CONDUCTING FIELD RESEARCH IN DEVELOPING COUNTRIES

Michael Kremer, Harvard University

Introduction

- Today's lecture focuses on methods
- Practical advice for conducting fieldwork in developing countries
- Two parts:
  - Part 1: general advice for fieldwork
  - Part 2: advice for Randomized Control Trials (RCTs)

Part 1

FIELD RESEARCH IN DEVELOPING COUNTRIES

Outline

1. **Whether to conduct fieldwork in development**
2. How to do field research
3. Data collection and management
4. Survey design
5. Consumption
6. Profits
7. Attrition
8. Mere measurement effects
9. Budgets and funding
10. Human subjects and permissions

Where to do field research?

- Always easier to do in a place that you already know
- Multiple projects in one country cuts down on fixed costs
- Collaborators, survey infrastructure
- Rural or urban areas
  - Potentially large transport costs in rural areas
  - Political or safety issues, higher salaries in urban areas

Why not collect field data

- Many data sets in developing countries are available in electronic form
  - Household: DHS and LSMS surveys, IFLS (Indonesia), REDS (India), MxFLS (Mexico)
  - Firm: World Bank Business Environment Studies
- If these are sufficient, use them
  - Collecting own data is time-consuming, risky

Why not collect field data

- Time-consuming, risky, drawn-out process
  - High turnover of civil servants → may lose your advocate in an office/ministry → may lose permission to do research
  - Survey firm hold up
  - Poor data quality and need to return to field
  - Natural/human disasters
  - Lots of management
- BUT: Experience in developing countries is useful for:
  - Asking the right questions / generating interesting hypotheses
  - Assessing data issues

Outline

1. Whether to conduct fieldwork in development
2. **How to do field research**
3. Data collection and management
4. Survey design
5. Consumption
6. Profits
7. Attrition
8. Mere measurement effects
9. Budgets and funding
10. Human subjects and permissions

How to do field research?

- Many different methods of collecting data
  - Qualitative
    - Useful for defining hypothesis
  - Existing administrative records
    - Cheap, high coverage
    - Sometimes distorted by official incentives/lack of incentives
  - Surveys
    - Moderate cost
    - Ask exactly the questions you need
    - Cheap talk—easy to distort true preferences
  - Focus group discussions
    - Cheaper and faster than surveys

Methods of collecting data, II

- Coupons (e.g. fertilizer, outcome how many redeemed)
  - Cheap
  - Actual behavior
- Games
  - Correlated with actual outcomes (Karlan…)
  - Expensive and hard to do right
  - Be careful about what they are measuring
  - Are people trying to give you the right answer?
- Participatory exercises
  - Expensive
  - Test only a few outcomes with rich coverage
  - Costly for people to hide true preferences

Methods of collecting data, III

- Direct Observation e.g. village meeting on allocation of resources (do women talk?)
  - Need same event to occur everywhere
  - Must happen at a predictable time
  - Can't control what happens in front of you
  - Your presence might affect the community meeting (but surveying someone might too)
- Namelists
  - Checking attendance of students
  - Faster and more accurate than school records
- Peer Data
  - Asking people about behavior of others, instead of self-reported

Methods of collecting data, IV

- Participatory decision making
  - Offer choice of thank you gifts to community
  - Observe how they choose
- Anthropometrics
  - Health outcome data
  - Expensive equipment
  - Training is crucial
- Satellite
  - High coverage for reasonable cost
- GPS
  - Useful for measuring externalities

Survey implementation

- Pre-fieldwork
  - Form core team
    - Researchers and policy analysts
    - Policy makers and program managers
    - Data collectors and producers (interviewers, supervisors, data entry staff)
  - Draft budget, timeline, and secure financing
  - Human Subjects
  - Design and test instruments
  - Draw sample
  - Write training manuals for field staff (interviewers, supervisors, data entry staff) and train
  - Prepare fieldwork and data entry plan, including tracking protocol for non-respondents
  - Conduct pilot
- In the field
  - Entering checklist data in “real-time” to assess progress
  - Re-survey households in sample as a check
  - Adapt strategy as needed
- Post-fieldwork
  - Debrief and collect feedback from survey team

Outline

1. Whether to conduct fieldwork in development
2. How to do field research
3. **Data collection and management**
4. Survey design
5. Consumption
6. Profits
7. Attrition
8. Mere measurement effects
9. Budgets and funding
10. Human subjects and permissions

Data collection

- Contracting out vs. employing enumerators directly?
  - Survey firms often do low quality work
    - Need lots of supervision
  - Separate data collection/ data entry
  - Need to pre-specify acceptable error rate in contract

- If sampling a larger target population and want a representative sample, you will need a household-level census, either:
  - Use publicly available data
  - Conduct your own census, can organize community meeting to map households and characteristics of household members (e.g. age, gender)

- Are several stages of data collection necessary?
  - Possibly. if you need to measure the progression of effects over time, or if an intervention is being phased-in slowly
  - Good to get GPS locations of households, nicknames of respondents.
  - Questions that shouldn't change over time, e.g. gender

Data collection

- **Data entry format** must be clear and should not leave room for interpretation by the enumerator: Write out complete question, DK code (-999), skip pattern (-777)

- **Training** enumerators in these procedures is essential. Create **manuals** for all survey instruments.

- Names and dates of enumerators on forms and dates of survey
- Need daily or weekly **check** of *all* the forms by a supervisor, and a re-check on a random basis by the research manager.
- **Re-survey** sample of respondents on a random basis.
  - Special emphasis on variables that shouldn't change

- If more than one survey per household, use household code
  - Give household code relevant information: 772032 = 77 is region code, 20 is district code, 32 is household number

Data collection

- Can do some interesting randomizations during data collection (especially the pilot)
  - Order of questions
  - Characteristics of enumerators
  - Frequency of data collection
  - Form and value of compensation for respondents (if applicable)

- Check if these systematically affect responses
  - If they do, potential problems with measurement
  - Might not be measuring parameters you're intending to measure

Field team and supervision protocol

Grosh and Glewwe, 2005

- Field team includes interviewers and supervisors
- A supervisor role:
  - Coordinate interviewers
  - Edit and quality checks of surveys
  - Conduct unannounced visits to interviewers
  - Audits by randomly selecting and re-visiting interviewed households
  - Each supervisor should be responsible for a small number of interviewers
  - Update "checklist" to track completion rate
- Ideal team size under one supervisor depends on area of survey, length of questionnaire(8 to 10 typical)

Field team and supervision protocol

Grosh and Glewwe, 2005

- Payment structure for team and ensuring high data quality
  - e.g. pay per questionnaire vs. pay daily salary vs. pay for project?
  - Financial motivation techniques are risky to use, since work effort and quality are often impossible to monitor all the time
    - Bonus per survey: surveyors will want to rush through or falsify surveys
    - Bonus for low error rates: puts a lot of pressure on the field editors to overlook errors, and can create collusion
  - Payment for transport, per diem or costs?

Outline

1. Whether to conduct fieldwork in development
2. How to do field research
3. Data collection and management
4. **Survey design**
5. Consumption
6. Profits
7. Attrition
8. Mere measurement effects
9. Budgets and funding
10. Human subjects and permissions

Formulate survey objectives

- Define hypotheses to test first
- Which variables or combination of variables can test these key hypotheses?
  - Good to check existing surveys for ideas
  - Intermediate and final outcomes
  - Variables to discriminate between various models
- Seek advice of experts
  - Medicine and public health
  - Education

Consider constraints

- Financial resources – tradeoff between sample size and amount of information obtained from each household
- Human resource capacity of organization implementing the survey – research coordinators, interviewers
- Willingness and ability of respondents to provide desired information
  - Both length and content of survey can affect quality of response

Willingness and ability of respondents to provide desired information

- Willingness: Use objective measures if treatment or act of measurement may influence measurement
  - HIV/AIDS education → artificially high self-reports of safe sex → can look at pregnancy instead
  - Asking about income → people might answer in a way that increases chances of getting aid
- Ability: Perceptions vs. reality
  - Corruption in Indonesia (Olken 2006)
  - Get as close to reality as possible

Questionnaire design

- Use publicly available instruments
  - Living Standards Measurement Surveys
  - Demographic and Health Surveys
  - World Values Survey
- Use local knowledge
  - Qualitative research can generate local validity
    - Key informant interviews
    - Focus groups
    - Participatory exercises
  - E.g., asking about polygamy or dowry may be appropriate in one setting but absurd, offensive, and a waste of resources in another

Questionnaire design: Basic structure

- Organize by module
  - Typical modules: Metadata, household roster, consumption and prices, education, health, anthropometry, employment, transfers and other non-labor income, housing, community characteristics, fertility, migration, household enterprises, agriculture, savings, credit, time use
- Start questionnaire with straightforward modules
- Sensitive modules should come at the end, when rapport with respondent has been built
  - Both the obvious questions like sexual health, but also household income, consumption, etc.
- Almost all answers to almost all questions should be pre-coded
  - Reduces ambiguity
  - Bypasses coding step
  - Coding scheme should be consistent across questions (e.g., 1=Yes, 2=No)

Questionnaire design: General issues

- Poor people are more willing (or even happy) to be interviewed, but be mindful of length
  - Large-scale, multi-topic household surveys can take 5-6 hours per interview (Deaton, 1997)
    - Compare to developed countries
  - Increases not only cost but also likelihood of non-sampling errors
- Write out questions so that interviewer can read verbatim
  - Instead of: “Main occupation?”
  - Write: “During the past seven days, what kind of work did you do? If you had more than one kind of work, tell me the one for which you worked the most hours during the past seven days.”
- Keep questions simple and singular
  - Instead of: "During the past seven days, were you employed for wages or other remuneration, were you self-employed in a household enterprise, were you engaged in both types of activities simultaneously, or were you engaged in neither activity?"
  - Simplify language and break into two questions:
    - "During the past seven days, did you work for pay for someone who is not a member of this household?"
    - "During the past seven days, did you work on your own account, for example, as a farmer or a seller of goods or services?“

Questionnaire design: General issues

- Specify preferred respondent for given question/module
  - Child's mother or primary caretaker if asking about child health
  - Household head if asking about expenditure or assets
- Include precise definitions of all key concepts used in questionnaire ("household," "meal," "relative")
  - Make sure direction of the relationship is clear in the survey
- Include explicit instructions for enumerator within questionnaire
  - Use skip codes
  - Observe typographic conventions (e.g., instructions for enumerators in uppercase letters)
- Minimize calculations necessary
- Consider recall periods
- Use questions that have more variation (e.g. very likely, likely, not very likely, etc.) rather than just yes/no questions – might help with analysis later
- Asking iterations of the same question throughout the questionnaire could help you check whether there is variation in the response

Questionnaire design: General issues

- Translate into all major languages and back-translate to verify
- Ensure good survey format
- Field-test the instrument – an iterative process!
  - **Questionnaire as a whole:** Internally consistent? Full range of required information collected? Any double-counted variables? Appropriate length?
  - **Individual modules:** Any questions missing? Redundant or irrelevant questions?
  - **Individual questions:** Clear wording? Do any questions allow for ambiguous responses? Have all responses been anticipated and coded?
- Numerous options for electronic survey software / hardware. Tradeoff between cost and survey complexity

Importance of pilots

- **Pilots vary in size & rigor**
- Pilots & qualitative steps are important
  - Good to pilot *entire* process
  - Test to see if planned intervention will function
  - Test experimental protocols and monitoring
  - Test the cost-feasibility
- Something always goes wrong
  - Respondents might not understand a question
  - Enumerators might ignore a question
  - Inappropriate ranges offered for variables so that everyone falls into one category
- Better to find this out before study begins
- Often discover other interesting questions in process

Outline

1. Whether to conduct fieldwork in development
2. How to do field research
3. Data collection and management
4. Survey design
5. **Consumption**
6. **Profits**
7. **Attrition**
8. **Mere measurement effects**
9. Budgets and funding
10. Human subjects and permissions

Measuring consumption

- Income is standard measure of welfare in developed countries
  - Easy to measure: for a given respondent, typically comes from one or two sources (e.g., wages and pensions)
  - US Consumer Expenditure Survey costs 5X more per household than Current Population Survey (Deaton and Grosh 2000)
- In developing countries, consumption typically used instead
  - Consumption is smoother – less subject to shocks and seasonal factors (monthly paychecks, post-harvest income)
  - Respondents less reluctant to share information (e.g., may have concerns about income taxation)
  - More visible to all household members and thus less respondent-dependent
  - In informal sector, personal and business accounts difficult to disentangle

Measuring consumption

- Data required
  - Consumption
  - Prices
  - Household size
- Consumption should include:
  - All reported expenditures on individual goods and services
  - Non-market consumption of goods/services:
    - Produced at home (e.g., food)
    - Received in-kind from employers
  - From stock (hence even for market goods, consumption ≠ expenditure)

Measuring consumption

- Sources of price data
  - Price indexes are often based on urban data – useful only if urban respondent or if spatial price variation limited
  - Collect local market data through community-level questionnaire
  - Collect from household
    - Unit value ≠ price (doesn't take into account quality)
    - Quantification sometimes difficult – one "bunch" of vegetables

Sources of bias in measuring consumption

- Unit non-response
  - By virtually any measure of socioeconomic status, better-off households are less likely to participate (Thomas, Frankenberg, and Smith, JHR 2001; Deaton, 2005)
- Unit non-coverage
  - Students, military, institutionalized persons often excluded
- Item non-response
  - Refusal to respond (sensitivity)
  - Inability to respond (may be correlated with literacy/numeracy)
- Social desirability bias
  - Illegal or "immoral" activities (e.g., alcohol use) may be underreported
  - Respondents may believe that "right" answer or demonstrating more need will yield more aid
- Recall bias

Recall bias

- Tradeoff:
  - Accuracy of memory calls for shorter period
  - Minimization of variance calls for longer period
    - Short periods overestimate dispersion (inequality, poverty)
- In order to analyze across modules, must have consistent periods
- Optimal recall period depends on whether interested in averages or distribution

Recall bias

- Example: India LSMS comparing 30- and 7-day recall periods
  - 7-day recall group reported 30% higher food expenditure and 17% higher total consumption
  - "enough to reduce the measured headcount ratio by half, removing almost 200 million people from poverty" (Deaton 2005)

Measuring profits

De Mel, McKenzie, and Woodruff (2009)

- Challenges in getting accurate data on firms
  - Few or no records (90% of firms do not pay taxes)
  - Money fungible between household and business
  - Lags between expenses and related revenue (e.g., buying inventory)
  - Owners reluctant to share information
  - Valuing family time
  - Recall bias

Measuring profits

De Mel, McKenzie, and Woodruff (2009)

- Bookkeeping experiments on Sri Lankan microenterprises
  - Randomly assigned firms to keep ledger books
  - Assigned research assistants to observe sample of firms 15 times per month
  - Recorded all transactions and estimated actual revenues
  - Result: firms' reported revenues 31% lower than estimates
- Low correlation between reported profits and difference between reported revenues and expenses
  - Pearson correlations ranged from .043 to .468
  - Reported profits larger and less noisy
- "Simply asking profits provides a more accurate measure of firm profits than detailed questions on revenues and expenses."

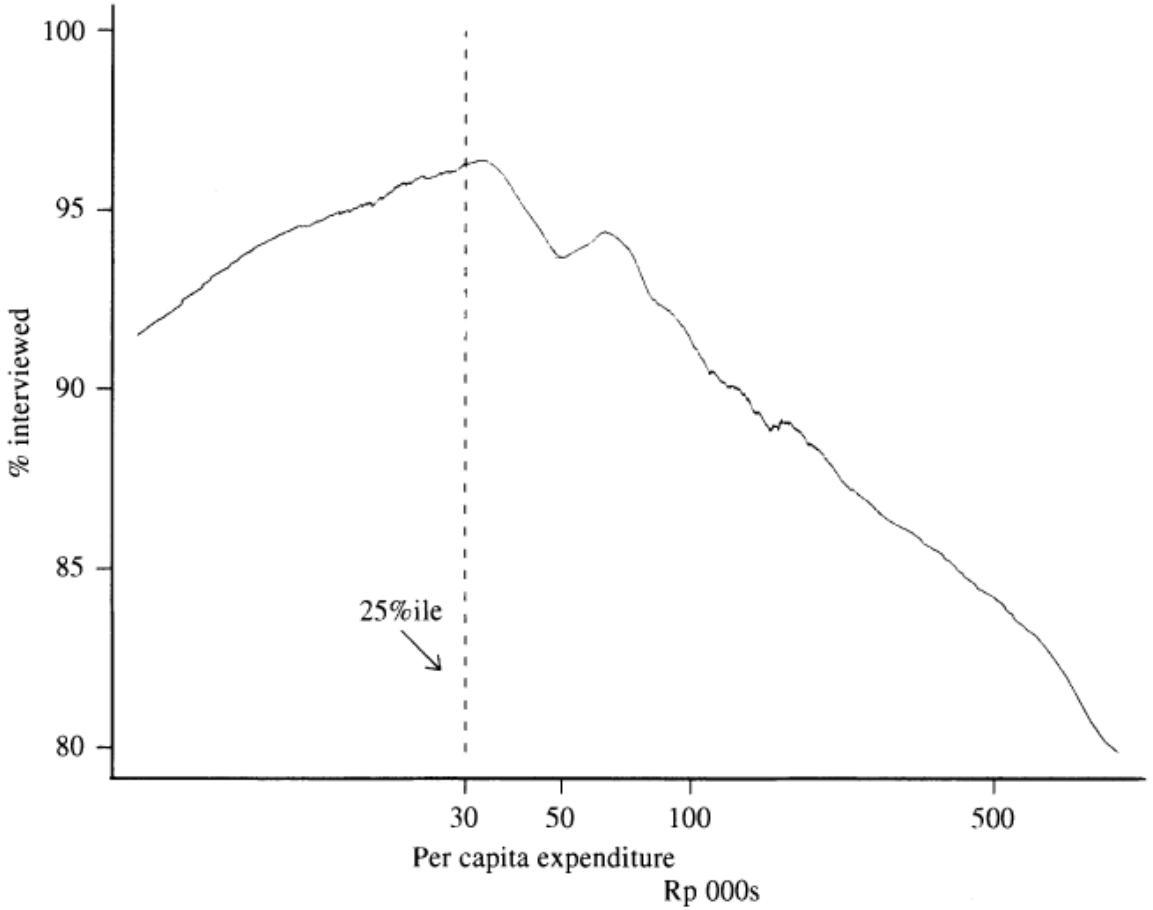
Managing attrition

Thomas, Frankenberg, and Smith (2001)

- Primary cause
  - In developed countries: respondent refusal
  - In developing countries: inability to physically find respondent
- Tracking is difficult
  - Household splitting/consolidation
  - Migration or displacement
  - Harsh geography
  - Poor codifying of addresses (e.g. “next to the large peepal tree”)
  - Reluctance of enumerators to approach people of higher status
  - Phones now make much easier
- Many surveys don’t track at all

Managing attrition

Thomas, Frankenberg, and Smith (2001)



Managing attrition

Thomas, Frankenberg, and Smith (2001)

- Ex ante
  - When designing the survey, work into the first wave the ability to find respondents later
    - Name of more than one person who might know respondent's whereabouts in a few years' time
    - Cell phone information
  - Develop a tracking protocol and budget for it
    - Example: distributing cell phones to all respondents

- Ex post
  - If experimental data, compare attritors with non-attritors in treatment and control groups along observable dimensions
  - If necessary, statistically adjust
    - Parametric (Hausman and Wise, 1979; Wooldridge, 2002; or Grasdal, 2001)
    - Non-parametric bounds (Manski, 1989; Lee, 2002)

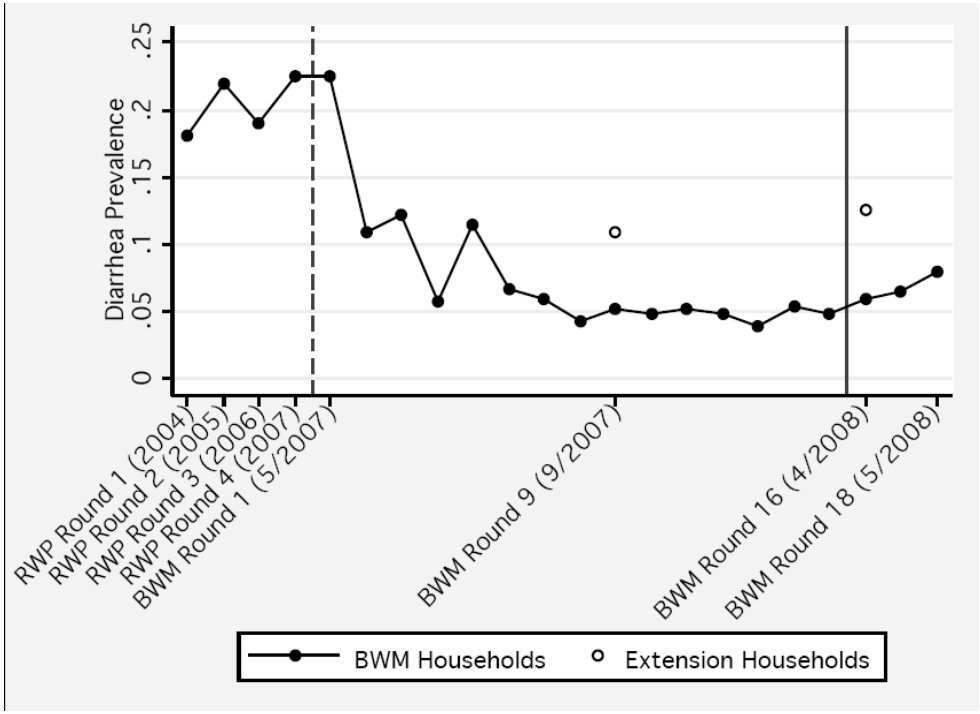
Mere measurement effects

- Look at evaluations that did the following
  - Random assignment of survey or frequency of survey
  - Surveyed and non-surveyed have opportunity to take up financial or health product
  - Have objective measure of take-up (rather than self-report)
- Survey increased likelihood of take-up in 4 out of 5 studies
  - Hospital insurance & health insurance (Philippines), microcredit (Morocco), microloans (India), home water treatment (Kenya)
- Example: Any insurance in Philippines (Gine, Karlan, and Zinman)
  - Survey increases take-up by 5.5 percentage points (from base of 37.2%)

Mere measurement effects

Kremer, Miguel, Null, Van Dusen, and Zwane (2009)

- Biweekly diarrhea prevalence measures and data collected annually from a comparable population
- Frequent data collection
  - Much lower reported diarrhea prevalence.
  - Higher levels of verified home water treatment



Outline

1. Whether to conduct fieldwork in development
2. How to do field research
3. Data collection and management
4. Survey design
5. Consumption
6. Profits
7. Attrition
8. Mere measurement effects
9. **Budgets and funding**
10. **Human subjects and permissions**

Budgeting

- Budget adequately
- Things go wrong
  - Exchange rate movements
  - Enumerator downtime
  - Resurveys needed
- New opportunities

Harvard Funding Sources

- **The Institute for Quantitative Social Science**
- **Weiss fund**
- **Warburg fund**
- **Weatherhead Center for International Affairs**
- The **Korea Institute**
- The **Minda de Gunzburg Center for European Studies**
- **Center for International Development**
  - up to $3,500 to support dissertation research on international development
- **The Harvard Graduate Student Council**
  - grants for conference attendance
  - summer research grant up to $1000.
- The **David Rockefeller Center for Latin American Studies**

Harvard Funding Sources cont'd

- The **Harvard Center for Population and Development Studies** runs the Saltonstall Population Innovation Fund
- The **Davis Center for Russian Studies** includes:
  - Merle Fainsod Prize for top incoming students
  - Foreign Language and Area Studies fellowships
  - research travel grants
  - dissertation completion fellowships.
- The **Fairbank Center for East Asian Research's** Shum Fellowship
  - $20,000 to spend a year studying in China
- The **Hauser Center**
  - full-time research and dissertation writing on a nonprofit sector/civil society topic

Outside Funding Sources

Non-Harvard affiliated programs

- The **Social Science Research Council**
- The **Organization of American States**
- **Fulbright**
- **Resources for the Future**
- **National Science Foundation**
- **Development specific sources:**
  - **Development Innovation Ventures**
  - **Global Innovation Fund**
  - **International Growth Center**
  - **PEDL grants from CEPR**

Development funding – writing applications

- Most care a great deal about dissemination and involvement of people from country
- Like co-financing
  - Sequence of funders
- Proposals likely read by people without economics background, best to avoid jargon
- Important to get on to letters of collaboration early
- Read rules carefully and look who has won funding before

Human subjects approval and Permissions

- Check what approvals needed
  - Country IRBs
  - University IRBs
- Allow adequate time
- Oral vs. written consent

- Permissions
  - National government
  - Local authorities
  - Relevant ministry

Making use of key informants

- Often have to go to authority figure to get permission to interview in an area
  - e.g. headmaster, village chief, district official
- They can help you/enumerators find people
  - Help provide listings of people (check not missing poorest), knows where to find people, summon them…
- May have to/want to ask them something
  - Can be considered insulting to ask poor people's view but not ask view of village chief
  - Use this to get at questions chief will know e.g. distance to local hospital or the capital, number of schools etc

RA work for faculty

- One way to get field experience is to work as an RA for a professor for a summer
- Occasional announcements to development lunch list, we may also know of other opportunities
- In some settings you may be able to negotiate for a personal side project
- Pros: Interesting work, learn from faculty, easier access to funding
- Cons: Doesn't lead directly to a Job Market Paper

END