

Survey at the crossroads: implementing electronic data collection on a large infrastructure survey in Kenya

3MCII, Chicago

July 27, 2016

Sarah M. Hughes, Mathematica Policy Research

“Collecting good data is one of the most powerful tools to end extreme poverty.”

Jim Yong Kim 2015



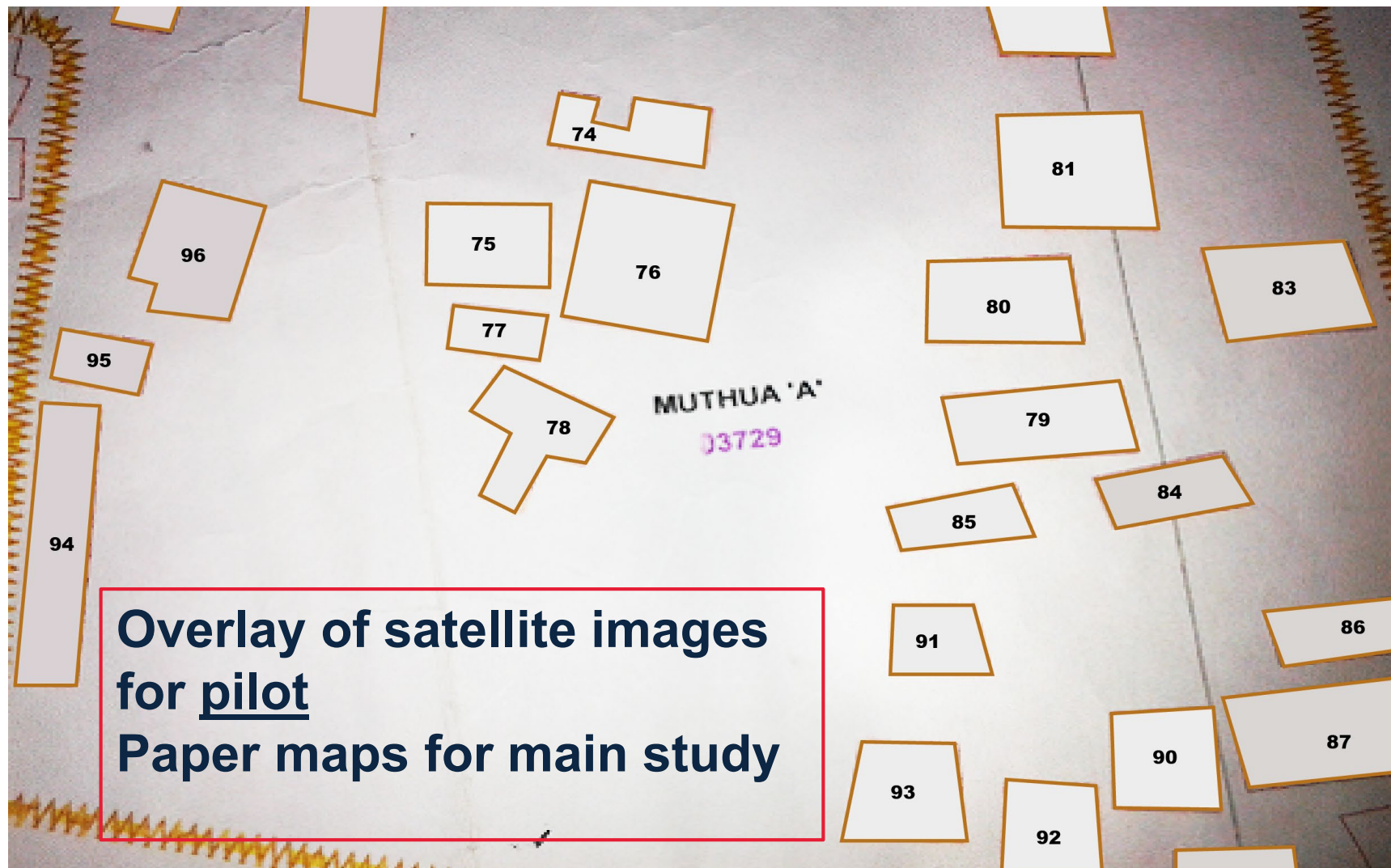
source: World Bank, 2015

Kenya Municipal Program baseline study

- **World Bank-sponsored 2012-2013**
- **Assessed infrastructure, living conditions and municipal service delivery**
- **15 municipalities**
- **Two-stage stratified cluster sampling design**
- **2,087 enumeration areas**
- **Full electronic listing of ~153,000 households**
- **30 minute tablet-programmed interviews**
- **14,581 households**

**Challenges: outdated sample frames,
rapidly urbanizing areas, refugees and
internal displacement**

Solutions: paper and digitized maps



Solutions: GIS and cluster guides

- “List and Go” with overnight sample selection
- Selected cases assigned to interviewer tablets



Challenges: uneven interviewer experience, incentives to cut corners

Solutions: interviewer training

Rate interviewers on ability to:

- Administer informed consent correctly
- Listen actively
- Use Show Cards
- Probe when necessary
- Read the question as it is written
- Read with correct voice inflection and tone
- Use question pacing
- Take and use instructions given
- Navigate through the tablet
- Know the subject matter
- Establish eye contact and build rapport
- Manage multiple tasks
- Effectively gain cooperation
- Apply professional interviewer skills
- Assure the respondents of confidentiality of their info
- Effectively manage difficult respondents
- Read interviewer instructions
- Take instruction

How would you rate trainee's overall performance?

Solutions: CAPI questionnaire

- Review item non-response rates
 - Don't know
 - Refuse
 - Not applicable
- Check for unusual filter question patterns
- Check for valid phone numbers
- Check for duplicate phone numbers
- Create consistency checks
- Review duration by interviewer, by supervisor

Solutions: CAPI sample management

- **Track by interviewer:**
 - Location of listed household vs. interviewed household (GIS)
 - Field result and final result
 - Contact attempts- time, day, date, fieldwork period, outcome
- **Track by supervisor team:**
 - Completes per day
 - Refusal rates
 - Location of clustered nonresponse
- **Frequent feedback to supervisors and separate data quality team conducting call-backs**

Results

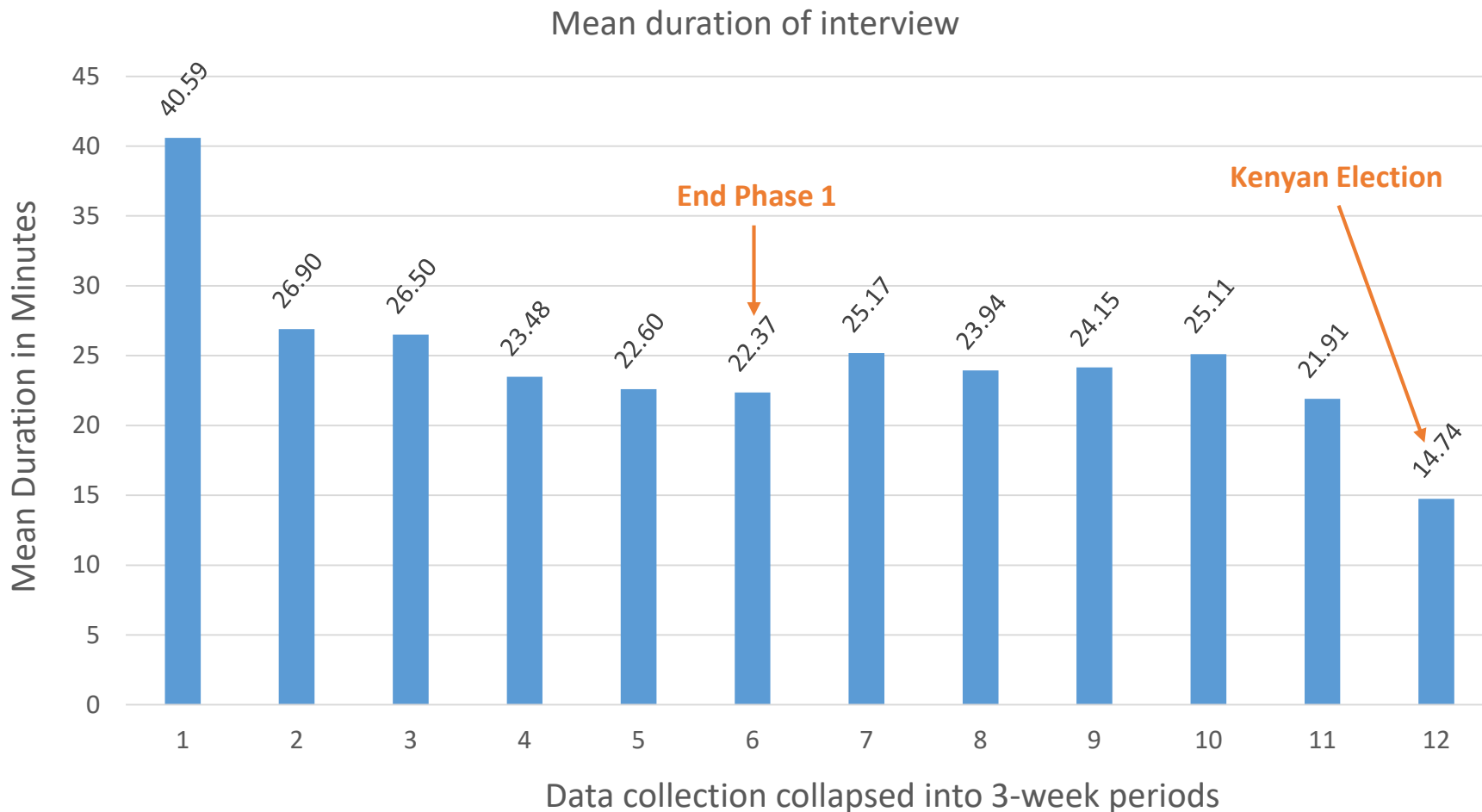
- **Listing task:**
 - Efficient use of CAPI for enumeration and selection
 - Little time to review contact results
- **Interviewing task:**
 - GPS comparison unwieldy and too prone to error for use during field
 - No variation in item nonresponse



Results

- Duplicate phone numbers appeared in the response data
- Post data collection review of filters didn't reveal obvious problems
- Durations were suspicious, but callbacks didn't reveal falsification

Mean duration of interview



Conclusions

- **Paradata can be useful for improving fieldwork, but a thorough review requires substantial labor and budget**
- **“Proof” of unacceptably poor data quality or faked cases very difficult to ascertain**

Further research needed:

- **Need to test whether poor adherence to survey methods, such as not reading questions precisely as they are written, affects estimates**

For More Information

- Sarah Hughes
 - SHughes@mathematica-mpr.com