# For Clueless Consumers, Can Records Improve Data Quality?

65th Annual Conference of the AAPOR

May 16, 2010

Chicago, IL

Eileen O'Brien

# Overview

- How EIA uses records differently. Why?
- Benefits and challenges of maximal records use
- Enhancing value and utility of data from records
- Conclusions, recommendations and future work

# EIA Uses Energy Bills in its Residential Energy Consumption Survey (RECS)

Design Features of RECS:

- Purpose: Identify factors driving U.S. energy demand
- Cross-sectional, quadrennial study conducted under CIPSEA 2002
- Mode: in-person interviews of ≈ 5,000 households*
- Sample Population: No oversampling of special groups
- Interview length ≈ 45 minutes (higher in 2009)
- Response rate ≈ 80 percent

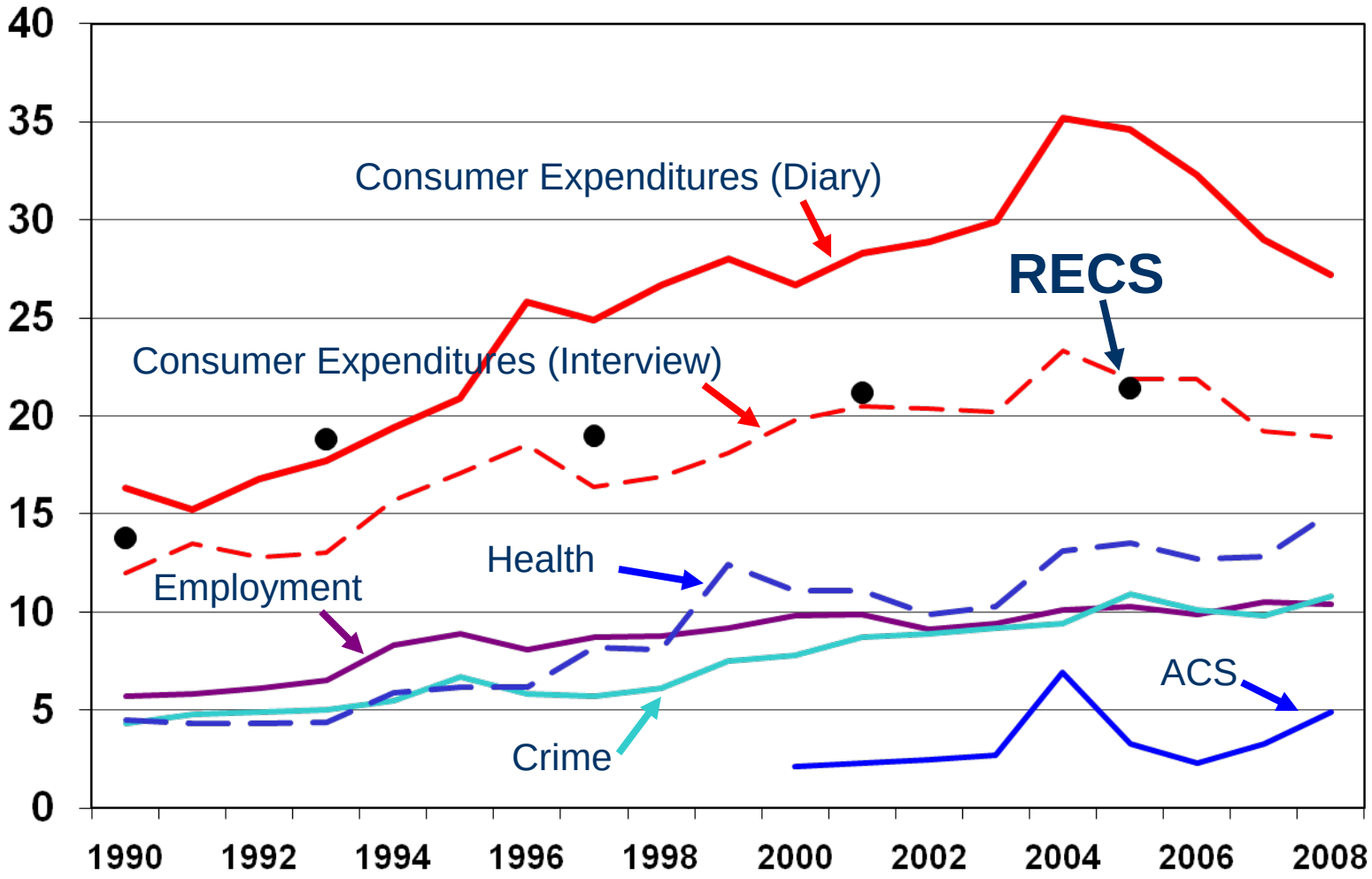
** RECS 2009 will interview about 15,000 households*

# Three principal methods we ***could*** use to collect energy data at the household level

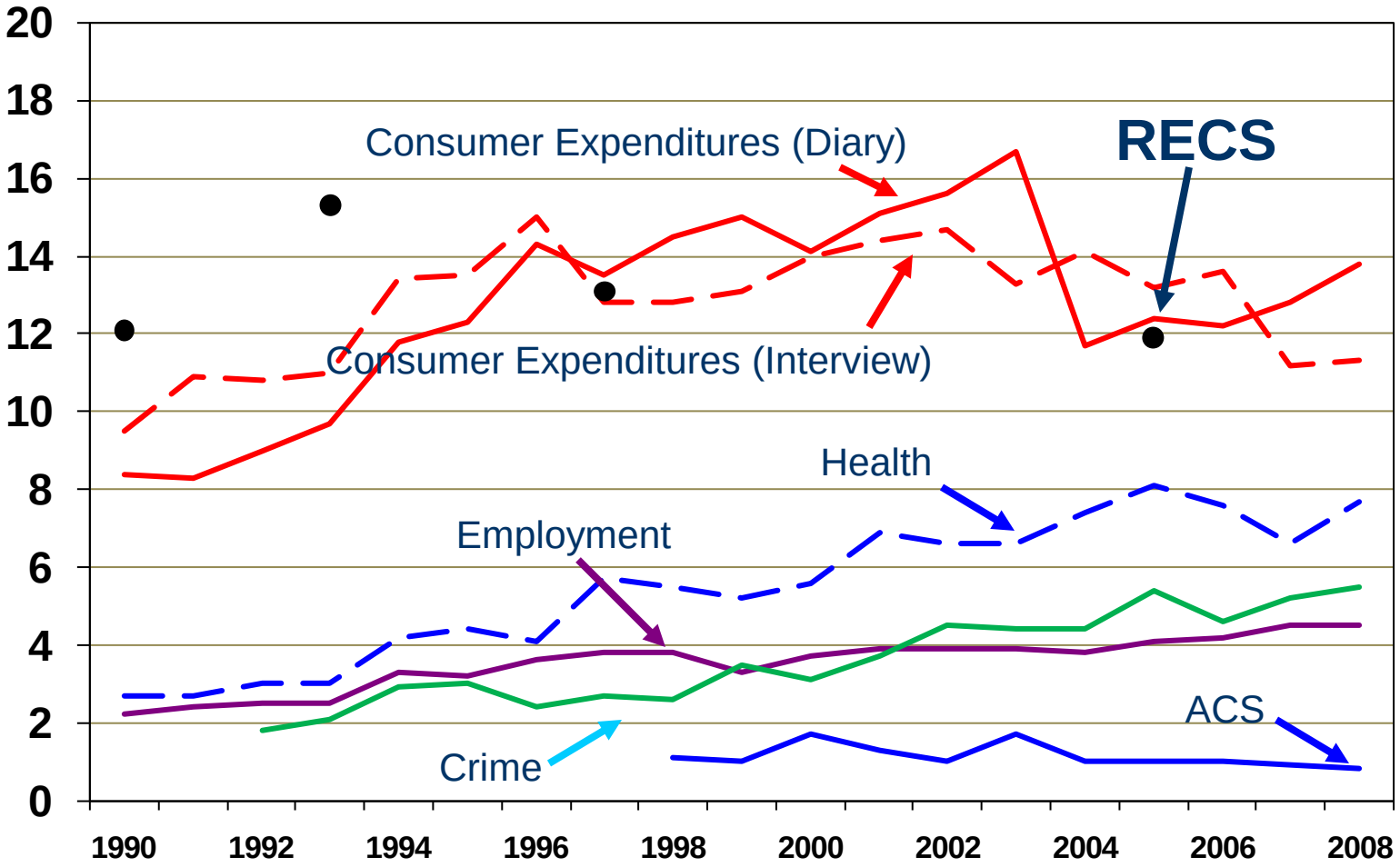
1. Have respondents use recall or rely on their bills in survey interviews (similar to American Housing Survey)
2. Scan/copy respondents' bills and synthesize whatever data we capture from them
3. Conduct two phase data collection (current strategy)– have energy companies report these data *FOR* respondents in a follow-on survey ("RECS Phase II")

# Option 1: Rely on respondents to use memory and records?

# RECS' nonresponse rates higher than federal household surveys with similar burden



# RECS' refusal rates about double the rate in comparable federal surveys



# Having respondents use records in the RECS interview would reduce overall data quality

- Interview length and burden would increase, affecting effort and response quality across the interview
- Unit nonresponse would increase further
  - A quadrennial survey done under contract uses many more novice interviewers; typically, they have more difficulty encouraging survey participation and effort

# Features of Energy Bills call for some expertise

Bills show--

- Special charges that don't apply to RECS
- Varying names for charges that are included
- Multiple bills within a bill
- Adjustments to previous/current charges in current bill
- Having two non-experts (interviewer and respondent) use complex bills would decrease response accuracy

Using bills requires higher levels of functional literacy and cognitive ability to extract and record requested information onto a standardized form

# Respondent limitations and data quality drive our data collection choice

- Reference period is too long for self reports (12-16 mos.)
- Energy bills are too complex for respondents to use
  - Most attend to amount due and ignore consumption information
- Respondents don't understand bill content (Payne, 2000)
  - Efforts to improve bill comprehension are mixed even for savvier respondents (Payne, 1996)
- Less and less interaction with bills and payment
  - Electronic delivery of energy bills: less use of bills and encoding
  - More automatic bill payments: limited knowledge of bill content

The quality of self records-use and self-reports varies. Comprehension, capacity for recall is limited when not encoded. The structure of response error changes.

# Option 2: Use respondent records alone?

- In this study we have scanned copies of bills
- Why not rely on them entirely?

# Bill content varies, is “incomplete” and limits use for annual estimation

| Account History * n=28 | Frequency | Percent |
|---|---|---|
| No account history on entire bill | 8 | 29% |
| Table of current month to previous year | 6 | 21% |
| Table of current to previous month, year | 2 | 7% |
| Consumption amount per month (12 mos.) | 1 | 3% |
| Bar graphs ONLY (12-14 mos.) | 8 | 28% |
| Bar graphs AND table data (13 mos.) | 2 | 7% |
| Two bar graphs for past 13 months – consumption and effective generation | 1 | 3% |

* Sampled 113 cases with 137 respondent bills. Only 28 were complete bills.

# We see a high rate of missing bills

Of about 4400 household interviews in RECS 2005…

- 2,794 respondents *said* they had an electric bill on hand
- 1,525 *said* they had a natural gas bill on hand
- But only 2,014 sample cases had at least one scanned energy bill from the interview

We will not impute half of our key estimates!

# Option 3: Have energy companies report for sampled households

# RECS collects data directly from energy suppliers: How?

1. Collect physical bills from household respondents
   - Examples of respondent bills by fuel are scanned into laptop
2. Identify their energy suppliers and contact info
   - Although interviewers enter *supplier contact info* and *account numbers* into Blaise, scanned bills can be used to validate this information and to compare against supplier-reported data
3. Collect billing data directly from their energy suppliers
   - We contact all energy suppliers for all interviewed households
   - Primary mode is a standardized paper questionnaire

Why does the supplier survey "work"?

**Supplier reporting is mandatory**

# RECS makes minimal use of respondents' knowledge, why?

- RECS produces estimates of annual energy consumption and expenditures
  - Respondents do NOT/cannot report these data even with records
  - Few bills and respondents provide data for the entire reference period, a year
- Lessens burden to household respondents
  - Only need provide *one* energy bill for each fuel type (e.g., an electricity bill, a natural gas bill, etc.)
- With the bills and reported supplier names, item response and coverage increase dramatically

# Using account records alters error structure in our favor

- Virtually no nonsampling error
- Little or no contribution to main survey unit nonresponse
- Supplier-level response is very high
  - Nonresponse would be clustered and affect a large area
  - In rare cases, household records cannot be found
- (Re)coding error possible, but more systematic and easier to detect and address than householder error
- Much less imputation necessary for key estimates

# Challenges to Using Data from Energy Companies

1. Linking customer data to household is tricky

- Primarily use address matching, but suppliers may need names or account numbers, too
- About 17 percent of name(s) on Authorization Forms do not match name(s) on bills
- Account numbers are difficult to record and use accurately-- vary in length (4-19 digits) and formatting is idiosyncratic

# Challenges (cont.)

2. Extends data collection period (two serial phases)

3. Maintaining response rates across two phase surveys

4. There is no standardized record format across suppliers
   - Numerous and varying terms for fees and surcharges (80 different terms in sample of 137 records)

5. Record keeping practices and ability to respond vary considerably between smaller bulk fuel suppliers and large utilities

# Sustaining records use is a moving target

- Company 'ownership' of records is changing, now includes 3rd parties. Beyond EIA's mandatory reach?
- Confusion over precedence of federal laws, state laws, and company promises to consumers is common
- Changes in billing systems, maintenance, and access affect nature of supplier respondent tasks and burden

# Getting more out of records

- Increase analytic value
  - Edit and impute data for data's 'natural' reference period, usually billing periods, to understand householders' consumption response to changes in weather, prices, and other factors
  - Produce estimates for calendar month to align with other economic data series
- Increase access
  - Share more data for statistical purposes under CIPSEA 2002
- Increase utility of one 'round' of a survey
  - Collect more monthly billing data to extend the data stream beyond reference period,
  - Produce inter-survey estimates and fill out the time series for key estimates

# Conclusions

- Using data producers (energy suppliers) reduces household burden, unit and item nonresponse for key estimates
- Burden on commercial entities increases
- Extends data collection schedule but improved data quality is worth it
- Using supplier data shifts Agency's effort to editing, modeling, imputation, estimation -- more complex, but interesting and valuable to overall mission

# Future Work

- Investigate data retrieval processes to understand how records are produced, tabulated and align with survey objectives
- Use findings to attempt some standardization for bulk of electronic data capture
- Understand the impact on data providers of requesting large numbers of customer records over long periods
- Make fuller use of natural granularity of records to model and forecast usage patterns and sharpen key estimates

# For more information

U.S. Energy Information Administration home page www.eia.gov

RECS Program http://www.eia.doe.gov/emeu/recs/contents.html

Eileen M O'Brien eileen.obrien@eia.doe.gov
Leader, Consumption Data Team