

Using Stata for Questionnaire Development

Theodore Pollari and Phil Schumm

Research Computing Group
Department of Health Studies
University of Chicago

August 23, 2004

(Supported by National Institute on Aging grant 1R01 AG021487-01A1)

Creating a Questionnaire

Standard Method

A Better Alternative

Explanation by Example

Schema for Individual Questions

Cross-References Between Items

Control Logic: Contingencies and Blocks

Discussion of Suggested Questionnaire Development System

Review of Our System

Review of Advantages of Our System

Future Plans

Standard method of creating a questionnaire

Questionnaires are usually created using a word-processor. This has several disadvantages:

- ▶ Editing is inefficient (e.g., changing response categories for multiple questions, re-ordering questions, specifying control logic, etc.)
- ▶ Not amenable to version control
- ▶ Must be manually re-entered into CAPI
- ▶ No mechanism for transferring metadata to final data set

A Better Alternative

Fixing Problems with a Systematic Process



- ▶ Create questionnaire as Stata .do file (or files) according to a specific schema
- ▶ Use .do files to create a Stata dataset representing a complete questionnaire along with all relevant metadata.

Explanation by Example
National Social Life, Health and Aging Project

- ▶ National survey of 3,000 older adults
- ▶ Supported by a grant from the National Institute on Aging
- ▶ Fielded by NORC using commercial CAPI software (SPSS MR)
- ▶ Interdisciplinary project involving sociologists, economists, psychologists, and physicians
 - ▶ Special challenges in integrating different research questions into a single instrument
 - ▶ Effective collaboration both essential and sometimes difficult to achieve

Example: An Individual Question

Stata Code Specification

- ▶ Represent questions as individual variables
- ▶ Use characteristics to represent question text and control logic
- ▶ Use notes and additional characteristics to represent metadata

Example: An Individual Question

Date of Birth : Stata Code

* Generate vars

```
gen dob = .
```

```
...
```

```
char dob[qt] In what month, day, and year were you born?
```

```
char dob[mask] _ _ _ _ _ (month) _ _ _ _ _ (day) _ _ _ _ _ (year)
```

```
lab var dob "date of birth"
```

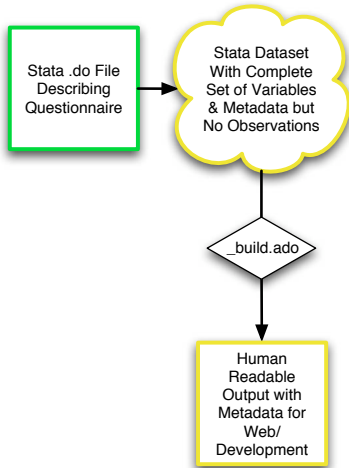
```
char dob[src] 1992 HRS
```

```
char dob[dom] sociodemog
```

```
note dob: 8/25/2003 - gpk
```

A Better Alternative

Fixing Problems with a Systematic Process



Example: An Individual Question

Date of Birth : Human Readable Form



Turn comments off

Up: [Table Of Contents](#) Next: [Physical Health](#)

Demographics

Version 2.2 (last edited by Genevieve Pham-Kanter on 27-Nov-03)

[Add comment](#)

Q1. In what month, day, and year were you born?

dob:

_____ (month) _____ (day) _____ (year)

LABEL: date of birth
SOURCE: 1992 HRS
DOMAINS: sociodemog

NOTES: 8/25/2003 - gpk

[Add comment](#)

Example: An Individual Question
Stata Code Specification

- ▶ Represent questions as individual variables
- ▶ Use characteristics to represent question text and control logic
- ▶ Use notes and additional characteristics to represent metadata
- ▶ Represent response categories as value labels

Example: Response Categories

Hispanic Origin: Stata Code

```
* Generate vars
gen hispanic = .
...

lab def yesno 1 "Yes" 2 "No" 98 "Don't know" 99 "Refused"
...

char hispanic[qt] Do you consider yourself Hispanic or Latino?
lab val hispanic yesno
lab var hispanic "Hispanic/Latino ethnicity"
char hispanic[src] 1992 HRS
char hispanic[dom] sociodemog
note hispanic: 8/26/2003 - gpk
```

Example: Response Categories

Hispanic Origin: Human Readable Form

Q3. Do you consider yourself Hispanic or Latino?

hispanic:

Yes [1]

No [2]

Don't know [98]

Refused [99]

LABEL: Hispanic/Latino ethnicity
SOURCE: 1992 HRS
DOMAINS: sociodemog

NOTES: 8/26/2003 - gpk

Add comment

Example: Setup for Cross-Referencing

Code from the .do File

```
* Generate vars
gen dob = .
gen hispanic = .

...

* Assign question numbers for referencing
local i 1
foreach v of varlist _all {
    char `v'[qnum] `i'
    local ++i
}
```

Example: Control Logic – Cross-Referencing & Blocks

Past Experience in Paid Work: Stata Code

```
* Generate vars
gen pastpaid = .

...

char pastpaid[qt] Have you ever worked for pay?
lab var pastpaid "past experience in paid work"
char pastpaid[block] pastblk
char _dta[pastblk] ASK THESE QUESTIONS ONLY IF RESPONDENT/*
  */ ANSWERS "Yes" to 'disab[qnum]' or 'othemp[qnum]' /*
  */ OR IF RESPONDENT ANSWERS BOTH "Yes" to /*
  */ 'homemk[qnum]' AND "No" to 'curremp[qnum]'
char pastpaid[src] modification of 1992 HRS
char pastpaid[dom] sociodemog
note pastpaid: 9/1/2003 - gpk
```

Example: Control Logic – Cross-Referencing & Blocks

Past Experience in Paid Work: Human Readable Form

RE: QUESTION 27: ASK THESE QUESTIONS ONLY IF RESPONDENT ANSWERS "Yes" to 23 or 26 OR IF RESPONDENT ANSWERS BOTH "Yes" to 25 AND "No" to 21

[Add comment](#)

Q27. Have you ever worked for pay?

pastpaid:

Yes [1]
No [2]
Don't know [98]
Refused [99]

LABEL: past experience in paid work
SOURCE: modification of 1992 HRS
DOMAINS: sociodemog

NOTES: 9/1/2003 - gpk

[Add comment](#)

Example: Control Logic – Contingencies

Self-Described 'Born-Again' Experience: Stata Code

```
* Generate vars  
gen bornagain = .
```

```
...
```

```
char bornagain[qt] Would you say you have been "born again" or have /*  
*/ had a "born again" experience--that is, a turning point in your /*  
*/ life when you committed yourself to Christ?
```

```
lab var bornagain "self-described 'born-again' experience"  
char bornagain[c] relig 2 3 5  
char bornagain[src] NHSLS  
char bornagain[dom] sociodemog  
note bornagain: 8/26/2003 - gpk  
note bornagain: 9/26/2003 - omitted per Linda's e-mail 9/24/2003 - gpk  
note bornagain: 11/8/2003 - reinserted per Ed's request 11/3/2003
```


Example: Control Logic – Contingencies

Self-Described 'Born-Again' Experience: Human Readable Form

ASK QUESTIONS 7-8 ONLY IF RESPONDENT ANSWERED "Protestant", "Roman Catholic", OR "Christian" TO QUESTION 4

[Add comment](#)

Q7. Would you say you have been "born again" or have had a "born again" experience--that is, a turning point in your life when you committed yourself to Christ?

bornagain:

Yes [1]
No [2]
Don't know [98]
Refused [99]

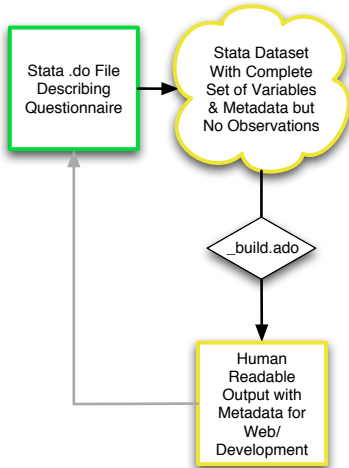
LABEL: self-described 'born-again' experience
SOURCE: NHSLs
DOMAINS: sociodemog

NOTES: 8/26/2003 - gpk
9/26/2003 - omitted per Linda's e-mail 9/24/2003 - gpk
11/8/2003 - reinserted per Ed's request 11/3/2003

[Add comment](#)

A Better Alternative

Fixing Problems with a Systematic Process



Brief Review

Create Questionnaires as Stata .do File (or Files)

- ▶ Represent questions as variables and response categories as value labels
- ▶ Use characteristics to represent question text and control logic such as contingencies, loops and blocks
- ▶ Use notes and additional characteristics to represent metadata
- ▶ .ado file available to generate web-accessible, human readable version of questionnaire
- ▶ Finished .do file creates a Stata Dataset with complete set of variables and metadata but no observations

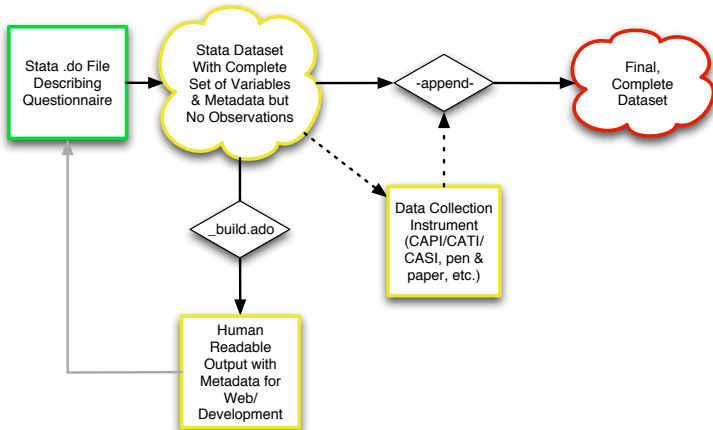
Advantages Over Traditional Methods

- ▶ Editing of questions is much more systematic and efficient
- ▶ Easy to use – proven useable on a major study with minimal training
- ▶ Flexible and powerful development and collaboration
- ▶ Questionnaire easily transfered to other systems and formats
- ▶ Preserves metadata along entire course of development and analysis

Advantages Over Traditional Methods

(Continued)

- ▶ Creates a ready-made documented dataset to append collected data onto



Future Plans

