Collecting Data Electronically in Developing Countries

Bansi Malde, Marcos Vera-Hernandez

Support from the ESRC/Hewlett Joint Scheme RES-183-25-0008 is gratefully acknowledged

PEPA is based at the IFS and CEMMAP
Motivation

• Explosion in collection of data by researchers in developing countries
• Driven by growing demand for evaluation of interventions and policies, and a lack of suitable existing data
• Desire to collect *high* quality data at *low* cost
• Primary method used in developing countries: Paper and pen
• But growing interest and efforts in using electronic data collection methods to collect data in developing countries
• This talk discusses the possibilities of using electronic devices, mainly PDAs, for data collection in developing countries
• Draws on our experience implementing a large household survey – The IFS/Mai Mwana Economic Survey in rural Malawi
Roadmap for Talk

- What are the possibilities for electronic collection methods?
- Considerations to Make
- Comparisons with paper-based surveys
- Our experience in Malawi:
  - Context and setting
  - Details on survey
  - Data Collection Set-Up
  - Interviewer Training
  - Implementation Experience
  - Interviewer Feedback
  - Data Quality
- Conclusions
THE POSSIBILITIES
Electronic methods:

• Allow for the collection of much of the same information collected via paper-based surveys
• AND that of information that cannot be collected using paper-based surveys only, such as:
  – GIS co-ordinates
  – Photographs
  – Videos
  – Sensitive information (e.g. Sexual partners, etc)
• Allow for the inclusion of in-built consistency checks → reduces errors, such as impossible values
• Automated routing of the questionnaire → the correct sections and questions answered
Possible Devices

- Personal Digital Assistants (PDAs) /Smart phones
- Laptops and Netbooks
- Tablet PCs
CONSIDERATIONS TO MAKE
Device Choice depends on:

- Cost
- Portability
- Battery life
- Field Conditions (e.g. heat, dust, electricity availability)
- Screen size
- Data storage and backup possibilities
- Data transfer capabilities
- Safety and security
- In-built functions (e.g. GPS receiver, camera, voice recorder)
Software

• Lots of available options (e.g. Entryware, Pocket Survey, Pendragon, and many more)

• Considerations to make:
  – Available support – very important
  – Costs
  – User friendly vs. Flexibility
  – Language possibilities
    • Important for languages with non-alphanumeric scripts (e.g. Hindi)
  – Complexity of the routing and types of questions you can programme
  – Possibility of consulting a database available in the device
    • Useful for longitudinal research
  – Automatic capture of GPS information
  – Format you want data in – Excel, SPSS, etc
Data Transmission Options

• Laptops
  – Data uploaded on the field and then manually collected together on a centralised server
  – Ideal in large geographic areas & in remote areas

• Mobile phone network
  – Not possible in remote areas, or areas with poor mobile networks
  – Improving though

• Automatic transmission of data from PDA to centralised online data server (via internet connection) – This will probably not work in remote areas
Consider a Kiosk Software

- Leave active only the applications needed for survey
- Reduces misuse, improves efficiency and makes PDA less attractive if stolen
Data Handling

• All the data is in soft format, so CRUCIAL to back up!
• Back-up on memory cards, USB sticks, etc
• Ensure that software used allows for response data to be backed up
• Train field staff on the importance of backing up the data!
COMPARISONS WITH PAPER-BASED SURVEYS
Advantages over paper-based surveys

- No need to print so many questionnaires – environmentally and budget friendly
- No additional data entry needed → Get data quicker, and reduces costs
- Avoid problems due to illegible writing
- Automatic routing → Faster interviews
- Only valid answers allowed
- Built in clock and time stamp → Useful for supervision
- Use of videos as part of questions
- Better quality data (Caeyers et al 2012 for comparison of consumption data)
Drawback over paper-based surveys

- All data is collected in soft format: No hard copies available
  - Data could be lost if device is lost or fails
  - Or if device has a virus
OUR EXPERIENCE IN MALAWI
Context

- Malawi – poor, mostly rural country in Southern Africa, between Tanzania, Mozambique and Zambia
- Survey was in Mchinji District, in the Central Region of Malawi
- Socio-economic conditions in Mchinji are similar or worse than the average for Malawi, overall still very poor in absolute terms: literacy rates of 60%, only 10% (2%) of households have access to piped water (electricity).
- Infrastructure in district very poor: only 2-3 tarmac roads in whole district and only 2% of population had access to electricity in 2004.
Survey

• Large household survey conducted by the IFS in collaboration with a health NGO, Mai Mwana, in Mchinji District, Malawi.

• Data collected to evaluate the economic effects of 2 health interventions implemented by Mai Mwana.


• Survey consisted of:
  – Individual questionnaire: ~ 150 questions asked on each household member including questions on education, work, self-reported health.
  – Household questionnaire: ~350 questions including a long consumption module, savings, assets, adverse events, family networks, information networks, transfers.
  – Main respondent questionnaire: ~40 questions on family planning and health knowledge.
  – Anthropometric measurements (height and weight) of all children < 6 years and their mothers.
Survey II

- Questionnaires included a range of open-ended quantitative and qualitative questions, categorical questions and multiple response questions

- Had some complex skip patterns, depending on:
  - Answers to previous questions
  - Age of individual
Data Collection Setup

- 24 interviewers, each with a PDA
- 3 supervisors, each covering 1/3 of the district, and each had a Laptop
- A Co-ordinator, who managed the central data server (amongst other things)
- The Mai Mwana Project handled all the fieldwork and related logistics (including training)
Equipment Used

- HP iPAQ 214 PDAs used (no mobile phone capability, large battery size, good screen size)
- Entryware Designer (to design questionnaire and manage data) and Entryware Mobile (on PDAs)
- Askarya Kioskbuilder – to allow access to certain programmes only
- BlueNext GPS (External GPS)
- Portable solar panel + battery provided
Interviewer ID number
(Note: this is the same as your PDA number)
PDAs in Action
Implementation Experience

• Excellent local partner – the Mai Mwana Project, who provided us with invaluable advice and local knowledge in hiring and training interviewers and setting up the fieldwork logistics

• Interviewers hired had completed secondary school and limited survey experience

• Interviewers had no IT knowledge, but most used mobiles – they picked up use of the PDA without too many problems

• Co-ordinator in wave 1 was highly competent and computer literate; while in wave 2 co-ordinator was one of wave 1 supervisors
Experience

• Data collection set up worked well (after initial learning period in wave 1)
• Far fewer technical issues with equipment than anticipated (but equipment was new)
  – 1 PDA failed in wave 1, and 2 in wave 2
• No thefts (luckily) during data collection
• Using an external GPS receiver connected to the PDA did not work well – recommend use of stand alone GPS units
• Data arrived very quickly once the survey started (once a month from Dec 2008)
• Data quality monitored closely during data collection → re-visits recommended in some cases
• Able to correct for errors in questionnaires, without having to re-print them (much quicker)
Interviewer Training

- Training was very important
- Intensive 2 week training period in wave 1, followed by 1 week of extensive piloting
- Retained 20/24 interviewers from wave 1 in wave 2 → training was somewhat easier – 2 weeks including piloting
- First introduced questionnaire, then allowed extensive practice on PDAs
- Data from pilots analysed to check for particularly problematic questions
- All this allowed to iron out errors in questionnaire programming, interviewer issues in using equipment
- Supervisors and co-ordinator held monthly meetings with interviewers to discuss issues
Interviewer Feedback

• Interviewers were very happy because they felt they were seen as smarter.
• Respondents were more interested and not threatened because of the PDAs.
• Most interviewers did not feel threatened (physical safety).
• Appreciated learning a new technology.
• No need to carry around a large number of papers.
• Interviews took less time than a paper-based interview would have taken (noticed by interviewers).
• We think that the use of PDAs increased the response rate because the interviewees felt curious about the new technology.
Data Quality

- Data is of generally good quality, with few invalid responses
- Interview length seems reasonable (median = 3820sec)
- But, there are a few issues to be aware of:
  - Using the in-built touch keypad can be difficult → need to be careful in entering numbers
  - For some important variables, e.g. identifiers, it is worth doing double entry (done in wave 2)
  - Automatic routing can lead to series of questions being wrongly asked or not asked due to errors – useful to give interviewers a paper version of the questionnaire to follow at the start
  - With roster questions, it can be difficult for interviewers to keep track of whose info has been recorded. Wave 2 questionnaire allowed for individual names to appear automatically to get around this issue.
Conclusion

• Using electronic data collection methods in poor, resource-constrained settings is not only feasible, but can also work well
• Summarise the available options
• Highlight some practical considerations to make in choosing equipment, data collection setup, software, etc
• Report from our experience implementing a large household survey in rural Malawi
Acknowledgements

• Many Thanks to staff at Mai Mwana, without whose advice and support this data collection exercise would not have been possible. We are particularly indebted to Tambosi Phiri, Hilda Chapota, Florida Banda, Sonia Lewycka and Mike Rosato.

• Many thanks to our survey co-ordinators, Andrew Mganga and Nicholas Mbwana, and supervisors Christopher Kamphinga, Aaron and Owen, and all the interviewers.

• Julia Behrman provided invaluable research assistance in setting up the first round of data collection.