# Reporting Standards for Social Science Experiments

Kevin Esterling
University of California - Riverside

Summer Institute
June 2014

BITSS
Berkeley Initiative for Transparency in the Social Sciences

# Introduction

- Reporting and disclosure are essential for contributing to the accumulation of knowledge
  - Minimize researcher degrees of freedom
  - Enables others to assess the plausibility of your identifying assumptions, validity, generalizability, etc.
- Disclosure is most important when you generate your own data
  - Today's focus will be on social science RCTs
  - For surveys, see AAPOR standards
  - For archival data, do your best!

# Costs and Benefits of Disclosure

- High costs
  - RCTs are "inconsistent with the human spirit"
    - Data collection is often messy
    - Studies do not go as planned
  - Time consuming to keep track of this in real time
- Benefits of disclosure:
  - Ethical norms; your own identity as an ethical person
  - Accumulation of knowledge; you are a member of society
  - Disclosure provides checklist of crucial design and analysis elements that you need to think about anyway
  - (*) Costly signal about the quality and merits of your research; your reputation and the venues for publication

# Illustration

## Contents


**A Introduction** 2

**B Experimental Design** 2
- B.1 Subject Recruitment and Selection . . . 2
  - B.1.1 RSVP and Assignment . . . 2
  - B.1.2 Treatment Compliance . . . 5
  - B.1.3 Administration of the Follow Up Survey . . . 6
- B.2 Background Reading Materials (BGM) . . . 7
- B.3 KN and Deviations from an Ideal Design . . . 11
  - B.3.1 Retaining the RSVP "No" Respondents in the Analysis . . . 11
  - B.3.2 Excluding the Chronic Nonresponders from the Follow Up Survey . 14
  - B.3.3 District Panel Sizes and Subcontracting for Subjects . . . 17

**C The Statistical Model** 18
- C.1 Identifying Causal Effects using Principal Stratification . . . 18
- C.2 Estimation . . . 21
- C.3 Additional Control Variables . . . 25
- C.4 Measurement Model Results . . . 28
  - C.4.1 Validity Test of the Compliance Indicators . . . 30
- C.5 Imputing Missing Data for the Endogenous Variables . . . 32
- C.6 Comparing Principal Stratification to Matching and IV . . . 32
- C.7 Identifying Direct and Indirect Effects in the "Causal Mechanism" Model 34
- C.8 Modeling the Compliance Process . . . 35

**D Scaling up the Experimental Design** 37

**E Conclusion** 41

# Reporting Standards

- Existing standards
  - CONSORT for biomedical research
  - CONSORT-SPI in development
  - Many others… (including AAPOR for certain survey experiments)
- Social science: Organized Section on Experimental Research of APSA (XPS)
  - Minimum reporting standards
  - Developed by the Experimental Standards Committee
  - Adopted by the *Journal of Experimental Political Science*
- XPS standards are a checklist with six sections
  - Hypotheses
  - Subjects and Context
  - Allocation methods
  - Treatments
  - Results
  - Other information

# Hypotheses

- Specific objectives or hypotheses
  - State the questions the experiment designed to address
  - What are the specific hypotheses to be tested?
    - Be sure to delineate which hypotheses and subgroup analyses were developed in advance of the data analysis
    - Also, note primary and secondary outcome measures

# Subjects

- Eligibility and exclusion criteria for participants
  - Why was this subject pool selected?
  - Who was eligible to participate in the study?
- Procedures used to recruit and select participants
  - Recruitment dates defining the periods of recruitment
  - What would result in the exclusion of a participant?
  - Were any aspects of recruitment changed (such as the exclusion criteria) after recruitment began?
- Sample size
  - Intended number of participants per cell or plan to stop recruitment
  - Best practice to conduct power analysis

# Allocation Method

- Procedures used to generate the assignment sequence (e.g., randomization)
  - Details of procedure (e.g., any restrictions, blocking)
  - Unit of randomization (individuals, groups, households, etc).
  - Provide evidence that assignment was successfully implemented, such as balance scores on pretreatment variables
- Blinding
  - Were participants, those administering the interventions, and those assessing the outcomes unaware of condition assignments?
  - If blinding took place, include a statement regarding how it was accomplished and how the success of blinding was evaluated

# Context

- Settings and locations where the data were collected
  - Field
  - Lab
  - Classroom
  - Online
- Other relevant specifics of the population
  - Large public university vs. small private university
  - Geographic location
  - Social networks or proximity of subjects
- Timeframes
  - When the experiments were conducted
  - Dates of any repeated measurements as part of a follow-up

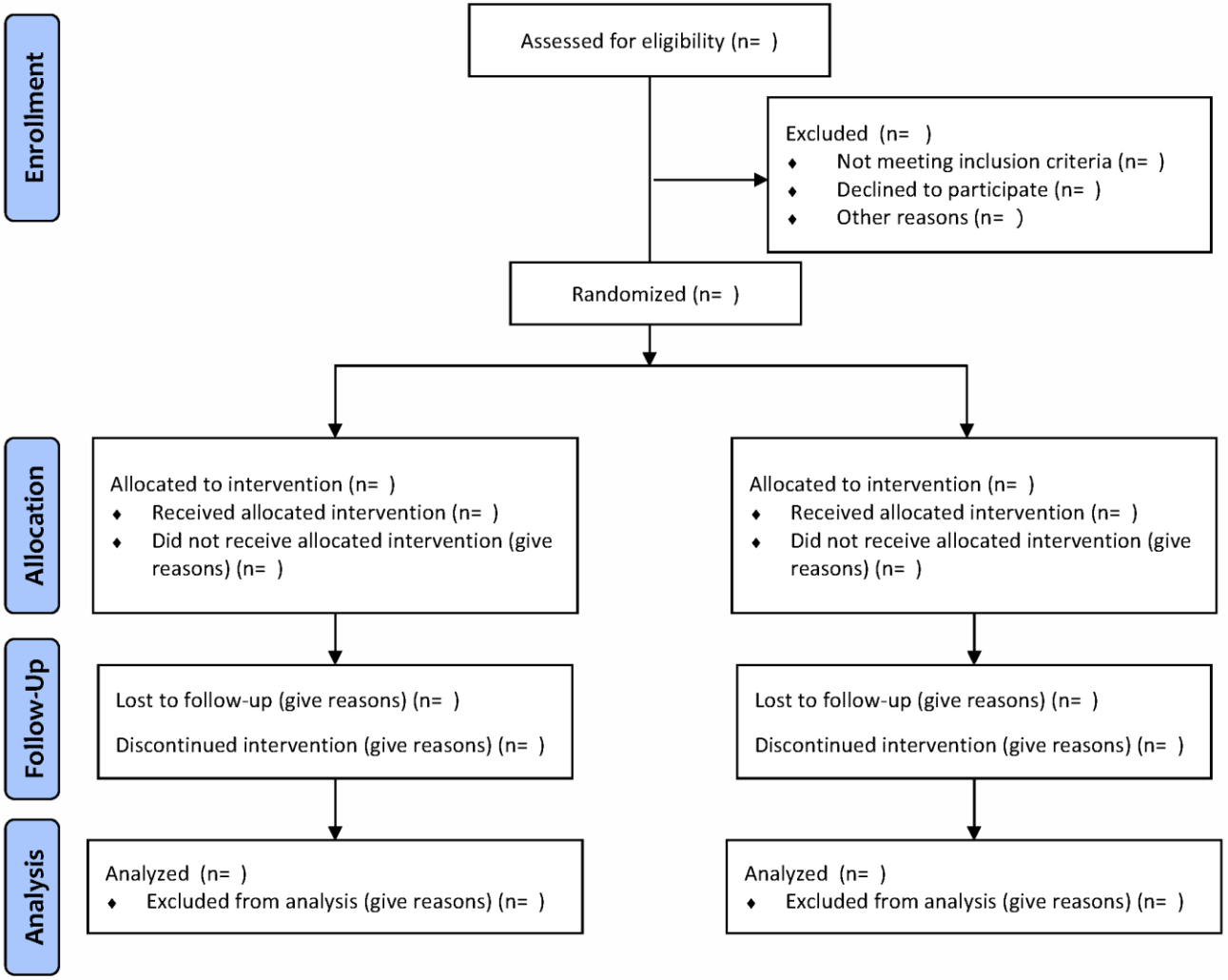
# Treatments

- Description of the interventions and their timing in each treatment condition
  - Method of delivery (paper, computer, face-to-face, telephone)
  - Was deception used? Were incentives given?
  - Manipulation checks; other evidence on whether the treatment was delivered as intended
  - Report any instructional anomalies or problems in administration
- For lab experiments (and other experiments, when relevant):
    - Report the number of repetitions, group rotation, ordering of treatments, piggybacking of other protocols
    - How long did each experiment last? How many sessions were subjects expected to attend? Amount of time between sessions
    - Were subjects given quizzes on the experimental instructions?
    - Were there practice rounds? If so, how many and what were the results?
    - Did subjects complete a post-experiment debriefing, interview, or questionnaire? If so, is there evidence that subjects understood the instructions and treatments?
- Descriptions should be sufficient to allow replication: verbatim treatment materials in appendix

# Results: Non-Compliance and Attrition

- Complete CONSORT Participant Flow Diagram (if non-trivial non-compliance or attrition)…

# Consort 2010 Flow Diagram



Here was our design going into the field:

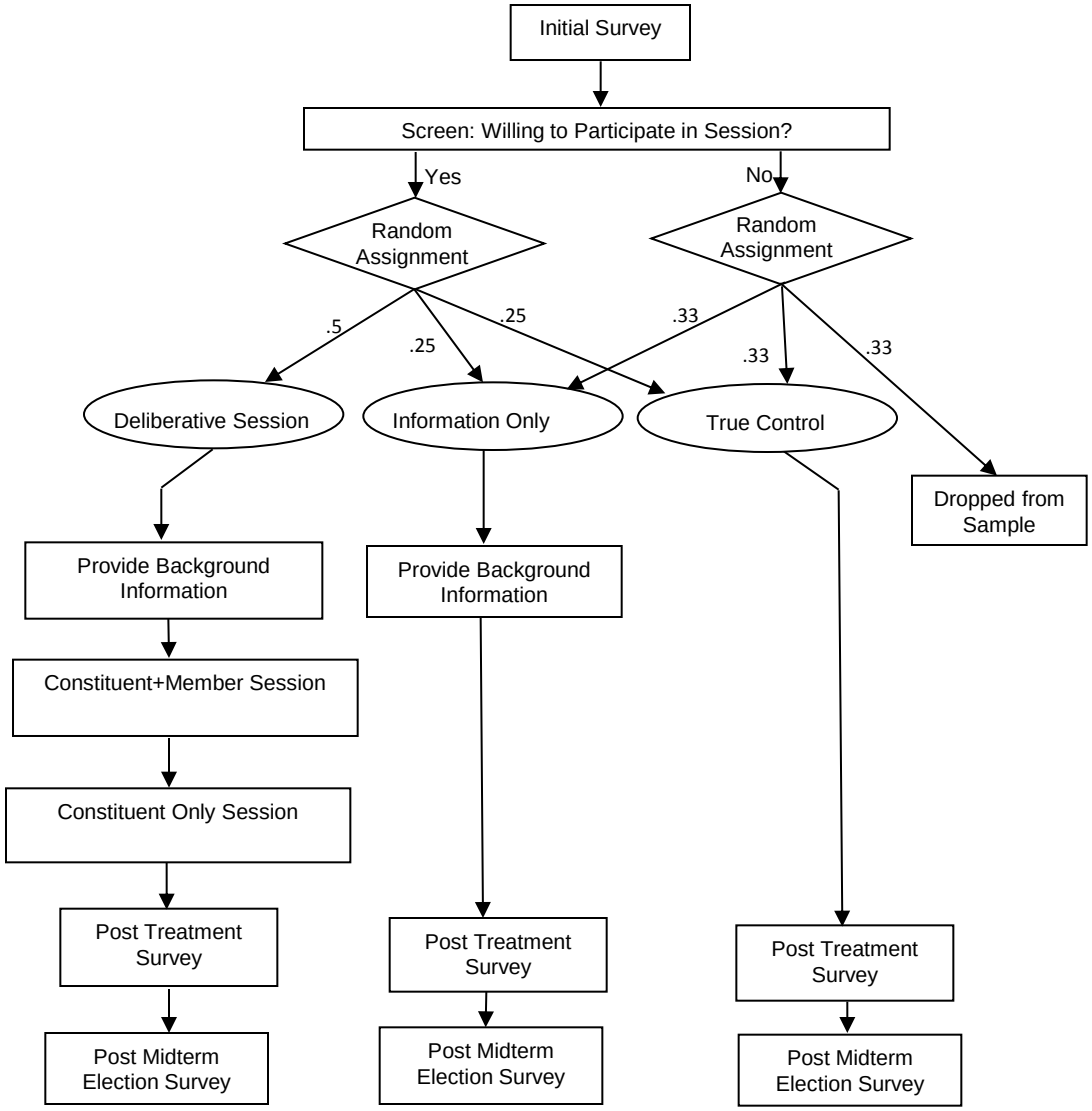
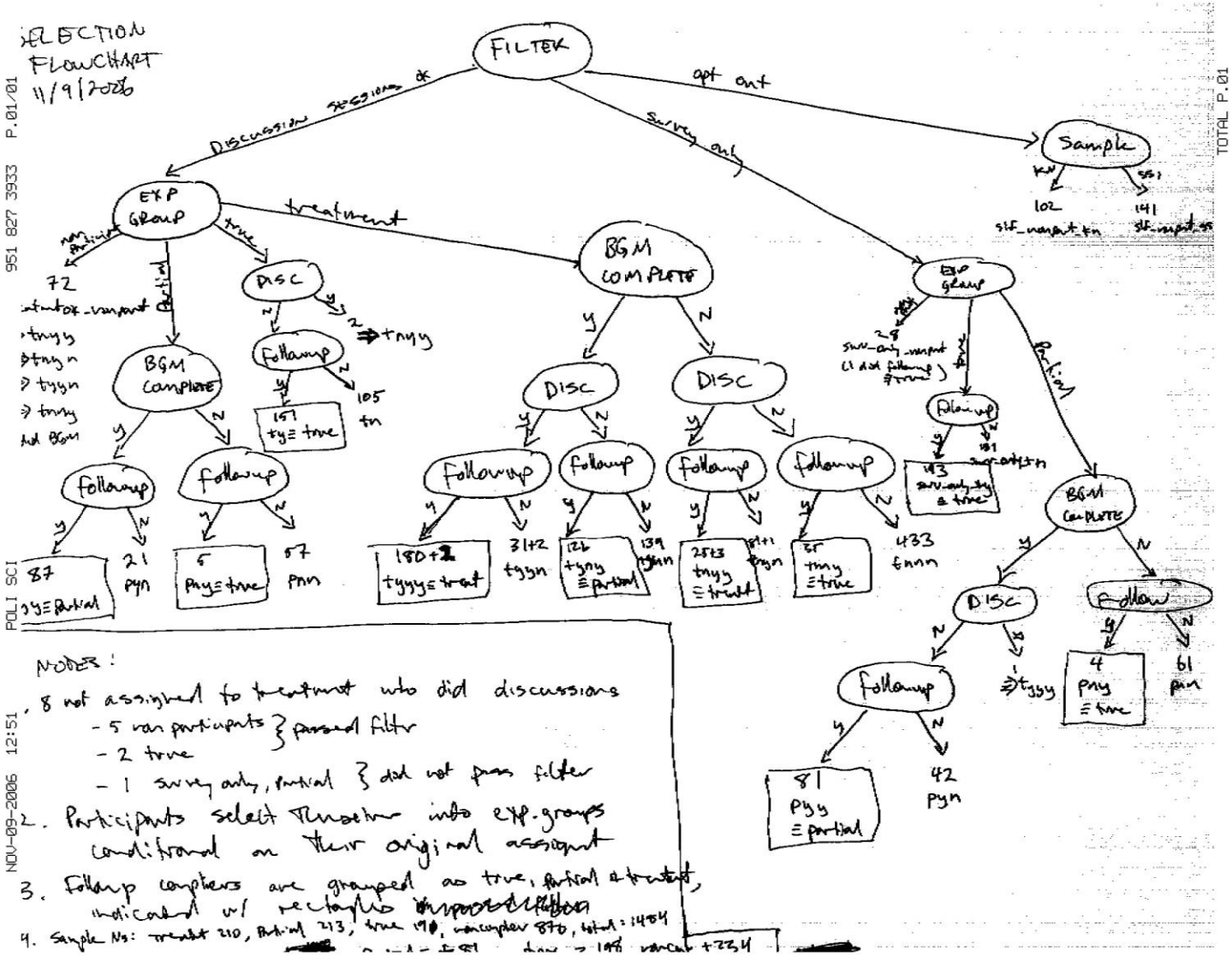



**Figure 1. Project Overview**

Here was the actual compliance among subjects:



Here was the actual compliance among subjects:

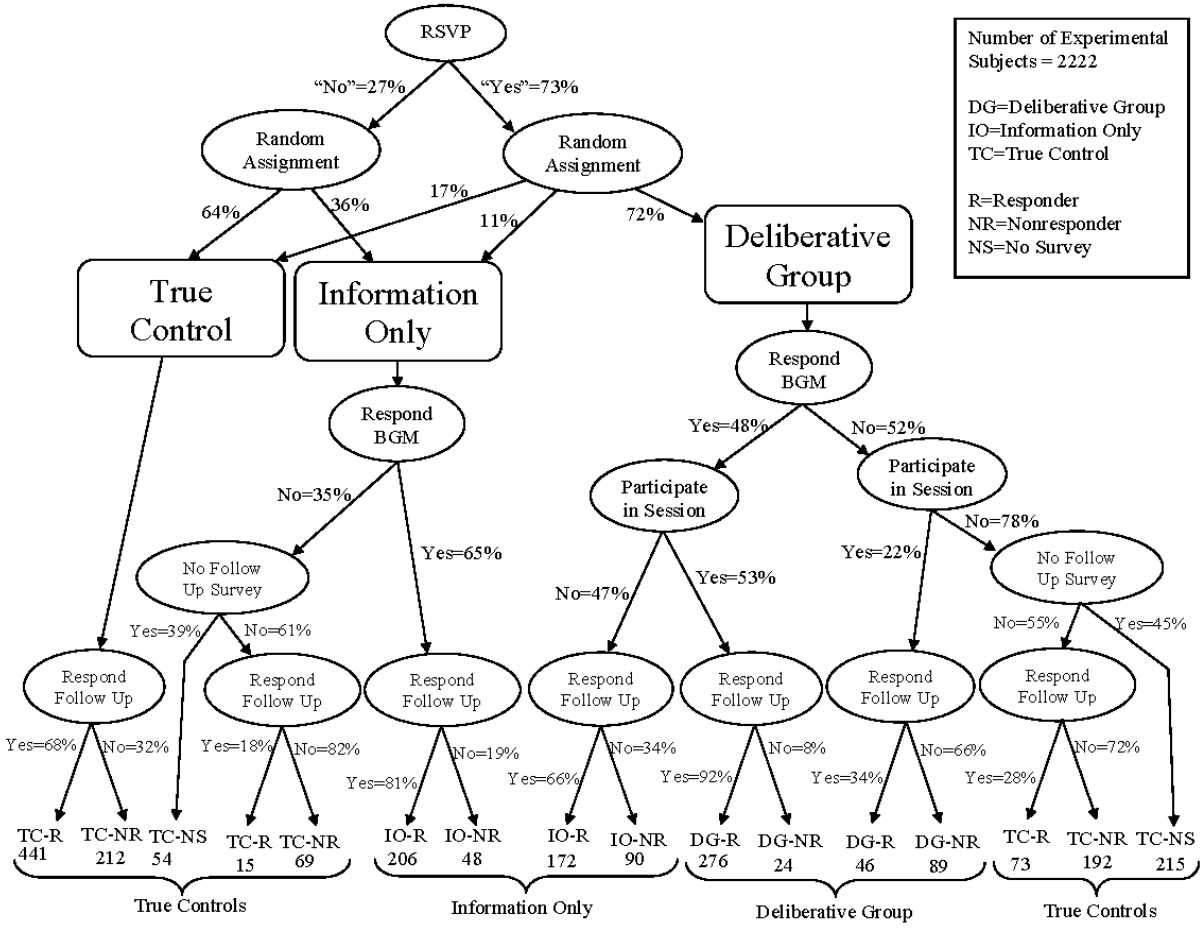



Figure A1: Assignment, Compliance and Response Rates

# Results: Treatment Effects

- Summarize Outcome Measures and Covariates
  - For indices, provide exact description of how they are formed
  - Clearly state which of the outcomes and subgroup analyses were specified prior to the experiment and which were the result of exploratory analysis
- Statistical Analysis
  - Report ITT and local effect estimates (reporting or weighting by blocks if appropriate) + identification strategy
  - Discuss reasons for noncompliance and attrition and examine if related to pretreatment variables
  - Report missing data by group and methods for addressing missing data
  - Note if level of analysis differs from level of randomization and estimate appropriate standard errors

# Other Information

- Was the experiment reviewed and approved by an IRB?
- If the experimental protocol was registered, where and how can the filing be accessed?
- What was the source of funding? What was the role of the funders in the analysis of the experiment?
  - Were there any restrictions or arrangements regarding what findings could be published?
  - Any funding sources where conflict of interest might reasonably be an issue?
- If a replication data set is available, provide the URL

# From the Dictionary of Useful Research Phrases

- "Three of the samples were chosen for detailed study…."
  - *The results of the others didn't make sense*
- "Typical results are shown…."
  - *The best results are shown*
- "A careful analysis of obtainable data…."
  - *Three pages of notes were obliterated when I knocked over a glass of beer*
- "While it has not been possible to provide definite answers to these questions…."
  - *An unsuccessful experiment, but I still hope to get it published*