-neonatal infant death</td>
<td>0.034</td>
<td>4,427</td>
</tr>
<tr>
<td>Post-infant child death</td>
<td>0.017</td>
<td>2,189</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td><strong>126,835</strong></td>
<td></td>
</tr>
</tbody>
</table>
## Descriptive Statistics: Distances

<table>
<thead>
<tr>
<th>CLUSTERS</th>
<th>Minimum Travel Distance, categorical</th>
<th>Mean</th>
<th>No.</th>
<th>Urban Mean</th>
<th>Rural Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimum distance to facility, &lt; 1 km</td>
<td>0.318</td>
<td>2,514</td>
<td>0.538</td>
<td>0.186</td>
</tr>
<tr>
<td></td>
<td>Minimum distance to facility, 1 – 1.9 km</td>
<td>0.111</td>
<td>869</td>
<td>0.169</td>
<td>0.074</td>
</tr>
<tr>
<td></td>
<td>Minimum distance to facility, 2 – 2.9 km</td>
<td>0.170</td>
<td>1,340</td>
<td>0.160</td>
<td>0.175</td>
</tr>
<tr>
<td></td>
<td>Minimum distance to facility, 3 – 4.9 km</td>
<td>0.116</td>
<td>915</td>
<td>0.058</td>
<td>0.150</td>
</tr>
<tr>
<td></td>
<td>Minimum distance to facility, 5 – 9.9 km</td>
<td>0.133</td>
<td>1,052</td>
<td>0.048</td>
<td>0.185</td>
</tr>
<tr>
<td></td>
<td>Minimum distance to facility, &gt; 10 km</td>
<td>0.153</td>
<td>1,211</td>
<td>0.027</td>
<td>0.229</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>7,901</td>
<td>3,346</td>
<td>4,555</td>
<td></td>
</tr>
</tbody>
</table>
## Descriptive Statistics: Covariates

<table>
<thead>
<tr>
<th>Mother-Level Covariates</th>
<th>Mean</th>
<th>SD</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wealth, quintiles</td>
<td>2.893</td>
<td>1.392</td>
<td></td>
</tr>
<tr>
<td>Education, none (1 = yes)</td>
<td>0.532</td>
<td></td>
<td>66,323</td>
</tr>
<tr>
<td>Education, primary (1 = yes)</td>
<td>0.271</td>
<td></td>
<td>33,777</td>
</tr>
<tr>
<td>Education, secondary (1 = yes)</td>
<td>0.176</td>
<td></td>
<td>21,890</td>
</tr>
<tr>
<td>Education, higher (1 = yes)</td>
<td>0.022</td>
<td></td>
<td>2,727</td>
</tr>
<tr>
<td>Maternal age, years</td>
<td>28.214</td>
<td>7.041</td>
<td></td>
</tr>
<tr>
<td>Marital status (1 = married)</td>
<td>0.865</td>
<td></td>
<td>107,875</td>
</tr>
<tr>
<td>Urban (1 = yes)</td>
<td>0.284</td>
<td></td>
<td>35,399</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster-Level Covariates</th>
<th>Mean</th>
<th>SD</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Average wealth, quintiles</td>
<td>2.889</td>
<td>1.066</td>
<td></td>
</tr>
<tr>
<td>Average education, highest level</td>
<td>0.682</td>
<td>0.616</td>
<td></td>
</tr>
<tr>
<td>Distance to primary school, km</td>
<td>1.724</td>
<td>4.822</td>
<td></td>
</tr>
</tbody>
</table>

\[ N \quad 124,719 \]
### Descriptive Statistics: Covariates

<table>
<thead>
<tr>
<th>Birth-Level Covariates</th>
<th>Mean</th>
<th>SD</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth order</td>
<td>3.876</td>
<td>2.651</td>
<td></td>
</tr>
<tr>
<td>Multiple birth (1 = yes)</td>
<td>0.027</td>
<td></td>
<td>3,383</td>
</tr>
<tr>
<td>Child sex (1 = female)</td>
<td>0.494</td>
<td></td>
<td>62,705</td>
</tr>
<tr>
<td>Time from birth to survey date, months</td>
<td>24.311</td>
<td>16.115</td>
<td></td>
</tr>
<tr>
<td><strong>N</strong></td>
<td></td>
<td></td>
<td><strong>126,835</strong></td>
</tr>
</tbody>
</table>
## Main Travel Distance Results

<table>
<thead>
<tr>
<th>Reference: &lt; 1 km</th>
<th>(1) Neonatal</th>
<th>(2) ANC Visits</th>
<th>(3) Delivery</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 km – 1.9 km</td>
<td>1.077</td>
<td>0.834***</td>
<td>0.920</td>
</tr>
<tr>
<td></td>
<td>(0.927 - 1.251)</td>
<td>(0.769 - 0.904)</td>
<td>(0.828 - 1.023)</td>
</tr>
<tr>
<td>2 km – 2.9 km</td>
<td>1.163**</td>
<td>0.825***</td>
<td>0.754***</td>
</tr>
<tr>
<td></td>
<td>(1.020 - 1.327)</td>
<td>(0.767 - 0.887)</td>
<td>(0.681 - 0.835)</td>
</tr>
<tr>
<td>3 km – 4.9 km</td>
<td>1.250***</td>
<td>0.779***</td>
<td>0.691***</td>
</tr>
<tr>
<td></td>
<td>(1.087 - 1.439)</td>
<td>(0.715 - 0.850)</td>
<td>(0.612 - 0.779)</td>
</tr>
<tr>
<td>5 km – 9.9 km</td>
<td>1.191**</td>
<td>0.713***</td>
<td>0.547***</td>
</tr>
<tr>
<td></td>
<td>(1.042 - 1.363)</td>
<td>(0.652 - 0.779)</td>
<td>(0.483 - 0.620)</td>
</tr>
<tr>
<td>&gt; 10 km</td>
<td>1.266***</td>
<td>0.612***</td>
<td>0.447***</td>
</tr>
<tr>
<td></td>
<td>(1.108 - 1.445)</td>
<td>(0.559 - 0.671)</td>
<td>(0.394 - 0.508)</td>
</tr>
</tbody>
</table>

| N                | 125,167      | 124,719        | 124,719      |

*** p < 0.01, ** p < 0.05, * p < 0.1
## Main Travel Time Results

<table>
<thead>
<tr>
<th>Reference: &lt; 10 min</th>
<th>(1) Neonatal</th>
<th>(2) ANC Visits</th>
<th>(3) Delivery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time: 10 min – 19.9 min</td>
<td>1.074</td>
<td>0.872***</td>
<td>0.794***</td>
</tr>
<tr>
<td>(0.952 - 1.212)</td>
<td>(0.814 - 0.933)</td>
<td>(0.722 - 0.873)</td>
<td></td>
</tr>
<tr>
<td>Time: 20 min – 29.9 min</td>
<td>1.157**</td>
<td>0.807***</td>
<td>0.732***</td>
</tr>
<tr>
<td>(1.015 - 1.319)</td>
<td>(0.745 - 0.874)</td>
<td>(0.659 - 0.814)</td>
<td></td>
</tr>
<tr>
<td>Time: 30 min – 59.9 min</td>
<td>1.223***</td>
<td>0.748***</td>
<td>0.602***</td>
</tr>
<tr>
<td>(1.078 - 1.389)</td>
<td>(0.692 - 0.809)</td>
<td>(0.538 - 0.674)</td>
<td></td>
</tr>
<tr>
<td>Time: &gt; 60 min</td>
<td>1.256***</td>
<td>0.688***</td>
<td>0.477***</td>
</tr>
<tr>
<td>(1.105 - 1.429)</td>
<td>(0.627 - 0.753)</td>
<td>(0.419 - 0.543)</td>
<td></td>
</tr>
</tbody>
</table>

| N | 125,167 | 124,719 | 124,719 |

*** p < 0.01, ** p < 0.05, * p < 0.1
# Check: In-Patient Facilities Only

<table>
<thead>
<tr>
<th></th>
<th>(1) ANC</th>
<th>(2) Delivery</th>
<th>(3) Neonatal</th>
<th>(4) Post-Neonatal</th>
<th>(5) Child 1-5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reference:</strong></td>
<td><strong>&lt; 1 km</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 km - 1.9 km</td>
<td>0.825***</td>
<td>0.904*</td>
<td>1.044</td>
<td>1.034</td>
<td>1.049</td>
</tr>
<tr>
<td></td>
<td>(0.760 - 0.896)</td>
<td>(0.808 - 1.012)</td>
<td>(0.896 - 1.217)</td>
<td>(0.879 - 1.218)</td>
<td>(0.860 - 1.279)</td>
</tr>
<tr>
<td>2 km - 2.9 km</td>
<td>0.801***</td>
<td>0.711***</td>
<td>1.211***</td>
<td>1.113</td>
<td>1.094</td>
</tr>
<tr>
<td></td>
<td>(0.742 - 0.865)</td>
<td>(0.638 - 0.793)</td>
<td>(1.054 - 1.392)</td>
<td>(0.964 - 1.285)</td>
<td>(0.913 - 1.310)</td>
</tr>
<tr>
<td>3 km - 4.9 km</td>
<td>0.736***</td>
<td>0.619***</td>
<td>1.314***</td>
<td>1.048</td>
<td>1.193*</td>
</tr>
<tr>
<td></td>
<td>(0.673 - 0.805)</td>
<td>(0.546 - 0.701)</td>
<td>(1.134 - 1.523)</td>
<td>(0.901 - 1.220)</td>
<td>(0.988 - 1.441)</td>
</tr>
<tr>
<td>5 km - 9.9 km</td>
<td>0.699***</td>
<td>0.543***</td>
<td>1.175**</td>
<td>0.931</td>
<td>1.013</td>
</tr>
<tr>
<td></td>
<td>(0.640 - 0.763)</td>
<td>(0.479 - 0.616)</td>
<td>(1.022 - 1.351)</td>
<td>(0.809 - 1.072)</td>
<td>(0.847 - 1.212)</td>
</tr>
<tr>
<td>&gt; 10 km</td>
<td>0.587***</td>
<td>0.435***</td>
<td>1.295***</td>
<td>1.108</td>
<td>1.108</td>
</tr>
<tr>
<td></td>
<td>(0.538 - 0.640)</td>
<td>(0.385 - 0.492)</td>
<td>(1.132 - 1.481)</td>
<td>(0.972 - 1.262)</td>
<td>(0.941 - 1.305)</td>
</tr>
</tbody>
</table>

| **N** | 124,719 | 124,719 | 125,167 | 87,289 | 83,176 |

*** p < 0.01, ** p < 0.05, * p < 0.1
## Check: Control School Distance

<table>
<thead>
<tr>
<th></th>
<th>(1) ANC</th>
<th>(2) Delivery</th>
<th>(3) Neonatal</th>
<th>(4) Post-Neonatal</th>
<th>(5) Child 1-5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reference:</strong></td>
<td>&lt; 1 km</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 km – 1.9 km</td>
<td>0.855***</td>
<td>0.856***</td>
<td>1.021</td>
<td>1.058</td>
<td>1.010</td>
</tr>
<tr>
<td></td>
<td>(0.782 - 0.935)</td>
<td>(0.762 - 0.961)</td>
<td>(0.866 - 1.203)</td>
<td>(0.881 - 1.271)</td>
<td>(0.811 - 1.260)</td>
</tr>
<tr>
<td>2 km – 2.9 km</td>
<td>0.845***</td>
<td>0.707***</td>
<td>1.163**</td>
<td>1.079</td>
<td>1.150</td>
</tr>
<tr>
<td></td>
<td>(0.776 - 0.920)</td>
<td>(0.630 - 0.794)</td>
<td>(1.000 - 1.353)</td>
<td>(0.911 - 1.278)</td>
<td>(0.938 - 1.409)</td>
</tr>
<tr>
<td>3 km – 4.9 km</td>
<td>0.774***</td>
<td>0.603***</td>
<td>1.273***</td>
<td>1.043</td>
<td>1.191</td>
</tr>
<tr>
<td></td>
<td>(0.694 - 0.864)</td>
<td>(0.521 - 0.698)</td>
<td>(1.079 - 1.501)</td>
<td>(0.874 - 1.243)</td>
<td>(0.953 - 1.489)</td>
</tr>
<tr>
<td>5 km – 9.9 km</td>
<td>0.739***</td>
<td>0.529***</td>
<td>1.200**</td>
<td>0.993</td>
<td>1.034</td>
</tr>
<tr>
<td></td>
<td>(0.661 - 0.826)</td>
<td>(0.456 - 0.614)</td>
<td>(1.029 - 1.399)</td>
<td>(0.846 - 1.166)</td>
<td>(0.844 - 1.266)</td>
</tr>
<tr>
<td>&gt; 10 km</td>
<td>0.571***</td>
<td>0.416***</td>
<td>1.240***</td>
<td>1.091</td>
<td>1.108</td>
</tr>
<tr>
<td></td>
<td>(0.506 - 0.644)</td>
<td>(0.356 - 0.485)</td>
<td>(1.062 - 1.447)</td>
<td>(0.942 - 1.265)</td>
<td>(0.914 - 1.343)</td>
</tr>
</tbody>
</table>

**N**  
95,108  
95,108  
95,300  
66,071  
62,972

*** p < 0.01, ** p < 0.05, * p < 0.1
Main Travel Time Results

<table>
<thead>
<tr>
<th>Time (min)</th>
<th>Facility Delivery</th>
<th>Neonatal Death</th>
<th>ANC Received</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 10 min</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 min - 19.9 min</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 min - 29.9 min</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 min - 59.9 min</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 60 min</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Odds Ratio with Confidence Interval
Interpretation of Results

- Stronger association for in-facility delivery than for ANC coverage
  - Women can better plan ANC visits compared to when going to deliver
  - ANC is repeated, but delivery is one-shot
- Reasons for null, insignificant findings in older children
  - Seeking neonatal care not as easily anticipated as seeking care for older child, who is less susceptible
- Composition effects – which type of women use facilities?
  - Women who plan ahead vs. women who do not plan
  - But we see no differences for non-migrating mothers
- No qualitative differences between spatial and temporal distance
Approach

• Calculate the expected value of the true explanatory variable:

\[
E[g(x_i) | m_i] = \int_x g(x)p(x|m_i)dx
\]

• Set \( g(x_i) = E[g(x_i) | m_i] + u_i \), where \( u_i \) is an error term with mean 0 and is independent of \( x_i \) and \( z_i \)

• Rewrite the estimating equation as:

\[
y_i = \alpha + \beta E[g(x_i) | m_i] + \gamma z_i + \nu_i
\]

where \( \nu_i = \beta u_i + \varepsilon_i \)

• This yields unbiased estimates of \( \alpha, \beta, \gamma \)
Calculating $E[g(x_i)|m_i]$

- Calculate the expected value of the true explanatory variable using Bayes’ Rule:

$$E[g(x_i)|m_i] = \int_x g(x)p(x|m_i)dx$$

$$= \int_x g(x) \frac{p(m_i|x)p(x)}{\int_x p(m_i|x)p(x)dx} dx$$

where $p(m_i|x)$ is the PDF of the error generation process and $p(x)$ is the PDF of the true values of the data, $x$
Calculating $E[g(x_i) | m_i]$

- In some cases, the integration needed to calculate the expectation is straightforward.
- In some cases, there may not be an analytic solution.
- Use numerical integration methods (sum over grid with interval $s = 0, \ldots, S$ and mesh $h$) to approximate the expectation.

$$\sum_{s=0}^{S-1} g(x_s) \frac{p(m_i | x_s)p(x_s)h}{\sum_{s=0}^{S-1} p(m_i | x_s)p(x_s)h} \approx \int_X g(x) \frac{p(m_i | x)p(x)}{\int_X p(m_i | x)p(x)dx} \, dx$$
A Possible Exception: Inversion

• Since we know the form of the measurement error, it may be possible to invert the distribution of perturbed data to generate the underlying distribution of the true data
  • Distributions of the true and perturbed variables are linked by a non-homogenous Fredholm integral equation of the first kind
  • Solution of this equation is well-studied
• But the inverse problem is generally not well posed
  • Cannot guarantee the existence or uniqueness of a solution
  • So then we require data on the underlying distribution