PANDA: Using Paradata to Improve Data Quality

Ari Teichman
Demographic Surveys Division
U.S. Census Bureau

Overview of Presentation

- PANDA Background
- Basic American Housing Survey (AHS) PANDA Functionalities
- User Feedback
  - 2007 AHS PANDA Users Debriefing
  - Enhancements for 2009
- Using 2007 National Averages as a Benchmark for AHS PANDA data
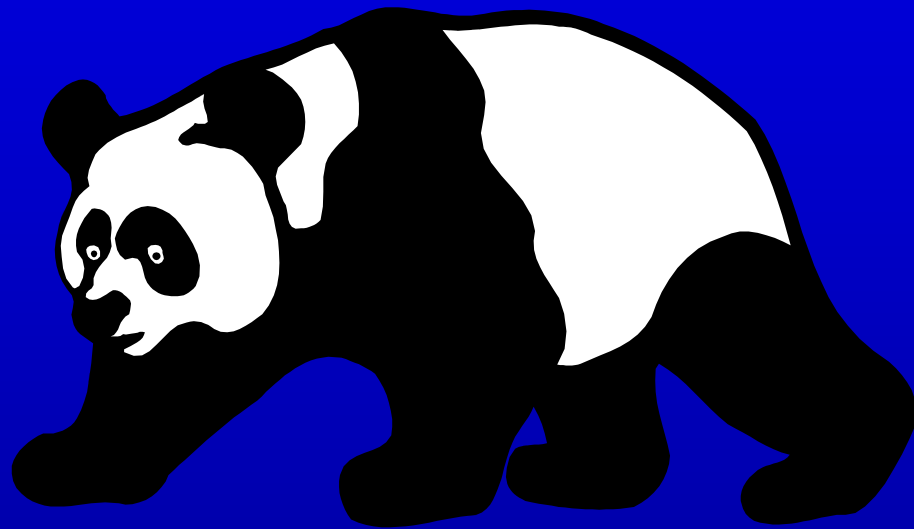- Conclusion/Questions

Performance

A

Nd

Data

Analysis tool

Acronyms in my Presentation

- DSD – Demographic Surveys Division
- AHS – American Housing Survey
- RO – Regional Office
- HQ - Headquarters
- FR – Field Representative
- PANDA – Performance and Data Analysis Tool
- CAPI – Computer Assisted Personal Interview

Development of PANDA

- Developed for the American Housing Survey
  - Long data collection period (approx. 4 months)
  - Data transmitted on a flow basis throughout data collection

- Transition from paper to CAPI
  - Completed cases transmitted directly, bypassing supervisors
  - Processing system redesigned to process data on flow basis immediately as it comes into Census HQ
  - Needed a tool to measure quality of data and identify potential problems in the data in a timely manner.

Purpose of PANDA

- Provide data at national, regional, FR, and individual case levels
- Provide an early warning of possible problems
  - FR difficulty with concepts
  - Falsification
- Provide up-to-date reports and tables
  - Enables ROs to address problems immediately

AHS PANDA Report Components

- Vacancy rates
  - High percentage of vacant interviews could mean FR does not understand when to conduct a vacant interview **OR**
  - FR could be falsifying and conducting a vacant interview to save themselves the time and trouble of conducting a regular occupied interview
- Average household size
  - Smaller household could indicate falsification (i.e. easier to conduct an interview with smaller household)

AHS PANDA Report Components

- Time of day interview started
  - Abnormal time of day interview started (i.e. 3:00 AM) might indicate falsification
- Minutes per case
  - Short interview might indicate falsification
- And More!

PANDA Training

- Video Teleconference with Regional Offices
- Help Screens Provided in PANDA
- Continual Technical Support from Headquarters

P@nda

home | reports | hq reports | trace files | search | settings | users | news | help | logout

Welcome Ari Teichman

February 15, 2007

Welcome to the American Housing Survey **Performance and Data Analysis** (PANDA) tool.

Use the navigation bar at the top of the page to navigate to the different sections of PANDA. The reports section will display several reports. The trace files section will allow you to retrieve a trace file either by control number or case id. The search section will allow you to query the database for specific criteria. The settings section will allow you to change your password and default settings. And last but not least, the help section will display information on how to use PANDA.

Data is available for cases received by October 02, 2007.

Your last login was on February 06, 2009 at 09:57:57 AM.

You have logged in 1 times this week and 29 times overall.

There are currently 2 users online.

If you need an account, please contact:

Laurie Jarboe
Clinton Borland
Medell Ford
@301-763-2185 Field

If you have problems with the system, please contact:

Ari Teichman
@301-763-5425 HSB

Diana Rubin
@301-763-5366 HSPB

Panda Version 1.2.8

Done   Trusted sites

Start | Novell-delivered Ap... | New Memo - Lotus ... | FEDCASIC | Microsoft PowerPoin... | Panda - Microsoft...   3:11 PM

Main Tabs in AHS PANDA

- Home
- Reports
- HQ Reports
- Trace Files
- Search
- Settings
- Help
- Logout

USCENSUSBUREAU

P@nda

Reports for AHS Total

Summary Reports

- RO/Area Totals
- Cumulative/Weekly Report
- Average Cases Per FR

- Special Living Sample RO/Area Totals
- Special Living Sample Cumulative/Weekly Report

- Cases Changed to Type A in Processing

Top 15 FR Reports*

- Highest Nonresponse Rate for Salary of Reference Person
- Highest Number of Regular Occupied Interviews Completed in Less Than 20 Minutes
- Highest Number of Cases Completed 12:00 a.m.-7:59 a.m.
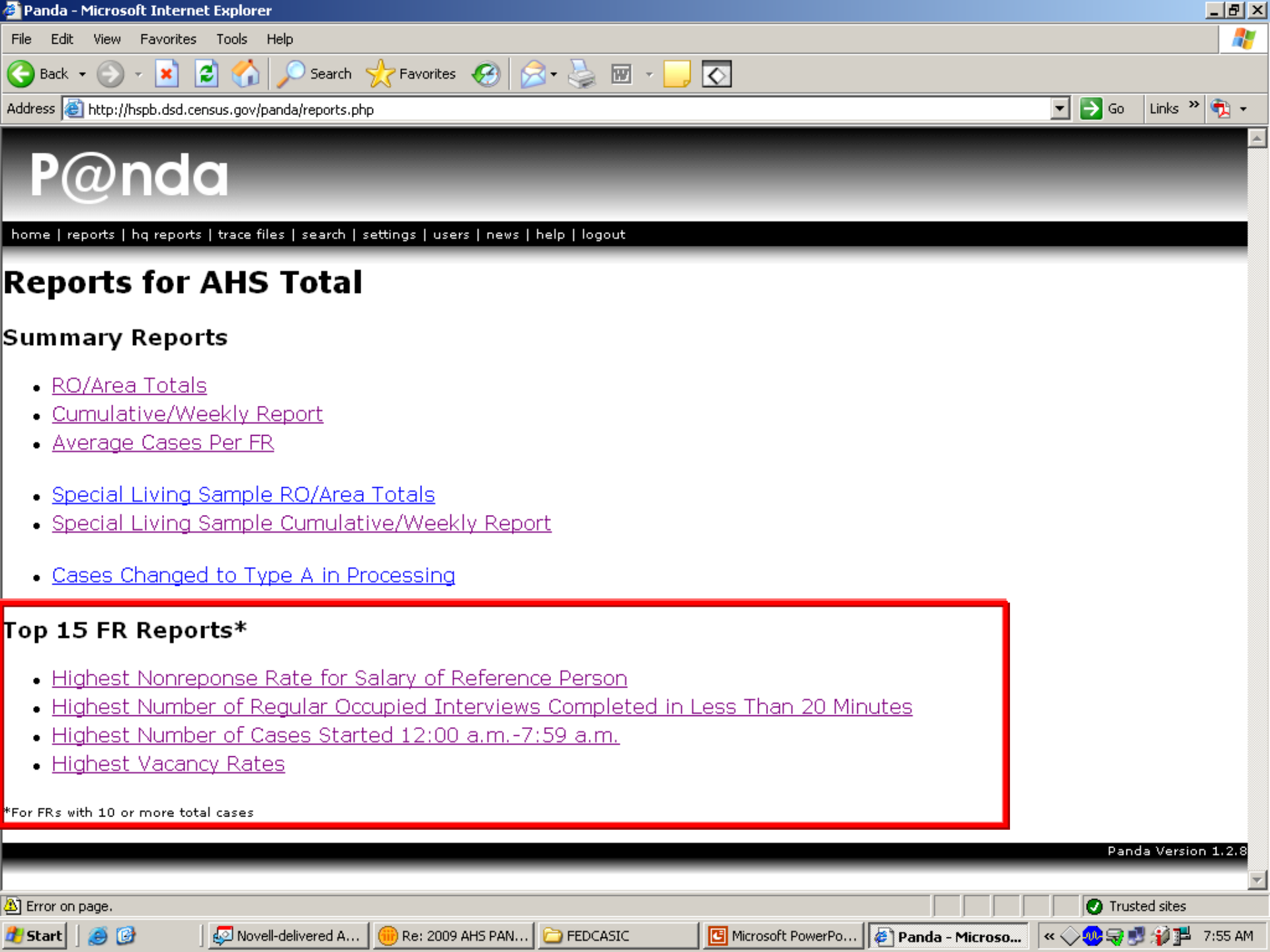- Highest Vacancy Rates

*For FRs with 10 or more total cases

P@nda

Cumulative/Weekly Report

cumulative | **indicator variables** | **minutes per case** | **time of day** | print | reports main

2005 AHS-N Weekly Report for Detroit (24) - Cases Received By: 06/03/2003
Most Recent Avg HH Size For RO = 2.46

Statistics by	Total	#PV	% Repl. HH	Type A	Type B	Type C	URE Int.	Par. Int.	Vac. Rate	Resp. Rate	Salary of ref. per.	Pres. of family int. inc.	Avg. HH Size for Reg. Occ. Int.	pie	reg. occ. <14	vac/ ure	reg. occ. 14-19	vac/ ure	reg. occ. 20-29	vac/ ure	reg. occ. 30-39	vac/ ure	reg. occ. 40-49	vac/ ure	reg. occ. 50+	vac/ ure	pie	a.m. 12:00-7:59	8:00-11:59	p.m. 12:00-5:59	6:00-9:59	10:00-11:59
AHS Total	1203	264	7.34	0.08	1.50	0.50	0.42	0.08	7.46	99.92	12.84	3.30	2.38	[pie]	39	24	39	25	177	24	301	8	262	2	272	5	[pie]	2	192	618	385	5
Detroit (24)	19	2	5.26	0.00	0.00	0.00	0.00	0.00	0.00	100.00	31.58	10.53	2.47	[pie]	1	0	2	0	8	0	5	0	1	0	2	0	[pie]	0	5	8	6	0
	1	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	1.00	[pie]	0	0	0	0	1	0	0	0	0	0	0	0	[pie]	0	1	0	0	0
	6	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	2.33	[pie]	0	0	2	0	4	0	0	0	0	0	0	0	[pie]	0	2	3	1	0
	4	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	75.00	0.00	4.25	[pie]	1	0	0	0	0	0	1	0	1	0	1	0	[pie]	0	2	1	1	0
	2	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	50.00	50.00	2.00	[pie]	0	0	0	0	0	0	2	0	0	0	0	0	[pie]	0	0	1	1	0
	1	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	2.00	[pie]	0	0	0	0	1	0	0	0	0	0	0	0	[pie]	0	1	0	0	0
	1	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	100.00	100.00	2.00	[pie]	0	0	0	0	0	0	1	0	0	0	0	0	[pie]	0	0	0	1	0
	4	0	25.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	25.00	0.00	1.75	[pie]	0	0	0	0	2	0	1	0	0	0	1	0	[pie]	0	0	2	2	0

Cumulative/Weekly Report

2005 AHS-N Weekly Report for Detroit (24) - Cases Received By: 06/03/20
Most Recent Avg HH Size For RO = 2.46

Individual cases can be displayed for each FR.

Statistics by		Total	#PV	% Repl. HH	Type A	Type B	Type C	URE Int.	Par. Int.	Vac. Rate	Resp. Rate	Salary of ref. per.	Pres. of family int. inc.
												Salary of ref. per. (%DK, Ref, or No Data for Reg. Occ. Interviews)	Pres. of family int. inc.
AHS Total		1203	264	7.34	0.08	1.50	0.50	0.42	0.08	7.46	99.92	12.84	3.30
Detroit (24)		19	2	5.26	0.00	0.00	0.00	0.00	0.00	0.00	100.00	31.58	10.53
		1	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00
		6	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00
[tra] [rpt]		1	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00
[tra] [rpt]		1	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00
[tra] [rpt]		1	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00
[tra] [rpt]		1	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00
[tra] [rpt]		1	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00
[tra] [rpt]		1	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00

P@nda

home | reports | hq reports | trace files | search | settings | users | news | help | logout

Reports for AHS Total

Summary Reports

- RO/Area Totals
- Cumulative/Weekly Report
- Average Cases Per FR

- Special Living Sample RO/Area Totals
- Special Living Sample Cumulative/Weekly Report

- Cases Changed to Type A in Processing

Top 15 FR Reports*

- Highest Nonreponse Rate for Salary of Reference Person
- Highest Number of Regular Occupied Interviews Completed in Less Than 20 Minutes
- Highest Number of Cases Started 12:00 a.m.-7:59 a.m.
- Highest Vacancy Rates

*For FRs with 10 or more total cases

Panda Version 1.2.8

P@nda

Trace Files

Download by Control Number

Control Number: [   ] [Download]

Download by Case ID

Case Id: [██████████] [Download]

P@nda

Search

Use the form below to query the data for cases that match the criteria you enter. To query the data, select the criteria using the drop down boxes provided. To search an entire category, leave the criteria blank. For instance, if you want to search for all cases that were done in less than 20 minutes. Select "Less Than 20 Minutes" from the minutes per case field, and leave the rest of the fields blank.

Multiple Criteria

RO Code: [▼]

FR Code: [ ]

Minutes Per Case: [▼]

Time of Day Started: [▼]

Salary of Reference Person: [▼]

Interview Type: [▼]

Special Living Flag [ ]

[Clear] [Search]

Control Number / Case ID

Control Number: [ ]

Case ID: [ ]

[Clear] [Search]

User Feedback

- Held a debriefing for 2007 AHS PANDA users
- Asked a variety of questions of the users to get their feedback
- What did Census regional offices like and dislike about the 2007 PANDA system?

User Feedback
2007 AHS PANDA Users Debriefing

- What effect, if any, does PANDA have on the management decisions you make regarding AHS?
  - A great deal in terms of reassigning cases and placing FRs in supplemental reinterview
  - Increased access to information leads to easier decision making
  - The tool provides relevant, timely info to better monitor progress and performance and be aware of problems

User Feedback
2007 AHS PANDA Users Debriefing

- What is most useful about PANDA?
  - Identifying cases started between midnight and 7:59am, length of interview, and other outliers
  - Targeting potential falsification
  - Ability to perform searches and produce trace files

User Feedback
2007 AHS PANDA Users Debriefing

- How do you research/address the findings in the messages from Field?
  - Use PANDA searches for detailed info on FRs work, coupled with local knowledge of assignment areas
  - Managers address potential problems with FRs through appropriate mgmt. structure
  - FRs are retrained and/or placed in reinterview

USCENSUSBUREAU

User Feedback
2007 AHS PANDA Users Debriefing

- What do you find least useful about PANDA?
    - Micromanaging
    - Incorrect results to searches
    - Focus on missing response to one field when other data has been successfully collected

Enhancements for 2009

- Address 2007 programming bugs (i.e. more accurate 'cumulative time of interview' measurement)

- Better trace file access

- Individual FR reports available on own page

- Confidentiality statements throughout PANDA system to reinforce data stewardship

- Enhancement to the visual design of reports (i.e. working on trying to have static column headers or color coded columns to make reports easier to read – Print to Microsoft Excel?)

USCENSUSBUREAU

Using 2007 National Averages as a Benchmark for PANDA Data

- Can use final 2007 data as a benchmark when looking at 2009 PANDA data

- For example, can look at:
  - Length of interview
  - Average household size
  - Time of day interview started
  - Item nonresponse rates
  - % of vacant interviews

Using 2007 National Averages as a Benchmark for PANDA Data

- In 2007, on average, it took:
  - 56 minutes to complete a regular occupied interview
  - 30 minutes to complete UREs (Usual Residence Elsewhere) and Vacants
- Compare to PANDA distributions to detect potential length of interview issues

Using 2007 National Averages as a Benchmark for PANDA Data

- In 2007, the average household size for regular occupied interviews was approximately 3.03 people
- Compare to PANDA household sizes to detect potential household size issues

Using 2007 National Averages as a Benchmark for PANDA Data

- In 2007, on average, interviews were started between the following times:

Time of Day	% of 2007 cases started
12:00AM - 7:59AM	0.76%
8:00AM - 11:59AM	18.94%
12:00PM - 5:59PM	48.43%
6:00PM - 9:59PM	30.67%
10:00PM - 11:59PM	1.20%

- Can compare this national distribution to PANDA distributions to look for anomalies

USCENSUSBUREAU

Using 2007 National Averages as a Benchmark for PANDA Data

- In 2007, on average, 18.65% of regular occupied cases had either a "don't know", "refused" or "missing" response for the salary of the reference person

- Can compare this national item nonresponse rate with PANDA salary item nonresponse rate

Using 2007 National Averages as a Benchmark for PANDA Data

- In 2007, on average, we conducted vacant interviews for 11.99% of cases

- Can compare this average to PANDA percentage to determine whether appropriate percentage of vacant interviews are being conducted

The Future of AHS PANDA

- Expanding to other DSD surveys
  - The National Health Interview Survey (NHIS) is using PANDA.
    - NHIS uses functionalities like a case-level time report, outlier flags and a multiple indicator report.
    - These functionalities may very well be incorporated into AHS PANDA as well.
- Always considering other reports and enhancements

Conclusion

- Very useful tool to monitor survey progress

- Can use 2007 national averages as a benchmark for 2009 PANDA data

- Further reports/enhancements are being looked into for the future

US CENSUS BUREAU

Acknowledgements

Census Bureau Colleagues
- Chris Stringer
- Thom Iseman
- Diana Rubin
- Laurie Jarboe
- Rick Downs
- Yiwen Hastings

Contact Information

If you would like a copy of my presentation, feel free to e-mail me at:

Ariel.L.Teichman@census.gov

USCENSUSBUREAU

Questions?