

# Paradata within the Total Survey Error Framework Successes, Challenges, and Gaps

Frauke Kreuter

Federal Computer-Assisted Survey Information Collection (FedCASIC) Workshop

March 27, 2012

# Summary

- We have seen successful uses of paradata to **gain efficiency**, and to **alert for errors**.
- We face serious challenges to expand the **concurrent** analytic use of paradata, the **tailored** collection of paradata, and the transfer **across** modes, surveys, and survey organizations.
- We might benefit from widening the scope to **other error** sources, through linkage with **cost** data and others, and from the use of paradata in **modelling**.

# Outline ...

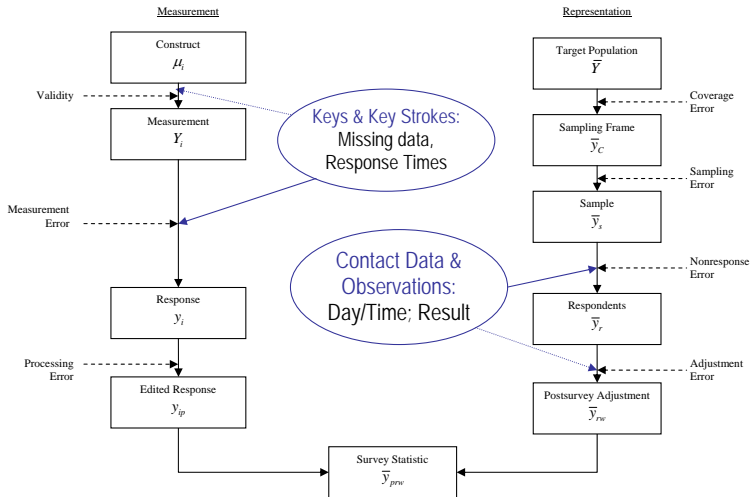
- 1 Current Activities and Typical Applications
- 2 Challenges in Collection, Analysis, and Communication
- 3 Paradata inside the Total Survey Error Framework

# Outline

- 1 **Current Activities and Typical Applications**
- 2 Challenges in Collection, Analysis, and Communication
- 3 Paradata inside the Total Survey Error Framework

# Prevalent Paradata in TSE Framework

Groves et al. 2004



# Response Time

- Substantive Use

- Attitudes as object-evaluation model

Fazio et al. 1986, Dovidio & Fazio 1992.

- Post-hoc Use - Focus on Error

- Characteristics of Instrument and Setting:

(+) poor wording, poor layout, length, complexity

(-) logical order, practice, correct answers, decreasing motivation

e.g. Bassili 1996, Draisma & Dijkstra 2004, Tourangeau, Couper & Conrad 2004, Yan & Tourangeau 2008

- Interview administration

e.g. Olson & Peytchev 2007, Couper & Kreuter 2012, Schafer 2012

Interview falsification

Clements 2001; Penne, Snodgrass & Baker 2002

- Concurrent Use - Focus on Error

- Intervention if respondents answer too fast Conrad et al. 2009

or too slow Conrad, Schober & Coiner 2007

# Call record data

- **Post-hoc Use** - Focus on Efficiency

- Optimal call schedules **Example**

e.g. Weeks et al. 1980, Greenberg & Stokes 1990, Bates 2003, Laflamme 2008, Durrant et al. 2010

- Predictors of response e.g. Campanelli et al. 1997, Groves & Couper 1998, Lynn 2003, Bates & Piani 2005, Bates et al. 2008, Durrant & Steele 2009

- **Concurrent Use** - Focus on Efficiency

- Call scheduling (CATI)
- Monitoring **Example**

- **Post-hoc Use** - Focus on Error

- Nonresponse bias analyses **Example**

e.g. FedStat Surveys - since OMB Standard and Guidelines 2006

- Nonresponse bias adjustment Politz & Simmons 1949, Kalton 1983, Beaumont 2005, Biemer & Link 2006

- **Concurrent Use** - Focus on Error

- Interventions **Example**

# Outline

- 1 Current Activities and Typical Applications
- 2 Challenges in Collection, Analysis, and Communication**
- 3 Paradata inside the Total Survey Error Framework



## Now one could ask:

- Why do we not see more research on key stroke data?  
⇒
- Why do analysts struggle with f2f contact protocol data?  
⇒
- Why do FRs shy away from call record protocols?  
⇒
- Why are interviewer observations not used for adjustment?  
⇒
- Why are adjusters disappointed about interviewer observations?  
⇒

## And one could suggest:

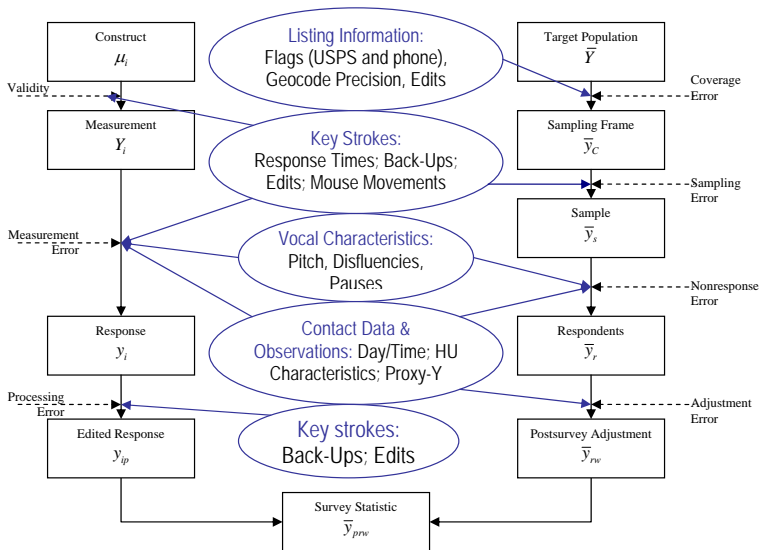
- Why do we not see more research on key stroke data?  
⇒ **Code repository shared across surveys and organizations**
- Why do analysts struggle with f2f contact protocol data?  
⇒ **Statistical methods development**
- Why do FRs shy away from call record protocols?  
⇒ **“There’s an app for that”**
- Why are interviewer observations not used for adjustment?  
⇒ **Flexible software to allow tailored indicators**
- Why are adjusters disappointed about interviewer observations?  
⇒ **Use of auxiliary information**

# Outline

- 1 Current Activities and Typical Applications
- 2 Challenges in Collection, Analysis, and Communication
- 3 Paradata inside the Total Survey Error Framework**

# Paradata inside TSE Framework

Groves et al. 2004



# I am hopeful

- U.S. Census Bureau - Survey Analytics
- NORC & SRO research on GIS to improve field efficiency
- Westat research on display and communication
- SRO research on quality of interviewer observations
- RTI research on quality of contact data
- LMU research on modelling
- ...
- You ...

Thank You!

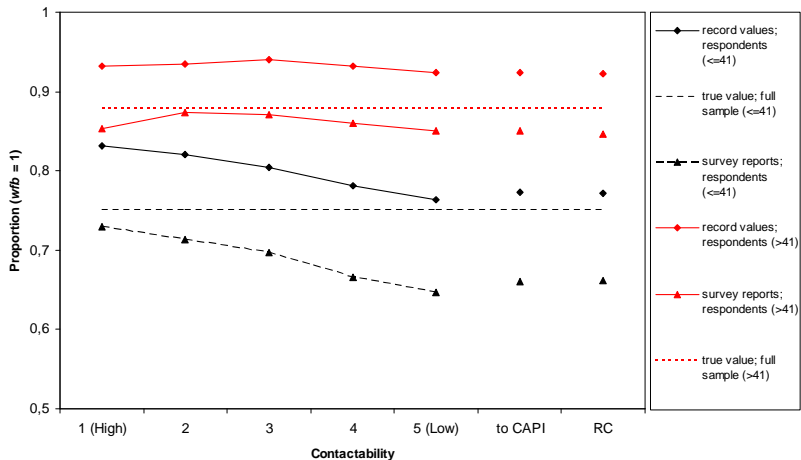
fkreuter@survey.umd.edu

# Contact Rates by Hour and Day in NSFG Wagner 2012

hour	Sunday		Monday		Tuesday		Wednesday		Thursday		Friday		Saturday	
	All	Elig	All	Elig	All	Elig	All	Elig	All	Elig	All	Elig	All	Elig
9	0.30	0.34	0.25	0.35	0.30	0.32	0.24	0.31	0.23	0.30	0.27	0.33	0.30	0.35
10	0.32	0.40	0.31	0.38	0.28	0.33	0.29	0.34	0.30	0.36	0.27	0.34	0.31	0.39
11	0.36	0.43	0.30	0.38	0.31	0.38	0.31	0.39	0.31	0.39	0.32	0.40	0.35	0.43
12	0.37	0.44	0.32	0.42	0.32	0.38	0.32	0.40	0.30	0.37	0.31	0.38	0.34	0.42
13	0.37	0.45	0.32	0.42	0.24	0.31	0.29	0.38	0.30	0.38	0.32	0.39	0.34	0.43
14	0.38	0.46	0.34	0.43	0.33	0.40	0.32	0.40	0.32	0.39	0.33	0.40	0.35	0.43
15	0.39	0.48	0.35	0.44	0.32	0.40	0.33	0.42	0.33	0.41	0.33	0.41	0.36	0.46
16	0.39	0.49	0.36	0.45	0.37	0.46	0.36	0.45	0.35	0.43	0.34	0.42	0.35	0.45
17	0.39	0.49	0.40	0.49	0.38	0.46	0.38	0.47	0.36	0.46	0.34	0.43	0.33	0.43
18	0.37	0.44	0.38	0.47	0.39	0.48	0.37	0.47	0.36	0.45	0.33	0.42	0.35	0.44
19	0.37	0.44	0.39	0.47	0.37	0.45	0.37	0.46	0.35	0.44	0.31	0.42	0.35	0.43
20	0.40	0.44	0.38	0.45	0.39	0.45	0.38	0.46	0.37	0.45	0.32	0.40	0.36	0.44

Back

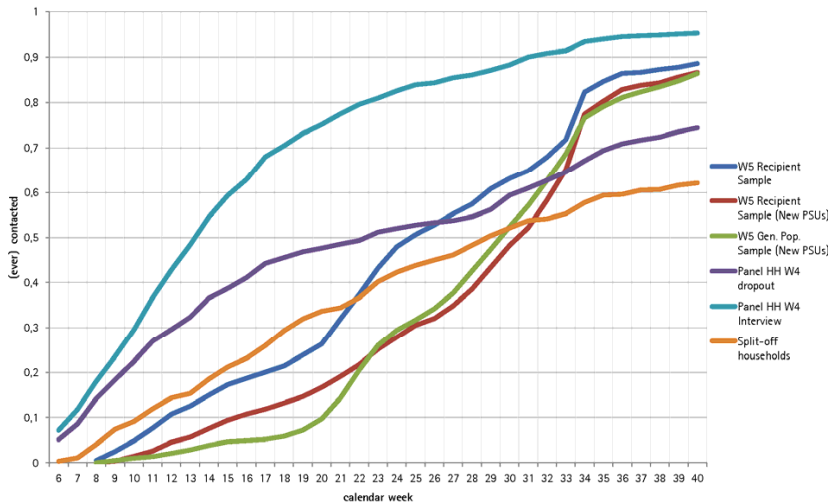
# Nonresponse Bias in PASS Kreuter, Mueller, Trappmann 2010



... though often no comparison [Back](#)

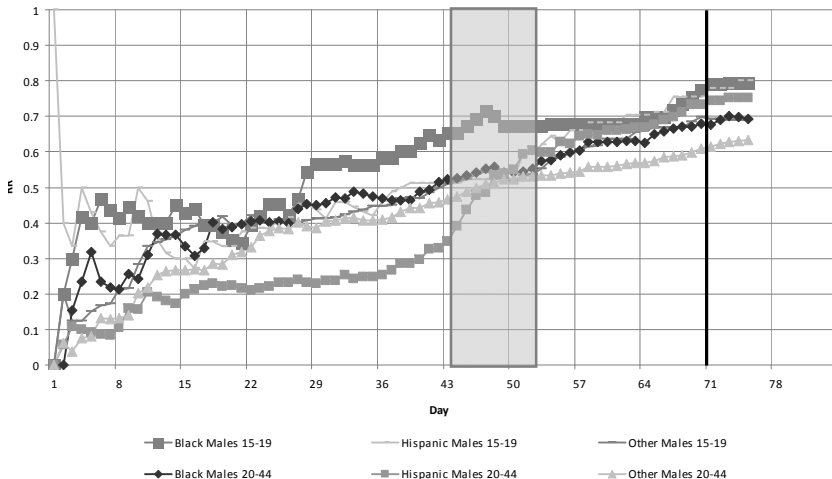


# Monitoring Effort in PASS Wave 5 Mueller 2011



Back

# Response Rates by Subgroup in NSFG Lepkowski et al. 2012



Back

# Automatically Generated Data

**Figure 1: Excerpt from an audit trail file**

```
"1/11/2004 9:15:50 AM", "Enter Form:1", "Key:XXXXXXXX "
"1/11/2004 9:15:50 AM", "Metafile name:C:\WINCM\DATA\STUDIES\CEQ_BA01\e-inst\inst.bmi"
"1/11/2004 9:15:50 AM", "Metafile timestamp:Wednesday, December 03, 2003 8:47:42 AM"
"1/11/2004 9:15:50 AM", "WinUserName:FR"
"1/11/2004 9:15:50 AM", "Enter Field:Front.Start", "Status:Normal", "Value:"
"1/11/2004 9:16:13 AM", "Leave Field:Front.Start", "Cause:Next Field", "Status:Normal", "Va:
.....
"2/11/2004 5:52:21 PM", "Enter Field:Sect03.ANYRENT", "Status:Normal", "Value:"
"2/11/2004 5:52:24 PM", "Leave Field:Sect03.ANYRENT", "Cause:Next Field", "Status:Normal", '
.....
"1/11/2004 6:16:42 PM", "Enter Field:Back.Appt.verify_info", "Status:Normal", "Value:"
"1/11/2004 6:16:43 PM", "Leave Field:Back.Appt.verify_info", "Cause:Next Field", "Status:N
"1/11/2004 6:16:44 AM", "Leave Form:1", "Key:XXXXXXXX "
```

Back

# ESS ROC data – manual entries

... data are incomplete and error prone

	CntryC <sup>e</sup>	visit	result	fromESS
31544.	30101118	1	NO contact at all	.
31545.	30101118	2	NO contact at all	.
31546.	30101118	3	NO contact at all	.
31547.	30101118	4	NO contact at all	.
31548.	30101118	5	NO contact at all	.
31549.	30101118	6	NO contact at all	.
31550.	30101118	7	Contact with R NO interview	.
31551.	30101118	8	NO contact at all	.
31552.	30101118	9	NO contact at all	.
31553.	30101118	10	NO contact at all	.
31554.	30101118	.	Interview	1

... times are all but randomized

... interviewer have their own preferences

Back

# World Health Survey 2002 – paper entry

## 0350. Contact record

Number of calls	S0350 CALL #1			S0351 CALL #2			S0352 CALL #3		
A. Date (day / month / year)	_ / _ / _			_ / _ / _			_ / _ / _		
B. Day of week									
C. Exact time began									
D. Interviewer I.D.									
E. Contact with	Respondent 1	Informant 2	No One 3	Respondent 1	Informant 2	No One 3	Respondent 1	Informant 2	No 3
F. Mode of contact	Personal 1		Telephone 2	Personal 1		Telephone 2	Personal 1		Telephone 2
G. Tel. Number if obtained									
H. Household Unit listing obtained	Yes 1	No 5		Yes 1	No 5		Yes 1	No 5	
I. Detailed description of contact or attempt to contact									
J. Result code									
<hr/>									
Number of calls	S0355 CALL #6			S0356 CALL #7			S0357 CALL #8		
A. Date (day / month / year)	_ / _ / _			_ / _ / _			_ / _ / _		
B. Day of week									
C. Exact time began									
D. Interviewer I.D.									
E. Contact with	Respondent 1	Informant 2	No One 3	Respondent 1	Informant 2	No One 3	Respondent 1	Informant 2	No 3
F. Mode of contact	Personal 1		Telephone 2	Personal 1		Telephone 2	Personal 1		Telephone 2

# CHI - Census Bureau

Contact History Instrument v5.8.5 Created 08/25/2004

Forms Answer Navigate Options Help

CHI

**• CONTACT STRATEGIES ATTEMPTED**

- Select the categories that describe the strategies used on this contact attempt.
- Enter all that apply, separate with commas.

<input type="checkbox"/> 1. Advance letter given	<input type="checkbox"/> 14. Contacted property manager
<input type="checkbox"/> 2. Scheduled appointment	<input type="checkbox"/> 15. Visited county assessor / post office / permit office
<input type="checkbox"/> 3. Left note / appointment card	<input type="checkbox"/> 16. On-line tracking database
<input type="checkbox"/> 4. Left promotional packet / informational brochure	<input type="checkbox"/> 17. Sought help from SFR / RO
<input type="checkbox"/> 5. Called household	<input type="checkbox"/> 18. Reassignment
<input type="checkbox"/> 6. Left message on answering machine	<input type="checkbox"/> 19. Offered incentive
<input type="checkbox"/> 7. FR will request No One Home Letter	<input type="checkbox"/> 20. CED double placement
<input type="checkbox"/> 8. FR will request Refusal Letter	<input type="checkbox"/> 21. Used MAF or ALMI
<input type="checkbox"/> 9. FR will request Better Understanding Letter	<input type="checkbox"/> 22. None
<input type="checkbox"/> 10. Called contact persons	<input type="checkbox"/> 23. Other - specify
<input type="checkbox"/> 11. Stake-out	
<input type="checkbox"/> 12. Checked with neighbors	
<input type="checkbox"/> 13. Contacted other family members	

Strategies attempted

00000001 STRATEG5 6/26/2006 12:07:04 PM Monday CTRL NUM : 123456789012345678901234

Back



<u>Variable set</u>	<u>Correlation with participation</u>	<u>Correlation with survey variables (absolute values)</u>	
		<u>Average</u>	<u>Maximum</u>
<b>Cooperation</b>			
Factor 1: Time concerns	-.25	.03	.09
Factor 2: Privacy or content concerns	-.27	.02	.09
Factor 3: General resistance	-.47	.02	.12
Factor 4: Gatekeeper issues	-.30	.02	.08

Back

# Measurement Error: Gender McCulloch & Kreuter 2012

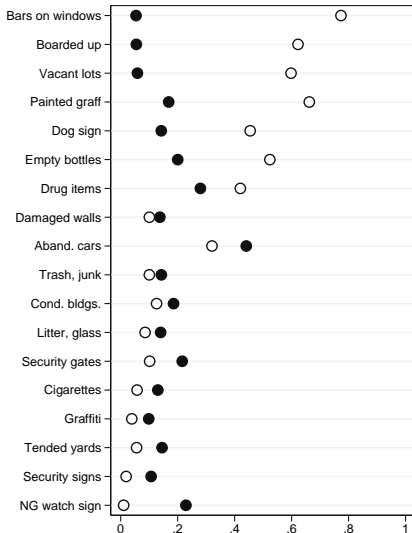
Pooled data: 28 CATI surveys (n=25,635), Marist College (MIPO):

	Respondent male	Respondent female	Total
Guess - male	97.36	13.87	49.63
Guess - female	2.64	86.13	50.37
Total	100.00	100.00	100.00

	White	Afr. Am.	Hispanic	Asian	Other
Gessed Correct	92.1%	87.4%	92.1%	92.0%	92.7%
Gessed Incorrect	7.9%	12.6%	7.9%	8.0%	7.3%



# $\rho$ Interviewer vs. $\rho$ Census Tract Casas-Cordero et al. 2012



# Measurement Error: Young Children West & Kreuter 2012

NSFG Interviewer Observations from 16 Quarters (n=15,044):

