Transitioning from Paper to Digital Data Collection for Non-Techies

Dr. Christopher Robert, Dobility/SurveyCTO
@ USAID Global Innovation Week
Introductions

- Me: Christopher Robert
- My company: Dobility, Inc.
- Our product: SurveyCTO
Introductions

- Me: Christopher Robert
- My company: Dobility, Inc.
- Our product: SurveyCTO
Digital data collection: what is it?

Photo: Emily Tomkys / OXFAM
Mobile data collection

- Interviewing people (CAPI), inspecting facilities, observing activities
- Recording data on tablets or smartphones
- Filling out electronic forms
Mobile case management

- Interviewing people (CAPI), inspecting facilities, observing activities
- Recording data on tablets or smartphones
- Filling out electronic forms
Other methods of digital data collection

- Interactive voice response (IVR)
  - Press 1 for yes, 2 for no
- Short message service (SMS)
  - Respond with YES or NO
- Computer-assisted web interviewing (CAWI)
  - Fill out this web survey
- Computer-assisted telephone interviewing (CATI)
  - Enumerators phone respondents and key their responses into a computer system

(not going to talk about this stuff)
Recent evolution in digital tools
GROUP JobGroup

Fields
Job ENG "Which of the following best describes your occupation?"
  ESP "¿Cuál de las siguientes opciones describe mejor su ocupación?": typJob
  JobSpec_01 "...(specify)" "...(Especifique)" : TOthSpec

RULES
Job
JobSpec_01
IF Job = Other THEN
  JobSpec_01 <> EMPTY "Please fill in the specify."
    "Por favor especifique."
ELSE
  JobSpec_01 := EMPTY
ENDIF
ENDGROUP  //Jobgroup
<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>start</td>
<td>start</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>end</td>
<td>end</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>deviceid</td>
<td>deviceid</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>subscriberid</td>
<td>subscriberid</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>simserial</td>
<td>simid</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>phonenum</td>
<td>phonenum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>note</td>
<td>intronote</td>
<td>Welcome to the sample cascading-select form. Please swipe forward to continue.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>select_one</td>
<td>yesno</td>
<td>consent</td>
<td>Would you like to continue?</td>
<td>yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>select_one</td>
<td>region</td>
<td>survey_region</td>
<td>In which region are you filling out this survey?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>select_one</td>
<td>country</td>
<td>survey_country</td>
<td>In which country?</td>
<td>$(consent)=1</td>
<td>yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>select_one</td>
<td>city</td>
<td>survey_city</td>
<td>In which city?</td>
<td>$(consent)=1</td>
<td>yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>text</td>
<td>name</td>
<td>What is your name?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>integer</td>
<td>age</td>
<td>How old are you?</td>
<td>&gt;3 and &lt;130</td>
<td>Please enter a valid age to continue.</td>
<td>$(consent)=1</td>
<td>yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>note</td>
<td>confirmnote</td>
<td>Your name is $(name) and your age is $(age). Thank you.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

...to Excel...
...to auto-formatted Excel...

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>type</td>
<td>name</td>
<td>label</td>
<td>hint</td>
<td>default</td>
<td>appearance</td>
<td>constraint</td>
<td>constraint message</td>
<td>relevance</td>
<td>disabled</td>
</tr>
<tr>
<td>2</td>
<td>start</td>
<td>starttime</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>end</td>
<td>endtime</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>deviceid</td>
<td>deviceid</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>subscriberid</td>
<td>subscriberid</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>simserial</td>
<td>simid</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>phonenumber</td>
<td>devicephonenum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>note</td>
<td>intronote</td>
<td>Welcome to the sample cascading select form. Please swipe forward to continue.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>select_one yesno</td>
<td>consent</td>
<td>Would you like to continue?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>select_one region</td>
<td>survey_region</td>
<td>In which region are you filling out this survey?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>select_one country</td>
<td>survey_country</td>
<td>In which country?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>select_one city</td>
<td>survey_city</td>
<td>In which city?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>text</td>
<td>name</td>
<td>What is your name?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>integer</td>
<td>age</td>
<td>How old are you?</td>
<td></td>
<td>&gt;3 and &lt;130</td>
<td>Please enter a valid age to continue.</td>
<td></td>
<td></td>
<td></td>
<td>yes</td>
</tr>
<tr>
<td>15</td>
<td>note</td>
<td>confirmnote</td>
<td>Your name is $(name) and your age is $(age). Thank you.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SurveyCTO
Powered by Debilly
...to drag-and-drop
Also in data quality: from cryptic...

display "Displaying percent DK/RF..."
foreach var of varlist _all {
    capture confirm numeric variable `var'
    if _rc == 0 {
        scalar miss = .
        scalar dk = .a
        scalar rf = .b
    }
    else {
        scalar miss = ""
        scalar dk = "don't know"
        scalar rf = "refusal"
    }
    quietly count if `var' != miss
    local nonmiss = r(N)
    quietly count if `var' == dk
    local dkn = r(N)
    local dkrate = `dkn' / `nonmiss'
    quietly count if `var' == rf
    local rfn = r(N)
    local rfrate = `rfn' / `nonmiss'
}

bysort enumerator: egen totdkrf = total(inlist(`var', dk, rf))
quietly count if totdkrf >= 3
if `dkrate' >= 0.025 | `rfrate' >= 0.025 | r(N) > 0 {
    describe `var'
    display "DK: " string(100 * `dkrate', "%5.1f") 
    display "RF: " string(100 * `rfrate', "%5.1f")
    tabulate enumerator `var' if inlist(`var', dk, rf) == 1, missing
}
drop totdkrf
}
Quality checks
You do not yet have any quality checks configured. Click the button to the right to add a quality check now.

Create a new quality check
Select the type of quality check:

- Value is too low
- Value is too high
- Value is an outlier
- Mean is too low
- Mean is too high
- Value is too frequent
- Value is too infrequent
- Group mean is different
- Group distribution is different

Warn when field value < this value:

- Critical quality check: flag warnings as high-priority

But ignore these values (separate with a comma):
And in data security: from cryptic...

openssl: Error: ‘?’ is an invalid command.

Standard commands
asn1parse  ca  ciphers  crl  crl2pkcs7
dgst       dh  dhpam  dsa  dsaparam
ec         ecpam  enc  engine  errstr
gendh      gendsa  genrsa  nseq  ocsp
passwd     pkcs12  pkcs7  pkcs8  prime
rand        req  rsa  rsautl  s_client
s_server    s_time  sess_id  smime  speed
spkac       verify  version  x509

Message Digest commands (see the 'dgst' command for more details)
md2         md4  md5  mdc2  rmd160
sha         sha1

Cipher commands (see the 'enc' command for more details)
aes-128-cbc aes-128-ecb aes-192-cbc aes-192-ecb aes-256-cbc
aes-256-ecb base64  bf  bf-cbc  bf-cfb
bf-ecb      bf-ofb  cast  cast-cbc  cast5-cbc
cast5-cfb   cast5-ecb  cast5-ofb  des  des-cbc
des-cfb     des-ecb  des-ede  des-ede-cbc  des-ede-cfb
des-ede-ofb des-ede3  des-ede3-cbc  des-ede3-cfb  des-ede3-ofb
des-ofb     des3  desx  rc2  rc2-40-cbc
rc2-64-cbc  rc2-cbc  rc2-cfb  rc2-ecb  rc2-ofb
rc4         rc4-40  seed  seed-cbc  seed-cfb
seed-ecb    seed-ofb

chrismac-2:scripts crobert$ openssl genrsa -out ~/Files/MyPrivateKey.pem 2048
Generating RSA private key, 2048 bit long modulus
........................+++  
...........+++  
e is 65537 (0x10001)
chrismac-2:scripts crobert$ openssl rsa -in ~/Files/MyPrivateKey.pem -inform PEM -out ~/Files/MyPublicKey.pem -outform PEM -pubout
writing RSA key
...to simple

Step 1. Choose a name for your private key

Please enter a short name for your encryption key. This will be used to name the encryption key files that you will download in the following steps.

Key name:

MyPersonalKey

Next

Step 2. Download the private key

Please now download the private key file. This is the file that you must safeguard most diligently as it is the key used to decrypt your data. Anybody with access to both this key file and your data can decrypt your data. And if you lose this key file, you will be totally unable to decrypt your data.

Download private key file

Back

Next

Step 3. Download the public key

Please now download the public key file. This public key is the one that you will attach to your forms, in order to encrypt their data. (The public key is public, used to encrypt data, and the private key is private, used to decrypt data. Anybody can encrypt data, but only you or your team should have the private key necessary to decrypt data.)

Download public key file

Back
How it works
Key concept: **forms**

- What is filled out to collect data. Can be anything:
  - A questionnaire
  - An inspection checklist and protocol
  - Ground-truthing for remote sensing, ...
Key concept: fields

- The individual units of data-collection within a form. Can be:
  - Questions (with numeric, text, photo, video, audio, etc. responses)
  - Scanned barcodes, captured GPS positions, calculations, ...
Key concept: relevance

- When a field (or group of fields) should appear. Can be based on:
  - Earlier responses
  - Calculations
  - Pre-loaded data
Key concept: **constraint**

- **What responses are valid.** Can be based on:
  - Earlier responses
  - Calculations
  - Pre-loaded data
Advanced concept: audio audits

Back-checks are hard. Why not listen in?

SurveyCTO
Powered by Dobility
**Advanced concept: speed limits**

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>appearance:</td>
<td>(none) - just an ordinary select_one field</td>
</tr>
<tr>
<td>required message:</td>
<td>Custom &quot;a response is required&quot; message (if any).</td>
</tr>
<tr>
<td>publishable:</td>
<td>For encrypted forms, leave this field publishable on the server?</td>
</tr>
<tr>
<td>minimum_seconds:</td>
<td>If using speed limits, how many seconds should user spend on this field?</td>
</tr>
<tr>
<td>media:image:</td>
<td>Display an image as part of this field.</td>
</tr>
</tbody>
</table>
Advanced concept: real-time publishing

Executive Education - Decision-making Simulation
Publishing to: Google Spreadsheet

Form publishing options

GOOGLE SPREADSHEETS
Executive Education - Decision-making Simulation

Publishing data to Google Spreadsheet "Executive Education - Decision-making Simulation"

Field mapping:
- SubmissionDate
- starttime
- endtime
- duration
Advanced concept: real-time dashboards

http://blog.surveycto.com/hacking-google-sheets-for-real-time-dashboards/
Advanced concept: **real-time monitoring**

![SurveyCTO Data Explorer - Monitor](image)

<table>
<thead>
<tr>
<th>SUBMISSIONS</th>
<th>total: 18</th>
<th>not loaded: 0</th>
<th>excluded: 0</th>
<th>available: 18</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIELDS</td>
<td>total: 17</td>
<td>not loaded: 0</td>
<td>note fields: 0</td>
<td>available: 17</td>
</tr>
<tr>
<td>DURATION</td>
<td>total: 54m 32s</td>
<td>average: 3m 1s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHECKS</td>
<td>configured: 3</td>
<td>warnings: 6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Start by adding some field or relationship summaries...
Advanced concept: real-time monitoring
Advanced concept: **real-time monitoring**

![Data Explorer - Monitor](image)

**Household listing - round 1**

<table>
<thead>
<tr>
<th>Enumerator / Duration</th>
<th>Precious Rutendo</th>
<th>Rejoice Simelinkosi</th>
<th>Johanna Moyo</th>
<th>Mercy Wadzanai</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean of duration</td>
<td>233</td>
<td>178.33</td>
<td>200</td>
<td>147.6</td>
</tr>
</tbody>
</table>

Display: mean (duration)
Advanced concept: real-time monitoring
Advanced concept: real-time monitoring
Advanced concept: **real-time monitoring**
A quick demo

Photo: Laura Elizabeth Pohl / EGPAF
Questions?
COLLECT DATA YOU CAN TRUST.
Offline or online, in the field, on the street, or in the lab.
Use SurveyCTO to collect your best data ever.
Thank you!

Please reach out for any questions:
crobert@surveycto.com
Paper vs. digital: key trade-offs
Evidence for digital data collection
Evidence for CAPI

Table 1: Summary statistics on errors, interviewer and survey characteristics and sample size

<table>
<thead>
<tr>
<th></th>
<th>Full CAPI</th>
<th>Restricted CAPI</th>
<th>PAPI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Routing errors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average number of routing errors per HH (total)</td>
<td>0.9</td>
<td>0.6</td>
<td>10.4</td>
</tr>
<tr>
<td>Average nr of entries in to be skipped fields per HH</td>
<td>0.0</td>
<td>0.1</td>
<td>6.3</td>
</tr>
<tr>
<td>Average nr of missing entries in required fields per HH</td>
<td>0.0</td>
<td>0.5</td>
<td>4.0</td>
</tr>
<tr>
<td><strong>Impossible/Unlikely entries</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average nr of impossible entries per HH</td>
<td>0.0</td>
<td>0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>Average nr of unlikely entries per HH</td>
<td>0.6</td>
<td>1.1</td>
<td>1.6</td>
</tr>
<tr>
<td><strong>Errors/unlikely entries per survey period quartile</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st quartile</td>
<td>0.8 (1.1)</td>
<td>2.3 (2.2)</td>
<td>12 (10)</td>
</tr>
<tr>
<td>2nd quartile</td>
<td>0.8 (1.0)</td>
<td>2.1 (1.8)</td>
<td>12.2 (11.6)</td>
</tr>
<tr>
<td>3rd quartile</td>
<td>0.5 (0.9)</td>
<td>1.7 (1.6)</td>
<td>9.4 (10.2)</td>
</tr>
<tr>
<td>4th quartile</td>
<td>0.4 (0.7)</td>
<td>1.2 (1.2)</td>
<td>8.4 (7.4)</td>
</tr>
<tr>
<td><strong>GPS data</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% HHs &gt; 1 km from cluster centre (likely outliers given the small size of EAs)</td>
<td>0.6</td>
<td>1.3</td>
<td>6.6</td>
</tr>
<tr>
<td><strong>Time stamp data</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% surveys with problematic time stamps</td>
<td>0.9</td>
<td>0.3</td>
<td>23.8</td>
</tr>
<tr>
<td>% surveys conducted on day 1 of the cluster visit</td>
<td>45.6</td>
<td>50.0</td>
<td>49.1</td>
</tr>
<tr>
<td>Average survey duration</td>
<td>81 (24)</td>
<td>78 (23)</td>
<td>89 (25)</td>
</tr>
</tbody>
</table>

Table 2: Effect of CAPI and Checks on data quality

<table>
<thead>
<tr>
<th>LHS</th>
<th>Routing errors: Missing entries in required fields</th>
<th>Routing errors: Entries in fields that should have been skipped</th>
<th>Impossible entries</th>
<th>Unlikely entries</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>OLS</td>
<td>OLS</td>
<td>OLS</td>
</tr>
<tr>
<td>CAPI</td>
<td>-3.544***</td>
<td>-6.225***</td>
<td>-0.337***</td>
<td>-0.259***</td>
</tr>
<tr>
<td></td>
<td>(0.187)</td>
<td>(0.306)</td>
<td>(0.040)</td>
<td>(0.087)</td>
</tr>
<tr>
<td>Checks</td>
<td>-0.490***</td>
<td>-0.117</td>
<td>-0.148***</td>
<td>-0.461***</td>
</tr>
<tr>
<td></td>
<td>(0.149)</td>
<td>(0.244)</td>
<td>(0.032)</td>
<td>(0.069)</td>
</tr>
<tr>
<td>Const.</td>
<td>4.038***</td>
<td>6.344***</td>
<td>0.487***</td>
<td>1.347***</td>
</tr>
<tr>
<td></td>
<td>(0.132)</td>
<td>(0.216)</td>
<td>(0.028)</td>
<td>(0.061)</td>
</tr>
</tbody>
</table>

Panel 1

<table>
<thead>
<tr>
<th>LHS</th>
<th>Potentially missing values in non-consumption sections (OLS)</th>
<th>Potentially missing values in consumption section</th>
<th>Time Stamp Problems</th>
<th>GPS Problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAPI</td>
<td>-2.800***</td>
<td>-1.081***</td>
<td>-0.234***</td>
<td>-0.053***</td>
</tr>
<tr>
<td></td>
<td>(0.142)</td>
<td>(0.121)</td>
<td>(0.015)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Checks</td>
<td>-0.287**</td>
<td>-0.351**</td>
<td>0.006</td>
<td>-0.007</td>
</tr>
<tr>
<td></td>
<td>(0.113)</td>
<td>(0.096)</td>
<td>(0.012)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Const.</td>
<td>3.091***</td>
<td>1.434***</td>
<td>0.238**</td>
<td>0.066**</td>
</tr>
<tr>
<td></td>
<td>(0.101)</td>
<td>(0.085)</td>
<td>(0.011)</td>
<td>(0.007)</td>
</tr>
</tbody>
</table>

Notes: N=1480, *** p<0.01, ** p<0.05, * p<0.1
1. Standard errors are shown in parentheses
2. All estimates are robust to cluster and interviewer fixed effects, controlling for respondent characteristics of household size, income, household characteristics (size of household members) and interview day.

Channels by which CAPI improves quality

• Enforced logic constraints and skip patterns during interview
• Quicker access to data for monitoring and quality control
• Advanced survey auditing tools: GPS, audio, timing
• Dynamic data loading across questions and across surveys
• Dynamic choice options
### Evidence for Digital Technologies

#### Table 2: Pros and cons of mobile technologies for survey data collection

<table>
<thead>
<tr>
<th></th>
<th>Audio/Text</th>
<th>Self-Reporting</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SMS</strong> (Short Message)</td>
<td>Text</td>
<td>Yes</td>
<td>Low cost</td>
<td>Maximum 160 characters</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Requires literacy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Does not allow visual aids</td>
</tr>
<tr>
<td><strong>IVR</strong> (Interactive Voice Recognition)</td>
<td>Audio</td>
<td>Yes</td>
<td>No need for operators</td>
<td>Often viewed as annoying</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Medium cost</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Does not allow visual aids</td>
</tr>
<tr>
<td><strong>CATI</strong> (Computer-Assisted Telephone Interview)</td>
<td>Audio</td>
<td>No</td>
<td>Respondent can ask to clarify questions</td>
<td>Higher cost than SMS or IVR, mainly because: (i) voice is more expensive than text communications and (ii) operators salaries.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Does not allow visual aids</td>
</tr>
<tr>
<td><strong>USSD</strong></td>
<td>Text</td>
<td>Yes</td>
<td>No length limitations.</td>
<td>Requires close collaboration and approval by telecom companies. Not commonly marketed in LAC.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Does not allow visual aids</td>
</tr>
<tr>
<td><strong>Mobile internet</strong></td>
<td>Text</td>
<td>Yes</td>
<td>- No length limitations.</td>
<td>Limited GPRS coverage in LAC, Requires smart phones.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Lower cost than voice communications.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Allows use of visual aids</td>
<td></td>
</tr>
</tbody>
</table>

#### Table 9: Attrition by mode in Peru

<table>
<thead>
<tr>
<th>Wave</th>
<th>IVR</th>
<th>SMS</th>
<th>CATI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wave 1</td>
<td>80%</td>
<td>70%</td>
<td>49%</td>
</tr>
<tr>
<td>Wave 2</td>
<td>75%</td>
<td>75%</td>
<td>47%</td>
</tr>
<tr>
<td>Wave 3</td>
<td>78%</td>
<td>76%</td>
<td>49%</td>
</tr>
<tr>
<td>Wave 4</td>
<td>78%</td>
<td>75%</td>
<td>52%</td>
</tr>
<tr>
<td>Wave 5</td>
<td>84%</td>
<td>76%</td>
<td>53%</td>
</tr>
<tr>
<td>Wave 6</td>
<td>81%</td>
<td>79%</td>
<td>61%</td>
</tr>
</tbody>
</table>

Comparing Methods: **Template Design**

**CAPI**

- Template needs to be designed before data collection can begin
- Template will be loaded onto the devices data collectors use in the field
- Data will be recorded directly from the interview onto the device
Comparing Methods: Template Design

CAPI

• Template needs to be designed before data collection can begin
• Template will be loaded onto the devices data collectors use in the field
• Data will be recorded directly from the interview onto the device

PAPI

• Physical survey forms must be designed before data collection can begin
• Data-entry template needs to be designed before data from the paper surveys can be entered into a computer
Comparing Methods: Question Types

CAPI

• Forms can allow multiple-choice and write-in responses
• Forms can capture the current GPS location, scan barcodes, take photos, record audio or video, perform calculations, capture signatures, allow respondents to draw onto photos, and more
Comparing Methods: **Question Types**

**CAPI**
- Forms can allow multiple-choice and write-in responses
- Forms can capture the current GPS location, scan barcodes, take photos, record audio or video, perform calculations, capture signatures, allow respondents to draw onto photos, and more

**PAPI**
- Forms can allow both multiple-choice and write-in responses
- Forms can allow spaces for the respondent to sign or draw directly onto the form
Comparing Methods: Hardware

CAPI

• Smartphones, tablets, laptops, or netbooks will be needed for each of the data collectors to perform interviews and record responses
Comparing Methods: **Hardware**

**CAPI**

- Smartphones, tablets, laptops, or netbooks will be needed for each of the data collectors to perform interviews and record responses

**PAPI**

- Computers will be needed for data entry. If the project will do in-field data entry, laptops or netbooks will be needed
Comparing Methods: Software

CAPI

• Different software options are available
• Pricing, programming difficulty, support, and flexibility vary widely
Comparing Methods: **Software**

**CAPI**
- Different software options are available
- Pricing, programming difficulty, support, and flexibility vary widely

**PAPI**
- Data-entry software will be needed
- Some options are available at no cost (e.g., CSPro), while others are not (e.g., Microsoft Access)
Comparing Methods: **Printing**

**CAPI**

- No printing is required for a CAPI survey
Comparing Methods: Printing

CAPI
• No printing is required for a CAPI survey

PAPI
• Paper-based survey forms must be printed
  – If they are printed in bulk and then revisions are later required (a common situation), stickers or replacement pages may need to be printed and then inserted into existing forms
Comparing Methods: **Enumerator Training**

**CAPI**

- Since skip patterns and field validation are handled by the data-collection device, enumerators can be trained more quickly.
- However, those not used to using smartphone or tablet technology may need extra time to become comfortable using the devices.
Comparing Methods: Enumerator Training

CAPI

- Since skip patterns and field validation are handled by the data-collection device, enumerators can be trained more quickly.
- However, those not used to using smartphone or tablet technology may need extra time to become comfortable using the devices.

PAPI

- Complex or numerous skip patterns can be difficult for enumerators to master, and it may take much practice for enumerators to reliably enter valid responses into all fields.
- The necessary training period could be lengthy.
Aside: Paper Survey Examples
### B1 LAND OWNERSHIP AND CULTIVATION क्षेत्र का पता नहीं रखा?

<table>
<thead>
<tr>
<th>पता नहीं (सामय)</th>
<th>बनाया 2007</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### B2 RENTED, PAWNED AND SHARECROPPED LAND भेड़ या कुर्टी पता नहीं?

<table>
<thead>
<tr>
<th>मान / दर (करें)</th>
<th>रस्ता पता नहीं (करें)</th>
<th>रस्ता पता नहीं (करें)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Rs</td>
<td>Rs</td>
<td>Rs</td>
</tr>
</tbody>
</table>

### B3 TRANSFERS, DISASTERS, AND RELIEF रंग में नहीं किया?

<table>
<thead>
<tr>
<th>रंग में 12 अक्टूबर</th>
<th>रंग में 12 अक्टूबर</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>2</td>
</tr>
<tr>
<td>Rs</td>
<td>Rs</td>
</tr>
</tbody>
</table>

### R4 पृथक रूप से समाधान प्रदान कीजिये?

<table>
<thead>
<tr>
<th>पृथक रूप से मिलाए जाने वाले विषयों की संख्या</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
</tbody>
</table>

### R5 पृथक रूप से पक्ष पर चर्चा की?

<table>
<thead>
<tr>
<th>पृथक रूप से पक्ष पर चर्चा की?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

### R7 अन्य अवधि के लिए रंग में सामग्री दी?

<table>
<thead>
<tr>
<th>रंग में 12 अक्टूबर</th>
<th>रंग में 12 अक्टूबर</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>2</td>
</tr>
<tr>
<td>Rs</td>
<td>Rs</td>
</tr>
</tbody>
</table>

---

* अर. – (अरबी) लिए (क्रिंडो)
* परिम. – परिम. लिए (स्पष्ट)
Aside: CAPI Survey Examples
Comparing Methods: Transportation & Storage

CAPI

- Surveys are typically "transported" via USB connection, local wi-fi network, or the Internet, then "stored" on a server or local hard drive
- Proper precautions must be taken (generally using data encryption) to ensure that the data is secure both in transit and, to the greatest extent possible, at rest
Comparing Methods: **Transportation & Storage**

**CAPI**

- Surveys are typically "transported" via USB connection, local wi-fi network, or the Internet, then "stored" on a server or local hard drive
- Proper precautions must be taken (generally using data encryption) to ensure that the data is secure both in transit and, to the greatest extent possible, at rest

**PAPI**

- A secure system must be designed to bring the surveys from the field to the office, and to store the surveys in the office
- Once data is entered into the computer, it should be stored securely (generally using data encryption)
Comparing Methods: Data Entry

CAPI

• The data is instantly digitized as it is entered into the tablet
Comparing Methods: Data Entry

CAPI

• The data is instantly digitized as it is entered into the tablet

PAPI

• Data must be entered into the programmed template by trained data-entry operators

• For each form or questionnaire, typically two entries are completed, they are compared, and any differences between the two are corrected
Comparing Methods: Quality Control

CAPI

- The flow of questions is automated, so the enumerator does not need to decide which question comes next
- CAPI supports logical checks, pre-filling information, setting, and secretly recording audio for later review
- The data is available right away in a digitized format for review
Comparing Methods: Quality Control

CAPI

- The flow of questions is automated, so the enumerator does not need to decide which question comes next
- CAPI supports logical checks, pre-filling information, setting, and secretly recording audio for later review
- The data is available right away in a digitized format for review

PAPI

- Typically the survey team will include "scrutinizers," whose job it is to look carefully at the completed questionnaires and catch enumerator errors
- The questionnaire can then be sent back to the field for correction
Comparing Methods: **Data Cleaning**

**CAPI**

- The data cleaning process can begin after the first day of surveying
  - At this time the enumerators are still able to return to any household if they need to make a correction
Comparing Methods: **Data Cleaning**

**CAPI**
- The data cleaning process can begin after the first day of surveying
  - At this time the enumerators are still able to return to any household if they need to make a correction

**PAPI**
- Once the data entry is complete, the cleaning process can begin
Comparing Methods: Data Availability

CAPI

- In the typical set-up where field teams securely upload data to a server on a nightly basis, data is available by day two of data collection
Comparing Methods: Data Availability

CAPI

- In the typical set-up where field teams securely upload data to a server on a nightly basis, data is available by day two of data collection

PAPI

- Data typically becomes available only after all of it has been entered and cleaned
  - This can be weeks or months after the actual data collection
### Comparing Processes

<table>
<thead>
<tr>
<th>CAPI</th>
<th>PAPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questionnaire design</td>
<td>Questionnaire design</td>
</tr>
<tr>
<td>Coding digital questionnaire</td>
<td></td>
</tr>
<tr>
<td>Piloting and testing</td>
<td>Piloting and testing</td>
</tr>
<tr>
<td>Training (digital and paper)</td>
<td>Training (paper)</td>
</tr>
<tr>
<td>Survey data collection</td>
<td>Survey data collection</td>
</tr>
<tr>
<td>Monitoring data quality</td>
<td></td>
</tr>
<tr>
<td>Community feedback summary statistics</td>
<td>Coding data-entry template</td>
</tr>
<tr>
<td>Data analysis</td>
<td>Data analysis</td>
</tr>
</tbody>
</table>
## Comparing Costs (SurveyCTO)

<table>
<thead>
<tr>
<th></th>
<th>CAPI</th>
<th>PAPI</th>
<th>CAPI - PAPI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Software</strong></td>
<td>$99*</td>
<td>$0</td>
<td>+ $99</td>
</tr>
<tr>
<td><strong>Printing</strong></td>
<td>$60 (10% budgeted for backup)</td>
<td>$300</td>
<td>- $240</td>
</tr>
<tr>
<td><strong>Data entry</strong></td>
<td>$210 (10% budgeted for backup)</td>
<td>$2100 (approximate cost of 4 data-entry clerks and 1 supervisor for 15 days)</td>
<td>- $1,890</td>
</tr>
<tr>
<td><strong>TOTAL (excluding hardware)</strong></td>
<td><strong>$369</strong></td>
<td><strong>$2400</strong></td>
<td><strong>- $2,031</strong></td>
</tr>
</tbody>
</table>
Comparing Costs (SurveyCTO)

<table>
<thead>
<tr>
<th></th>
<th>CAPI</th>
<th>PAPI</th>
<th>CAPI - PAPI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hardware</strong></td>
<td><strong>$1000</strong> (based on 10 devices at $100/device)</td>
<td><strong>$0</strong></td>
<td>+ $1000</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>$1369</strong></td>
<td><strong>$2400</strong></td>
<td>- $1,031</td>
</tr>
<tr>
<td>(including hardware)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Adapted from Oxfam's experience shifting to SurveyCTO for several effectiveness reviews. We changed the currency to U.S. dollars, updated the software cost to reflect our $99 professional plan (versus £162 in the original table), and revised the projected hardware cost to $100/unit (versus £300 in the original table).

* Software cost: additional savings when more than one survey runs in the same month.
Deciding Between CAPI and PAPI

Questions you should ask yourself:

• How complex is my survey?
• How much time do I have before survey launch?
• How quickly do I need to receive incoming data?
• How technically-skilled are available trainees/staff?
• Can I procure or rent phones or tablets?
• What is my catchment area?
• What is the mobile data connectivity like? Road quality?
• What is my budget?
  – Digital instrument design time, devices, mobile data, shipping, generator and fuel, power strips and chargers, repairs, depreciation, theft
  – Can we pay software costs when actively collecting data?
Questions?
Thank you!

Please reach out for any questions:
crobert@surveycto.com
What is SurveyCTO?
SurveyCTO is a mobile data-collection platform that enables you to collect high-quality data, even when you’re offline.
Our story

Dobility grew out of several Harvard-based development economics projects based in Tamil Nadu, South India.

The same technical challenges seemed to plague every project, and it became apparent that better technology could dramatically simplify the lives of research team members, reduce project costs, and improve data quality.

So we created SurveyCTO...
Today, thousands of users in about 150 countries depend on SurveyCTO for their data collection.
How to create forms that collect exactly what you need...
Create visible fields...

What label should appear for this field?
Your question text here

What type of field should this be?

- text
- integer
- decimal
- select_one
- select_multiple
- geopoint
- barcode
- date
- datetime
- time
- image
- audio
- video
- file
- note

Cancel
Configure...
...and hidden fields

What type of hidden field would you like to add?

- calculate
- comments
- text audit
- audio audit
- speed violations audit
- speed violations count
- speed violations list
- calculate here

[Buttons: Cancel, Configure...]
Skip irrelevant questions

Add select one field

Save now
Cancel

Shortcuts:
- Required options
- Labels
- Relevance
- Constraint
- Other options

Add relevance - Step 1

Relevance

Help

A field will only display when it is deemed "relevant". Use the + and – buttons below to add and remove conditions.

Field is relevant if All of the following conditions are true:

Field: gender is equal to the text value

Field: age is greater than the numeric value 10

Add condition
Save

Cancel
Constrain responses

Add integer field

Add constraint - Step 2: Constraint parameters

- Allow decimals?
- Minimum value: (leave blank for none)
  - Value: 18
- Maximum value: (leave blank for none)
  - Value: 120
- Allowable exceptions: (comma-separated list)
  - Values: -888, -999
- Please enter the error message to display when the constraint is violated:
  - Message: Please enter a valid age for this respondent (18-120).
Capture GPS
Capture GPS

Add geopoint field

Shortcuts:
- Required options
- Labels
- Relevance
- Other options

GPS
GPS coordinates can only be collected when outside.

Replace Location

Show Location

Latitude: N 41°17'3"
Longitude: W 72°56'1"
Altitude: -26m
Accuracy: 14m
Capture images (and videos, barcodes + more)

Add image field

Save now
Cancel

Shortcuts:
Required options
Labels
Relevance
Other options

Other options

appearance:
(none) - just an ordinary image field
new - force user to take new picture, not select existing one
annotate - allow user to annotate the image
draw - have user draw something

required message:
signature - have user draw a signature
other - custom appearance (specify below)

Custom "a response is required" message (if any).

publishable:
For encrypted forms, leave this field publishable on the server?
Yes
No

minimum_seconds:
If using speed limits, how many seconds should user spend on this field?
Capture images (and videos, barcodes + more)
Capture images (and videos, barcodes + more)
Customize appearances

<table>
<thead>
<tr>
<th>photo</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Take a photo of the market</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>gps</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>GPS</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>fruit</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Which fruit do you like?</td>
<td></td>
</tr>
</tbody>
</table>
Customize appearances

<table>
<thead>
<tr>
<th>photo</th>
<th>Take a photo of the market</th>
</tr>
</thead>
<tbody>
<tr>
<td>gps</td>
<td>GPS</td>
</tr>
<tr>
<td>fruit</td>
<td>Which fruit do you like?</td>
</tr>
</tbody>
</table>

The screen shows a mobile application interface with options to customize appearances. The application appears to be collecting data about preferences, specifically asking which fruit the user likes.
Switch between languages

Edit form settings

languages
english, french

default_language
english

Save
Cancel

Edit
Add Language
Remove Language
Rename Language
OK
Switch between languages

Select the fruit that you like:

This is the question or prompt as it will appear to users of your form.

Sélectionnez le fruit que vous aimez:

This is the question or prompt as it will appear to users of your form.
Switch between languages

Test form for presentation

Sélectionnez le fruit que vous aimez:

- Banana
- Mango
- Apple

Next >

This is the question or prompt as it will appear to users of your form.
Access your data instantly...

...and continuously monitor for quality
Designed for safety
Designed for safety: **encryption and redundancy**

- All data in transit is encrypted with Secure Sockets Layer (SSL) technology

- Option to encrypt all of your data at the point of collection, using your own 2048-bit encryption keys

- All collected survey data is stored on the device's removable memory card AND backed up on the device's internal memory

- SurveyCTO's server databases are mirrored 24x7 across multiple of Amazon's data facilities

- Continuous backups are restorable to any minute of any day in the last month
Designed for safety: **encryption made easy**

---

**Your forms**

- Start new
- Upload form
- Tools
- Refresh
- Help

- Build constraint
- Test constraint
- Build relevance
- Build calculation
- Create new key

---

**Click here** to learn more about encrypting form data with your own encryption keys. When you're ready to begin, you can click the button below to open the key generator and generate a new public/private key pair right here in your browser. This will happen locally, on your computer, and the key pair will not be transmitted to the SurveyCTO server. You will need to closely guard the private key file, and you will need to add the public key to any forms you wish to encrypt.

- [Start key generator](#)
- Close

---

**SurveyCTO**

Powered by Debilly
Designed for safety: encryption made easy

To encrypt your form data, you will need your own encryption key, which you can get from the Tools...Create new key option at the very top of this "Your forms" section. Click here to learn more...

- Upload public key
- Paste public key text
Integrate (safely) with other platforms
Integrate (safely) with other platforms
Integrate (safely) with other platforms
Integrate (safely) with other platforms
Integrate (safely) with other platforms
Integrate (safely) with other platforms
All while accessing professional support...

...from experts in research and survey design
Questions?
Thank you!

Please reach out for any questions:
crobert@surveycito.com