Mixing Survey Modes: Challenges for Questionnaire Construction

Prof dr Edith D. de Leeuw

Share Survey Modes Workshop

Mannheim, 8 July 2009

Why and How Modes Differ

Self-Administered vs. Interviewer-Guided

Visual vs. Aural

Media-related customs

Modes & Measurement

- Measurement error occurs when a respondent's answer to a question is inaccurate (departs from the "true" value)
- Modes vary in terms of:
  - Interviewer versus self-administered questionnaires
    - Interviewer impact
  - Stimuli / manner in which survey question is conveyed to respondent (and response is recorded)
    - Information transmission
  - Knowledge about mode, usage, social customs
    - Media related factors

How Modes Differ

Overviews: De Leeuw 1992, 2005 and Dillman & Christian, 2005

- Empirical Evidence **Interviewer Impact**
  - More social-desirability in interview
    - E.g., drinking, fraud
    - More open in self-administered modes
  - More positive in interview
    - Less lonely, better health in interview
  - More acquiescence in interview
    - Tendency to agree
    - Easier to agree than disagree with another person
  - Less missing data/more detailed answers open questions in interview
    - In general interviewer probes help

How Modes Differ 2

- Empirical Evidence **Medium Impact**
  - Few systematic studies (Overviews De Leeuw, 1992, 2005) indicate **advantage of self-pacing**
    - Self-administered more consistent answers
    - SAQ higher psychometric reliability on scales
      - BUT all Paper SAQ vs. interview!
- Internet as medium still different (cf. Krug, 2006)
  - Multi-tasking
  - Scanning
  - Satisficing (close enough in stead of optimal)

Internet as Medium

- Empirical Evidence Medium Impact
  - Hardly any systematic studies
    - Satisficing (less differentiation in web, Fricker et al, 2005)
  - Psychological testing
    - Equivalence when no time pressure (De Leeuw et al, 2003)
  - Conveying sincerity of purpose and trust more difficult
  - **More research needed on nonresponse**
    - Web on average 11% lower (meta-analysis Lozar Manfreda, et al, 2008)
  - Research needed on **response to sensitive** questions
    - Influence of SPAM
    - Trustworthiness web
      - Panel should have advantage vs. one time web survey
      - Existing relationship vs one-time

How Modes Differ 3

- **Information transmission**: visual vs aural; spoken vs written vs typed; question by question or blocks (page)
- Some evidence recency effect in telephone surveys
  - More often last offered answer category is chosen
- Context and order effects less likely in self-administered (paper) than interview
  - Overview / segmentation
    - No empirical studies including web surveys
- Visual presentation & design & quality
  - Growing body of evidence that respondents use all information including visual cues to decide what answer they are going to report
    - Cf Dillman, 2007; Toepoel, 2008; Couper 2009

Good news, but….

- Literature reports that there are some mode difference
  - Not large
    - Except for more sensitive questions
  - But….
    - All empirical evidence is based on
      - Well conducted experiments
        - Controlling/adjusting population differences
        - **Equivalent questions and questionnaires!**

Lesson Learned

- To minimize mode effects one should:
  - Control/adjust for population differences
    - E.g., More younger, higher educated in web and more elderly, lower educated phone
  - Use equivalent questions and questionnaires!
    - Ensure measurement equivalence

Questionnaire Design

Traditional Designs for Specific Modes and the Implications for Mixed-Mode Surveys

Traditional Design F2F

- Face-to-face: Visual + Aural
  - Show cards with answer choices
    - Long lists of answers, long scales with each point labelled
  - Pictures may be used
  - Open-ended questions on wide variety of topics
    - Trained interviewers are carefully instructed to probe in order to get detailed and complete information
  - No opinion etc not explicitly offered, but accepted when given. Interviewers often trained to accept ‘no answer’ only after a standard ‘probe’
  - Transitional texts to guide interviewer and respondent to next block of questions

Traditional Design Tel

- Telephone: Aural only
  - Shorter answer scales (2-5 point scales)
  - Often only anchored end-points
    - On a scale from 1 to 5 with 1 being not at all satisfied and 5 being completely satisfied
  - Visual analogue questions
    - Imagine a ladder with 7 steps
    - Imagine a thermometer with a scale from 0 to 100
  - Unfolding for longer scales
    - Satisfied, dissatisfied or somewhere in the middle
    - Completely, mostly, somewhat (dis)satisfied

Traditional Design Tel2

- Telephone design
  - Difference with face-to-face
    - In general breaking up questions in parts to accommodate loss of visual channel
  - Like face-to-face
    - Open-ended questions and probes
    - No opinion / no answer not explicitly offered
      - But is accepted after probe by well-trained interviewer

Traditional Design Postal

- Mail survey: Visual only, no interviewer present
  - In general, no breaking up of questions in parts
  - But, use longer list of response categories in stead
  - Fully labelled scales
  - Check all that apply instead of yes/no answers
  - Only ‘no answer’ when person skipped question, in stead of interviewer coded ‘refused, do not know, no opinion’
  - Go back and forth: more context available
  - Use illustrations / visuals

Example Mail vs Telephone

Mail

- Is the home in which you live
  - Owned free & clear
  - Owned with a mortgage
  - Rented
  - Occupied under some arrangement

Telephone

- Do you own or rent a home?
  - Follow-ups accordingly, e.g. when owns a home
    - Do you have a mortgage or is it owned free and clear

Example Face to Face

Face-to-face using show card with response categories

- Is the home in which you live
  - Owned free & clear
  - Owned with a mortgage
  - Rented
  - Occupied under some arrangement

Or when quick doorstep only aural

- Do you own or rent a home?
  - Follow-ups accordingly, e.g. when owns a home
    - Do you have a mortgage or is it owned free and clear

Traditional Design Web

- Web survey:
  - Visual only, but audio potential
  - No interviewer, but intelligent computer system
- Many similarities with mail
- Differences
  - More sequential offering of questions
  - Check all that apply almost standard format
    - Radio buttons (but…)
    - Evidence Christian et al (2008) check-all-that apply not optimal
  - Grids often used for groups of questions
    - What is best visual design?

Traditional Design Web2

- Web survey many similarities with mail plus some additional features, such as,
  - Visual only, but audio potential
  - No interviewer, but intelligent computer system
- Also many differences in question design
  - Special formats
    - Slider bars
    - Drop down menus
  - Open questions influenced by box size, dynamic space (cf Dillman)

Visual Illustrations

- Visual Illustrations are attractive
- May motivate respondent
  - Cover mail survey positive on response (e.g. Dillman's work)
- Help question/word meaning
- BUT: May influence respondent's answer!!
  - Example "How often do you eat out "
  - Illustration 1: couple, candlelight, rose in vase
  - Illustration 2: picture of MacDonald
    - Visuals/illustrations and their influence (Couper, 2007,2009)

Designing for Mixed-Mode Two Cases

Naively Mixing Enhances Measurement Errors

- Different modes have a tradition of different formats, and question format has effect on response distribution
- Consequence: Designers may routinely enhance **unwanted** mode effects in mixed-mode survey
  - E.g. unfolding in one mode, full presentation of all response options in other mode
- What to do?
  - **Design Questionnaire for Mixed-Mode**
    - How?

Design for Mix

- Two Situations:
  - One main method that accommodates the survey situation best
    - Main method is used to maximum potential
    - Other methods auxiliary
      - Example: Nonresponse follow-up
  - Truly multiple mode design
    - All modes are equally important
      - Example: PPSM, International surveys, Longitudinal studies, Respondent is offered choice

Design for Optimal Mix 1

- One Main Method, other methods auxiliary (cf Biemer&Lyberg 2003)
- Identify main method
  - Use main method optimal and to its maximum potential
  - Auxiliary methods designed **equivalent**
    - To avoid measurement error
    - May be perhaps sub-optimal for auxiliary method
      - Example: less response categories
- Note: Dillman et al (2009) coined this ‘mode-enhancement-construction’

Example LFS

- Longitudinal face-to-face & telephone
- Identify main method
  - Main method not necessary first method
    - Main method **telephone**
    - **Face-to-face** auxiliary from longitudinal point of view
- Main design for telephone interview

cf. Biemer & Lyberg, 2003

Example LFS cont

- Design **longitudinal** questions for telephone use
  - Not full potential *face-to-face* used in face-to-face interview
    - No visuals, no show cards
    - Shorter scales, unfolding
    - Open questions
- Design **one-time** recruitment questions for face-to-face use (full potential visual)
- Ensure **data integrity** for longitudinal use!

One Main Method

- Telephone with Face-to-Face Mixes
  - If telephone main method
    - Relatively easy to design mix optimally
      - Interviewer assistance in both modes
      - Do not use the ‘extra’ visual channel in face-to-face
  - If face-to-face main method
    - Absence of visuals makes it more complicated
      - Carefully balance pro and cons
        - Optimize for one? (preferred-mode specific design, aka mode-enhancement construction)
        - Or use ‘uni-mode’ design?
    - Implement a small experiment within one mode if possible!

One Main Method 2

- **Self-Administered Questionnaires and Interviewer Mixes**
- SAQ or Interview Main Method?
  - Complexity of questionnaire
    - Big issue in mixes with paper-mail not in mixes interview with web
  - Are visuals essential?
    - Face-to-face in mix may accommodate visuals, phone does not
      - CAWI-CATI may have problems, CAWI-CAPI not
  - Sensitive questions
    - Social desirability differences, interviewer influence
  - Is interviewer probing essential or not?
    - Paper mail problems, but web can emulate some probes
      - NCES example

Example NCES

- RTI surveys for National Center for Educational Statistics
  - TSMII-paper Wine et al at www.rti.org
  - Original studies were done by telephone
  - Switch to Web with telephone follow-up
    - Highly Internet savvy population
    - So web good choice, but…

Example NCES 2

- Switch to Web with telephone follow-up
  - But, researcher did not want to lose advantages interviewer
    - (Non)Response conversion
    - Clarification, solving inconsistencies, coding, etc
  - Blend best features of both modes

Example NCES 3

- Start with web survey ‘enriched’
  - Offer incentive for early completion
  - Help desk with
    - Specially **trained telephone interviewers**
  - Telephone prompts by phone by trained interviewers help-desk instead of standard e-mail reminders
    - Directly or on answering machine
    - Reminding of early completion incentive

Example NCES 4

- Questionnaire equivalence
  - Field tested
    - Some adaptation to web questionnaire
    - To make situation more equivalent to telephone interview
  - Changes in web-questionnaire
    - No answer option equivalence with interview
      - Continue button in stead of explicit ‘no answer’
      - But generic pop-up after 3 consecutive no answers to remind of importance
      - Key-items redisplayed with tailored text
    - Sensitive questions and SAQ
      - Revision finance items to be less sensitive
  - Help text designed for web also helped interviewers

Truly Multiple Mode Surveys: Modes are Equivalent

Three Approaches in Design

Modes Are Equivalent

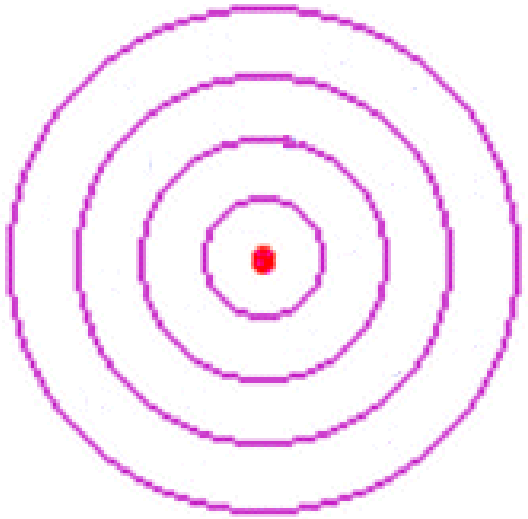
- Three schools of thought
  - Method Maximization
    - Optimize each mode ***separately***
  - Unified Mode Design or Uni-mode design
    - Provide the same stimulus (question format) in each mode, same ***offered*** stimulus
  - Generalized Mode Design
    - Purposively constructing questions to be different to achieve cognitive equivalence, same ***perceived*** stimulus
    - This can be seen as a sophisticated form of mode specific design (cf Dillman et al 2009)

I. **Method Maximization**

- Optimize each method **individually** as far as possible
  - If one method has an extra use it
- Rationale
  - Reduces overall error
  - Best of all possible worlds
- Assumption
  - Same concept is measured in both modes but with different accuracy only
  - Differences between methods only affect random error! (*no systematic bias*)

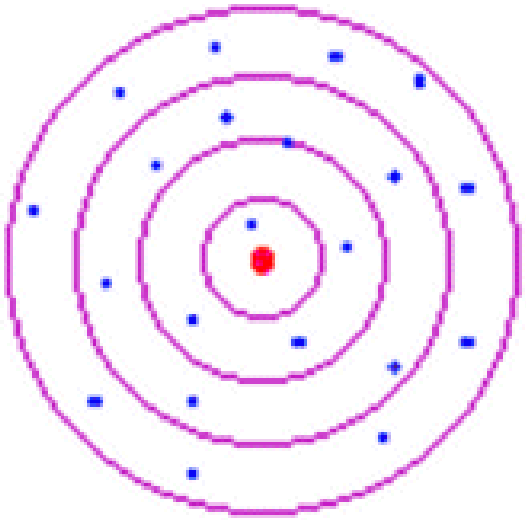
Reliability and Validity

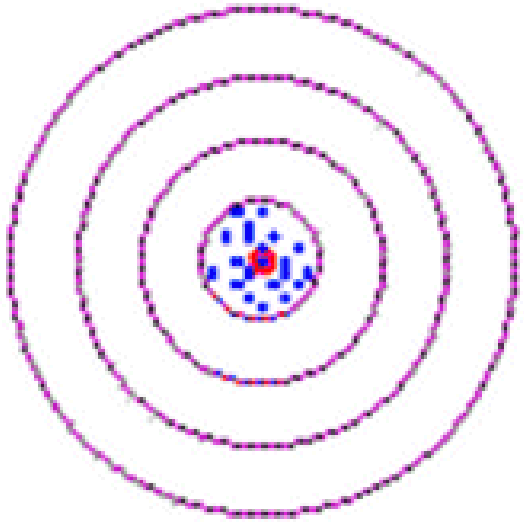
- Imagine an English County Fair and ….



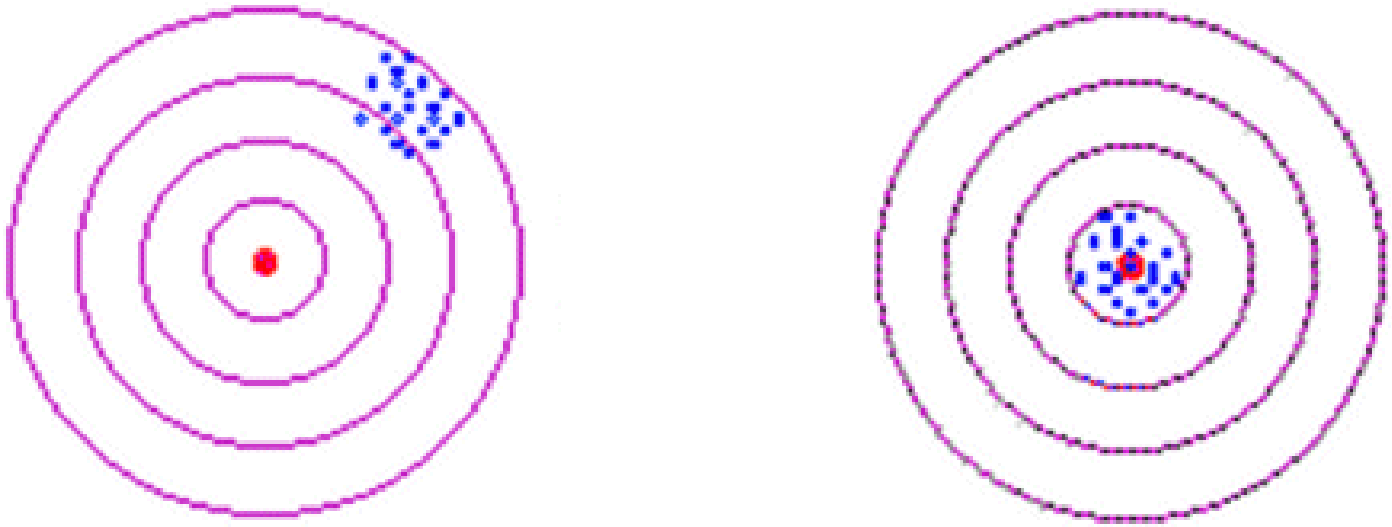
Variance vs Bias or Systematic Error

Low vs. High Reliability





Low vs. High Validity



Method Maximization continued

- Optimize each method individually
- Beware of Assumptions:
  - Differences between methods only affect random error!
  - M_1: $T+e_1$ $\quad e_1 \neq e_2$
  - M_2: $T+e_2$ $\quad e_1, e_2$ *random*
- But is this feasible?
- How about systematic error, bias?
  - **Danger of question format effects**
    - Example: check al that apply vs.. yes/no
    - Example: offer all response categories vs unfolding
- Burden of proof on designer

II. Unified Mode Design

- To minimize data integrity problems Dillman (2000) proposes UNI-mode design for all modes
- **Uni-mode design.** From **unified** or **uniform** mode design; designing questions and questionnaires to ***provide the same stimulus*** in all survey modes in order to reduce differences in the way respondents respond to the survey questions in the different modes.
  - Write and present questions the same or almost the same
  - Same offered stimulus in each mode
- How to do this, see Dillman (2000, 2006)

Uni Mode Design continued

- Unified or UNI-mode design for All Modes
  - Avoid inadvertently changing the basic question structure across modes in ways that change the stimulus.
  - Make all response options the same across modes and incorporate them into the stem of the survey question.
  - Reduce the number of response categories to achieve mode similarity.

(Dillman 2000, 2006, Chapter 6)

Uni Mode Design cont

- Unified or UNI-mode design for all modes
  - Use the same descriptive labels for response categories instead of depending on people's vision to convey the nature of a scale concept.
  - Develop equivalent instructions for skip patterns that are determined by answers to several widely separated items.
  - Reverse the order in which categories are listed in half the questionnaires to avoid recency/primacy effects
  - Evaluate interviewer instructions carefully for unintended response effects and consider their use for other modes.

(Dillman 2000,2006, Chapter 6)

Uni Mode Design cont

- Dillman, 200, 2006, chapter 6:

  "*Avoid question structures that unfold* "

- Comment:
  - Comes from paper mail survey-outlook.
  - One can and may unfold in both modes in CAWI-CATI design
  - Or in CAPI-CATI
  - Or in … any mix (but not in a mix with PAPI: paper mail)

Example UNI Mode Design

Mail, Telephone and Face-to-face interview

- Early attempt De Leeuw 1992, chap 4, p 37
  - http://www.xs4all.nl/~edithl/pubs/disseddl.pdf
- Response options the same across modes
- Same descriptive labels for response categories
- Reduced number of response categories
  - Maximum 7 pushing the limit for phone
  - But, used show cards in face-to-face
    - Equivalent with visual presentation mail
- Used simple open questions
- Interviewer instructions and instructions in mail questionnaire equivalent

In Sum: Uni-mode Design

- Designing for Mixed modes
  - Unified (uni-) mode questions to reduce mode effects
    - Question format
    - Response format
    - Instruction
- Uni-mode design for Mixed modes also implies
  - Besides questionnaire development
  - Questionnaire lay-out
  - Implementation procedures

Meeting the Challenge of Mixed-Mode Surveys

- Unified (uni-) mode questions to reduce mode effects
- Take it one step further, then designing questionnaire ….
- Do not think in traditions.
  - Example 1: RTI-Wine et all, 2006
    - Use telephone interviewers after training for web survey help-desk and for reminders

Do Not Think in Traditions

- Example
  - Question about employment
    - In CATI split up in two questions
      - Were you working for either pay or profit during the week of XXX?
      - If ‘no’ follow-up question: Were you not working for any of the following reasons: you were a student on paid work study, you were self-employed and did not get paid that week, you were on vacation from work or travelling while employed, you were on paid sick leave, personal leave or other temporary leave, or you were on a job that did not pay but had other benefits
      - Follow-up question was to make sure the respondent counted these experiences as employment

Do Not Think in Traditions continued

- Question about employment
  - To improve understanding CATI split up in two questions
    - Were you working for either pay or profit during the week of XXX? + follow-up question
    - Follow-up question was to make sure the respondent counted these experiences as employment
- Paper/Web traditionally no follow-up question
    - Paper/Web often incorporate the definition of employed in an instruction
  - But people do not read instructions and definitions carefully
  - Follow-up questions perhaps be better solution?

Meeting the Challenge of Mixed-Mode Surveys 2

- Step:1 unified (uni-) mode questions to reduce mode effects
- Step 2: Do not think in traditions
- Step 3: From unified to an integrated mode design?

III. Generalized Mode Design

- From unified (uni) mode design to an integrated, generalized mode design

- **Generalized mode design.**
  - Purposively constructing questions and questionnaires to be different in different modes with the goal of ***achieving cognitive equivalence*** of the perceived stimuli, thereby resulting in equivalent answers across modes.
  - The ***same*** offered stimulus is not necessarily the same ***perceived*** stimulus!

Generalized Mode Design continued

- Prerequisites integrated, generalized mode design
  - Designer understands
    - How differences between modes affect the question-answer process
    - How they affect the way respondents *perceive* the question, process the information and select and communicate the response
  - Designer does not think in traditions
  - Burden on the researcher to demonstrate that different questions elicit equivalent responses.

Understand What Happens

- To reach cognitive equivalence
  - Check with respondents, because Medium May Change Meaning
  - Same worded question can be perceived differently in different modes
    - Wording may be adequate in one medium and awkward in another

- Example Pierzchala et al, 2003
  - CATI: Are you {name}?
  - Web: Is your name {name}?

Example Generalized Mode Design

- Christian, Dillman & Smyth (2005)
  - CATI
    - When did you start attending WSU, when did you get your degree, etc
    - Interviewer probed for desired format
  - First Web-design
    - Same question text “When did you start attending WSU” ,“ When was your degree granted”, etc
    - With write in standard programming: mmyyyy

Date Degree Granted: ____ / ____ (MM/YYYY)

    - Too many errors

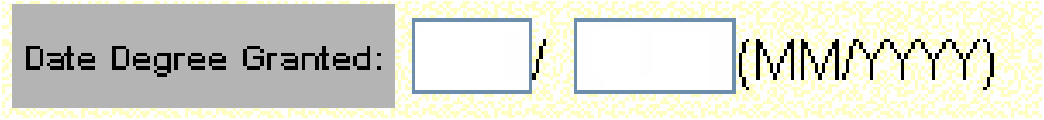
Example continued Generalized Mode Design

- In CATI
  - Interviewer probed for desired format
  - Interviewer is intelligent system
- Web Can Be Intelligent System too!
  - Try to emulate interviewer
- Christian, Dillman & Smyth (2005)
  - http://survey.sesrc.wsu.edu/dillman/papers/Month%20Year%20Technical%20Report.pdf
  - Redesigned size boxes, graphical language, symbols
    - Equivalence needed more than the same question wording!

Example continued Generalized Mode Design

- Web Can Be Intelligent System too!
  - Try to emulate interviewer
  - Christian et al (2005) redesigned size boxes, graphical language, symbols

In stead of



  - Equivalence needed more than the same question wording

Designing for Equivalence

- Need for more theory
- Need for empirical data
- Embedding small experiments / mode comparisons
  - Provide data to estimate mode effect
- Using these data for empirically based adjustment
  - Weighting
  - Propensity score adjustment

Embedded Experiments and/Adjustment

Build in overlap	**Method 1**	**Method 2**
Group X	**Main Data Collection**	**Some Data**
Group Y	**Some Data**	**Main Data Collection**

Suggested Readings

- Introduction Mixed-Mode
  - Edith D. De Leeuw (2005). To mix or not to mix data collection methods in surveys. JOS, Journal of Official Statistics, 21,2, 233-255 (also available on www.jos.nu)
- On quality, data collection, writing questions, online research, mixed-mode, analysis
  - De Leeuw, Hox & Dillman (2008). International Handbook of survey Methodology. New York: Lawrence Erlbaum/Psychology Press, Taylor and Francis Group

Suggested Websites

- Don A. Dillman's homepage
  - http://survey.sesrc.wsu.edu/dillman/
    - Papers
- Edith de Leeuw homepage
  - http://www.xs4all.nl/~edithl/
    - Additional material and readings accompanying the International Handbook of survey methodology
- University of Michigan-ISR
  - http://ccsg.isr.umich.edu/
    - Guidelines for best practice in cross-cultural studies
- US Bureau of the Census
  - http://www.census.gov/srd/mode-guidelines.pdf
    - Guidelines on developing demographic questions for administration in different modes

Suggested Websites 2

- Journal of Official Statistics
  - http://www.jos.nu
    - All articles that appeared in JOS online available, simple search menu
- Survey Practice (AAPOR)
  - http://www.surveypractice.org
    - Short practical articles
    - Autumn 2009, theme propensity score adjustment
- European Survey Research Association
  - http://www.surveymethodology.eu/journal/
    - Online journal on survey methodology