THE DOG THAT DIDN'T BARK… INTERPRETING NON SIGNIFICANCE

Diana Eugenie Kornbrot

d.e.kornbrot@herts.ac.uk

Rachel M. Msetfi

r.msetfi@herts.ac.uk

University of Hertfordshire, UK

Aims

- **Improve Statistical Reporting**
  - Focus on 'non-significant' results
  - Preliminary to Usable and Useful Guideline
- **Views of experts on realistic scenarios**
  - Compare proportions with 'high' blood pressure
  - Two independent groups
- **Do experts agree:**
  - on what is salient to report to professionals?
  - on implications of non-signifcant effects?
  - on future prediction of numbers for population/ later samples
- **Does discipline expertise & role matter?**

Background

- Importance of Non-Significant Effects
- Planned Research
  - Ensuring *Absence* of Side Effects
  - Protection from false positives, (vaccines)
  - Counterbalancing of confounding factors
- Exploratory Research
  - Multi-factor studies, Multiple regression
- Current Guidelines
  - Usually for significant effect
  - Need POWER for NS effects

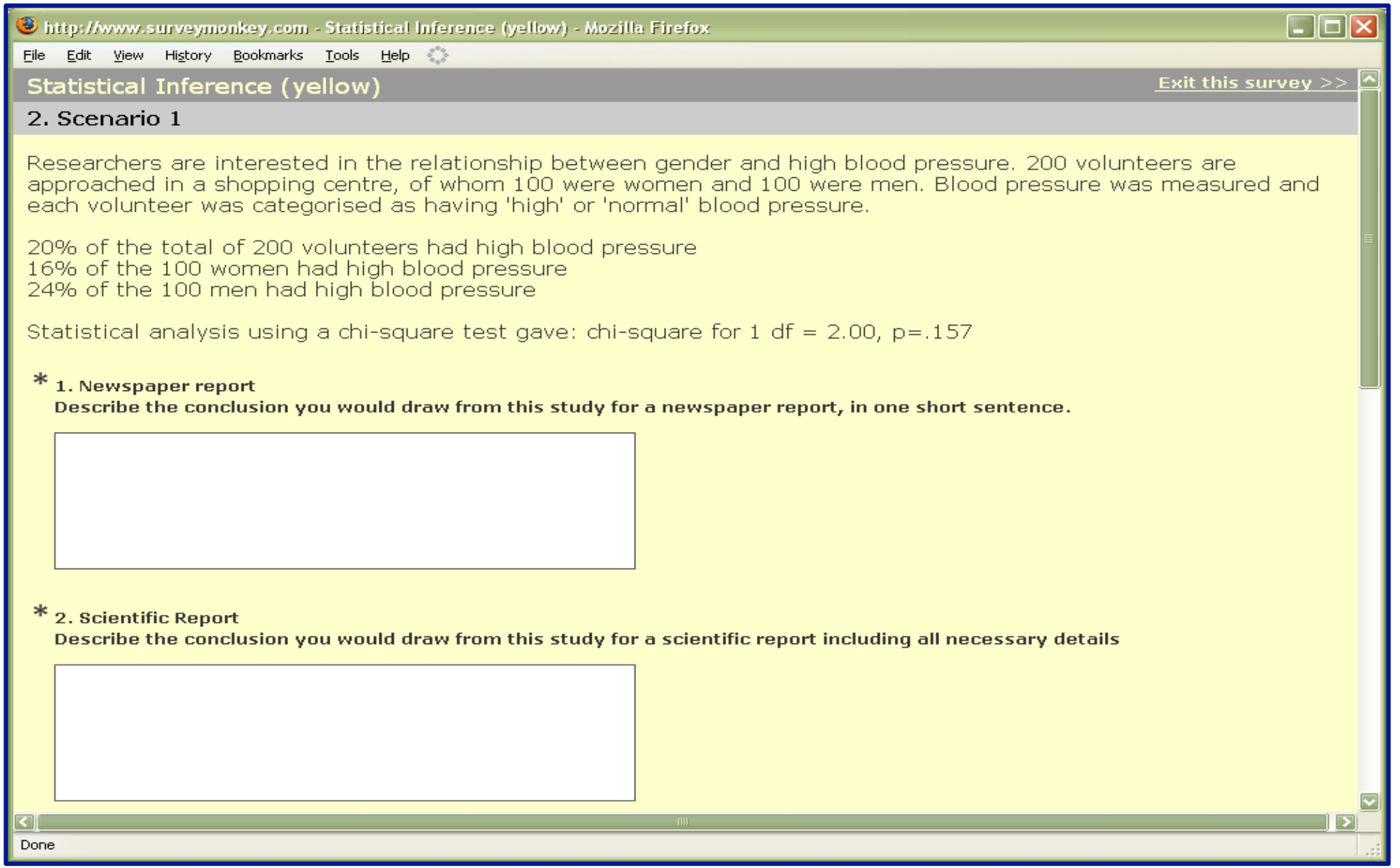
Journal Recommendations (NS)

- **Psychology**
  - APA, Psychonomic Science, APS, EPS
  - For NS: significance level + p(null)
  - Power – No mention
- **Education**
  - AERA
  - For NS: no specific recommendations
  - Power – No mention
- **Medical**
  - CONSORT group & associates
  - For NS: no specific recommendations
  - Power calculation for sample size
  - Power for Results – No mention

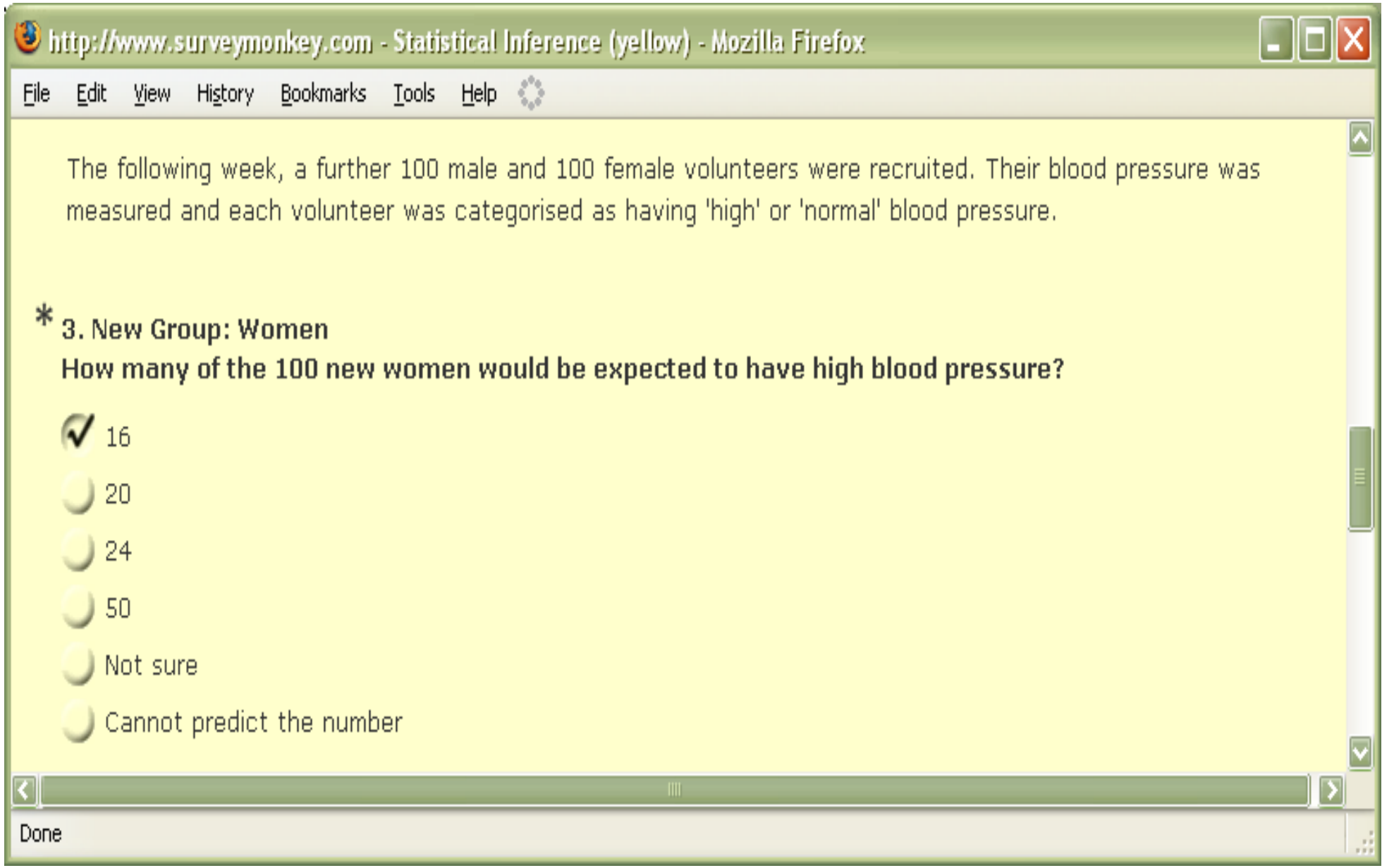
Method

- **Participants**
  - Convenience sample (230) recruited from email lists
- **Materials & Design: Internet Survey**
  - Two scenarios with non-significant chi-square
  - Participants are asked to make:
    - free form scientific report [categorized]
    - fixed form predictions of replication numbers
- **A priori Power**
  - Following OUR proposed guidelines…
  - Power .79-.98 for medium effect size

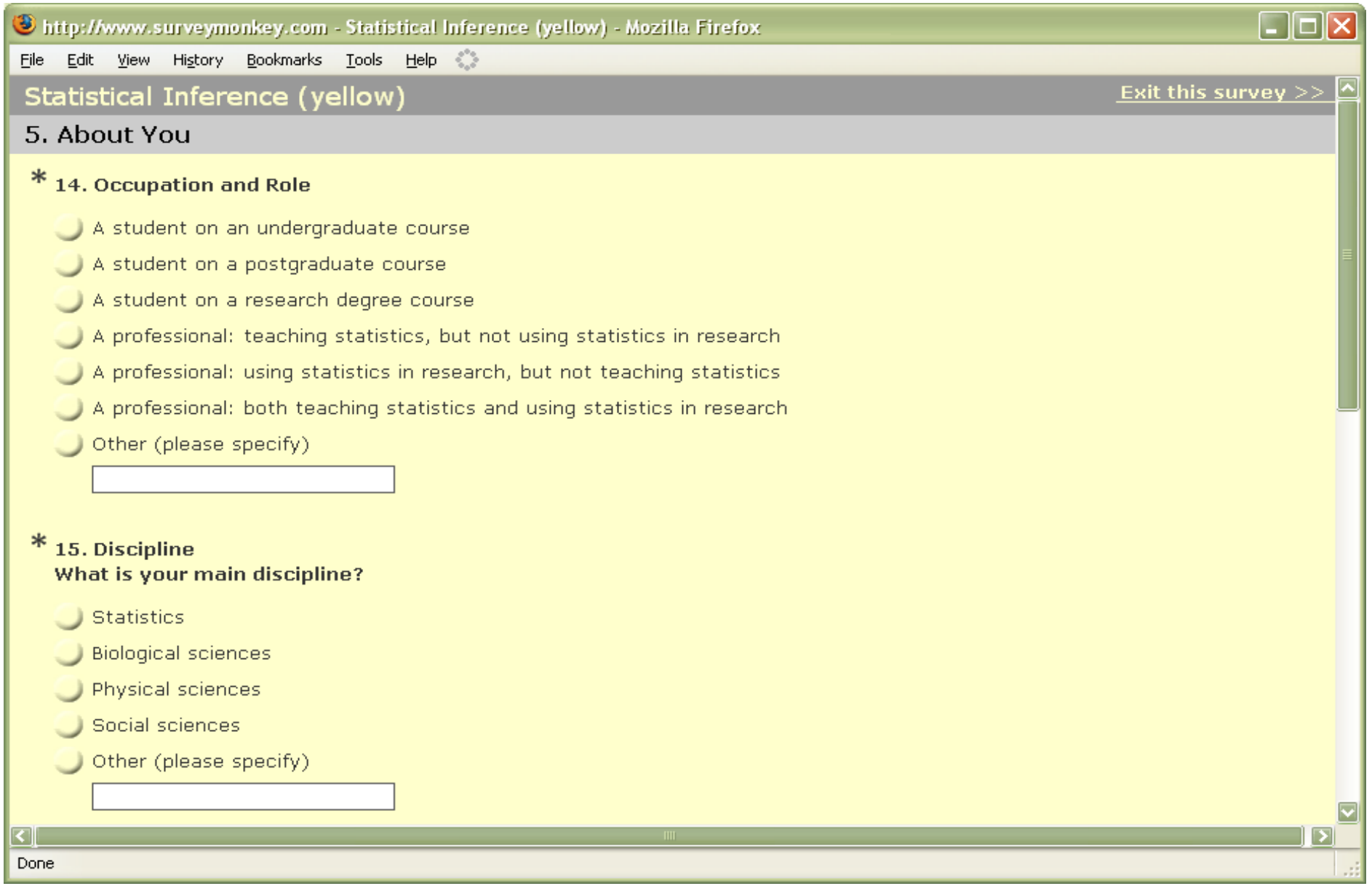
Free Form Scientific Report



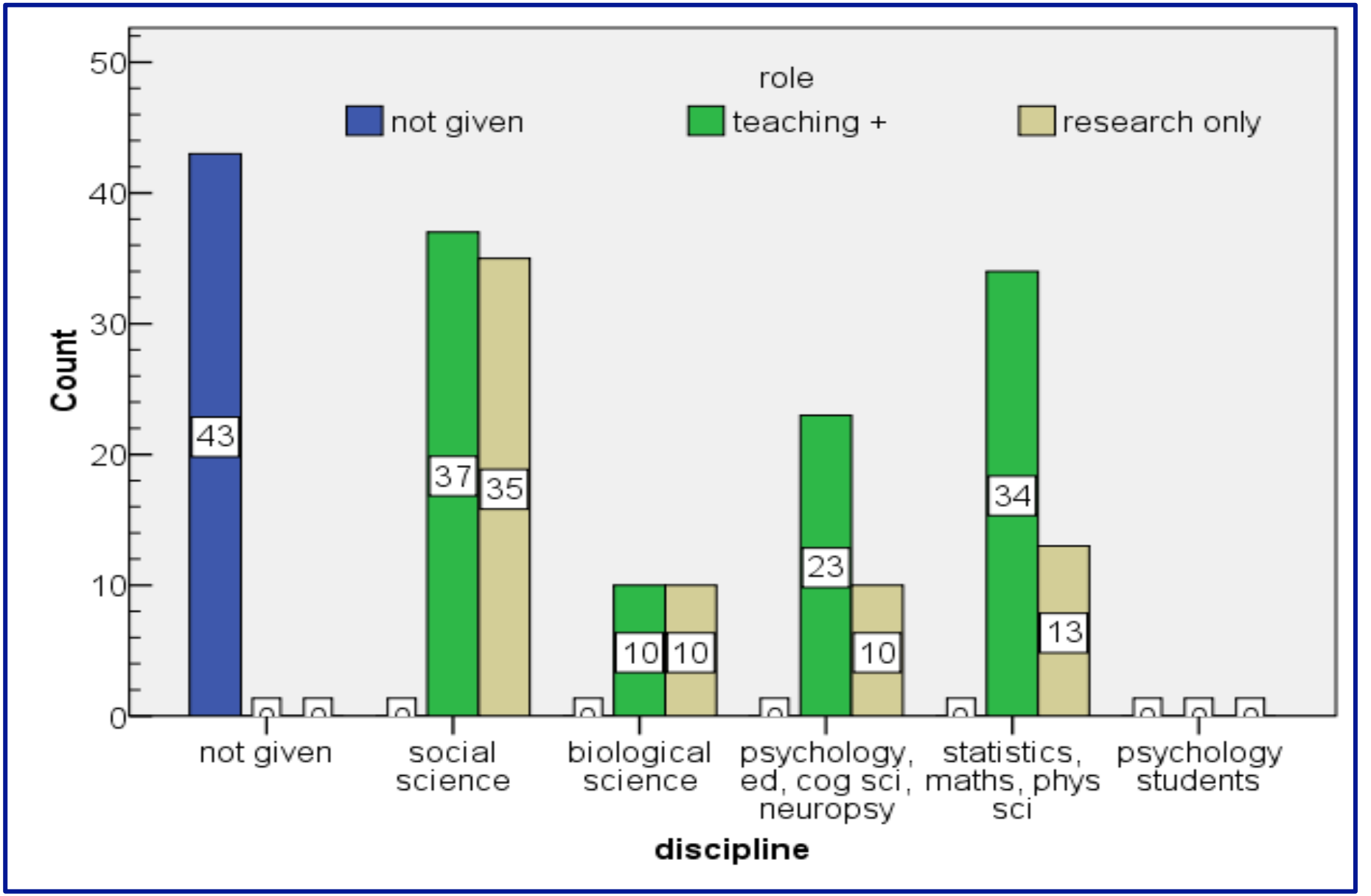
Fixed Form Future Number Predictions



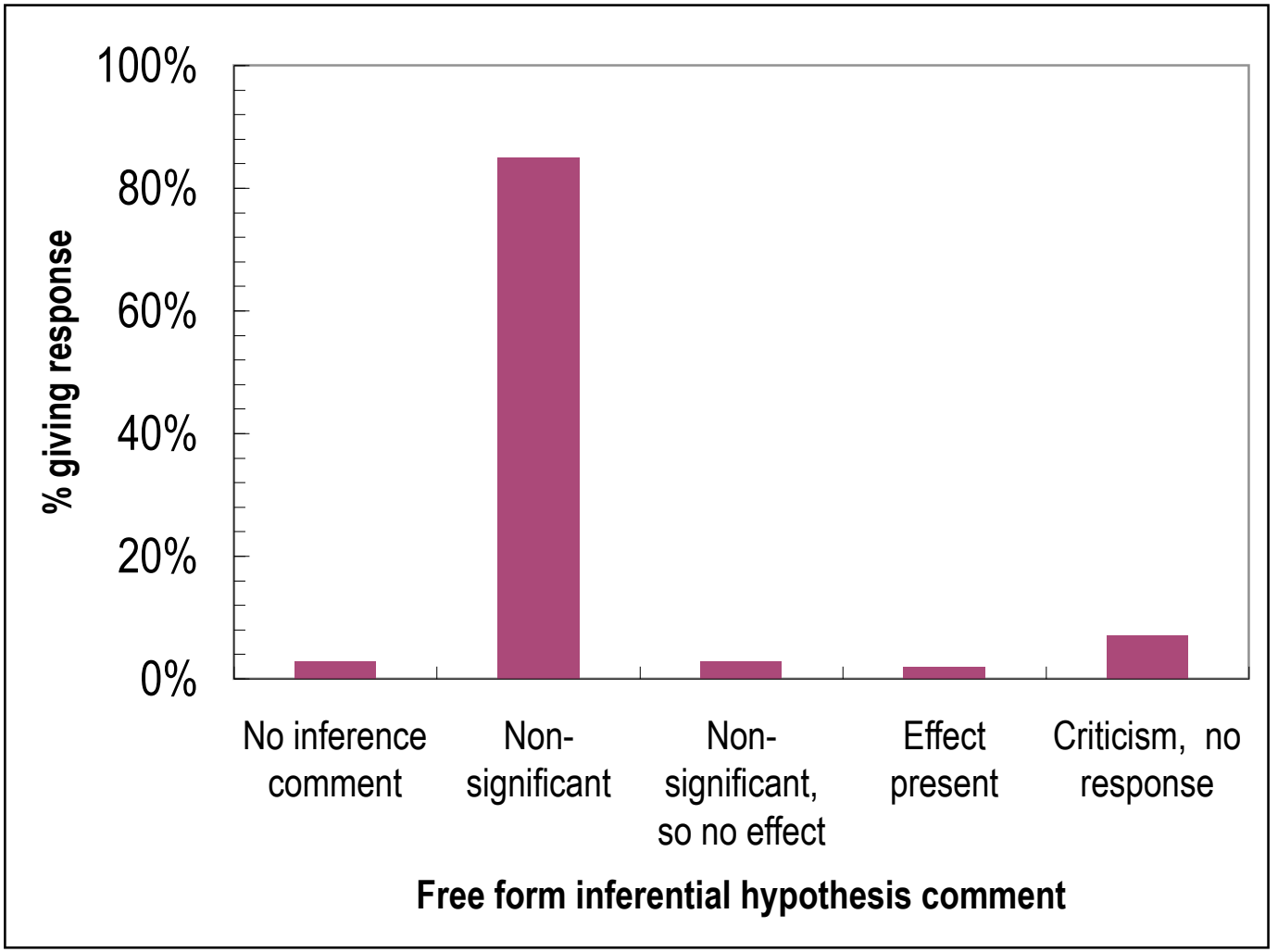
Expertise Role and Discipline



Participants: Discipline & Role

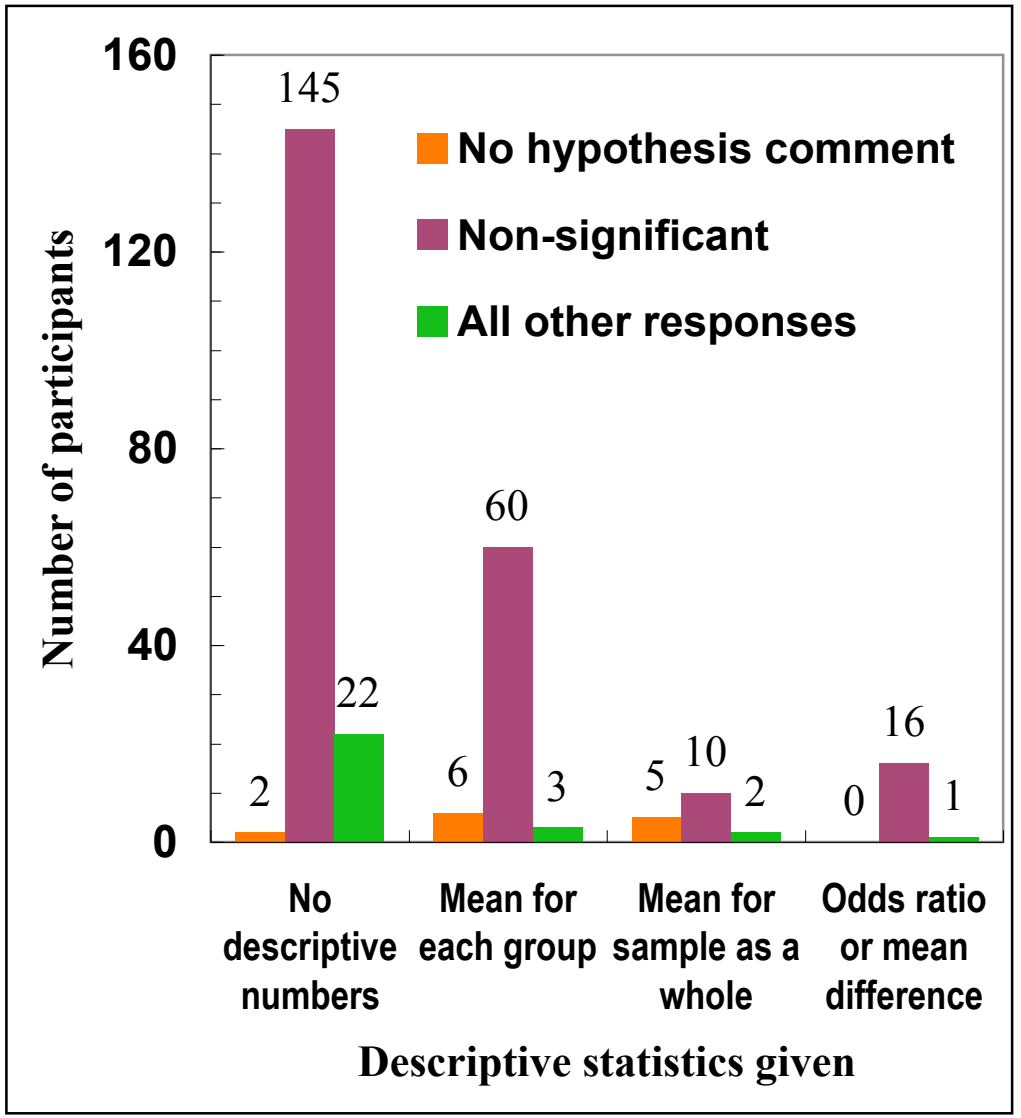


Free Form Scientific Report
All Participants: Inference Comments



81% of participants stated no significant effect or insufficient evidence

Free Form Report All Participants: Inference & Descriptive Comments



Majority DO give results of hypothesis test

Majority do NOT give any descriptive numbers

More give means for each separate group than for combined group.

Contrast future predictions, where about same number for separate and combined group means [slide 14]

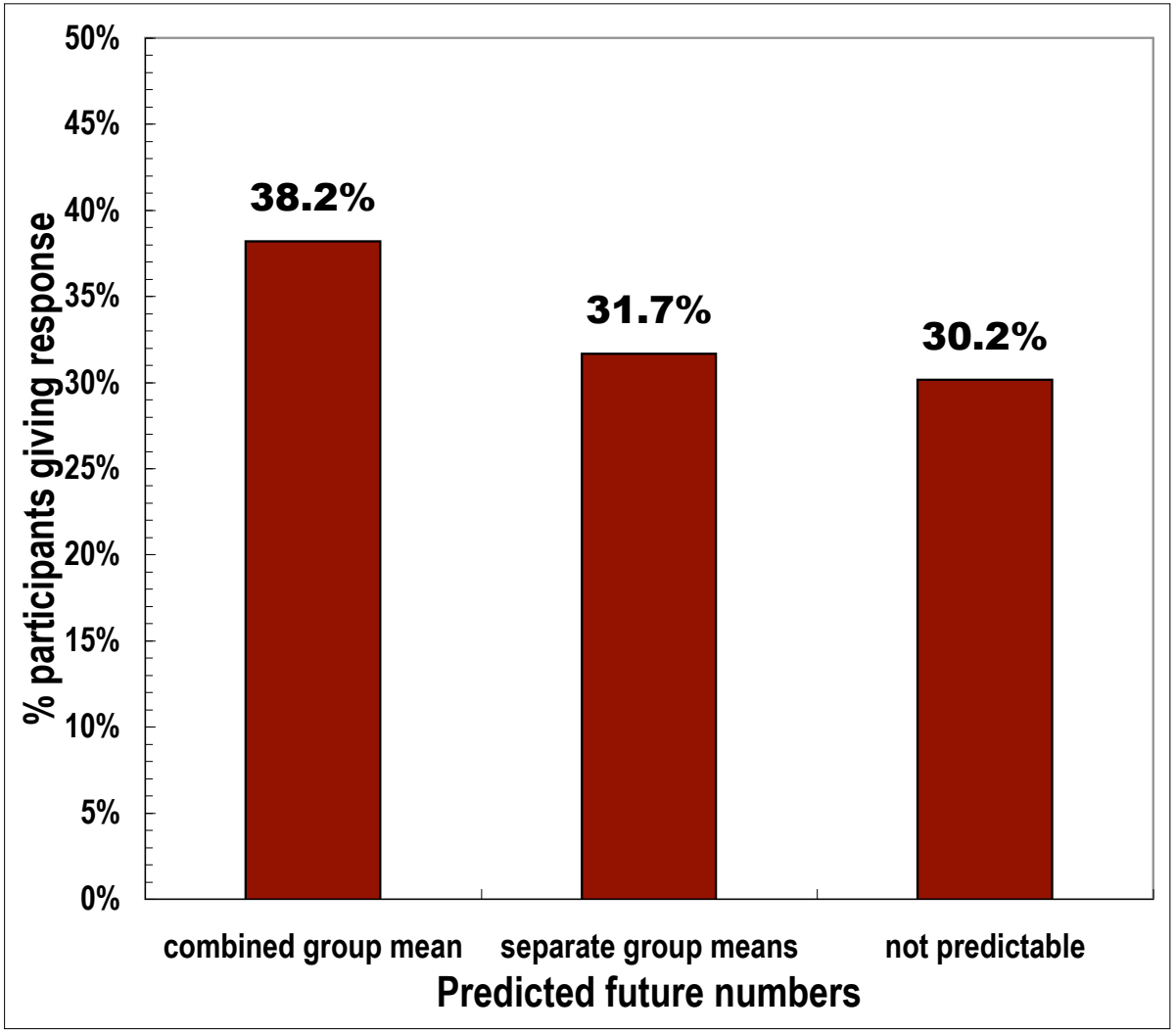
Free Form Scientific Report
All participants
Study Size (N) & Sufficiency for Replication

- Study size, N, availability from description
  - N not available 73%; N available 27%
- Sufficiency for Replication or Meta Analysis
  - Insufficient 86%; Sufficient 14%
- No effects of Expertise

Free Form Scientific Report
All Participants
on Power

- No mention at all — 76%
- States low power or low N or queries the power — 24%

Prediction of Future Replication
All Participants: Fixed Form



Many think no future predictions possible!

Separate group means and combine group means have similar frequencies

Unlike free form, where separate group means more frequent [slide 11]

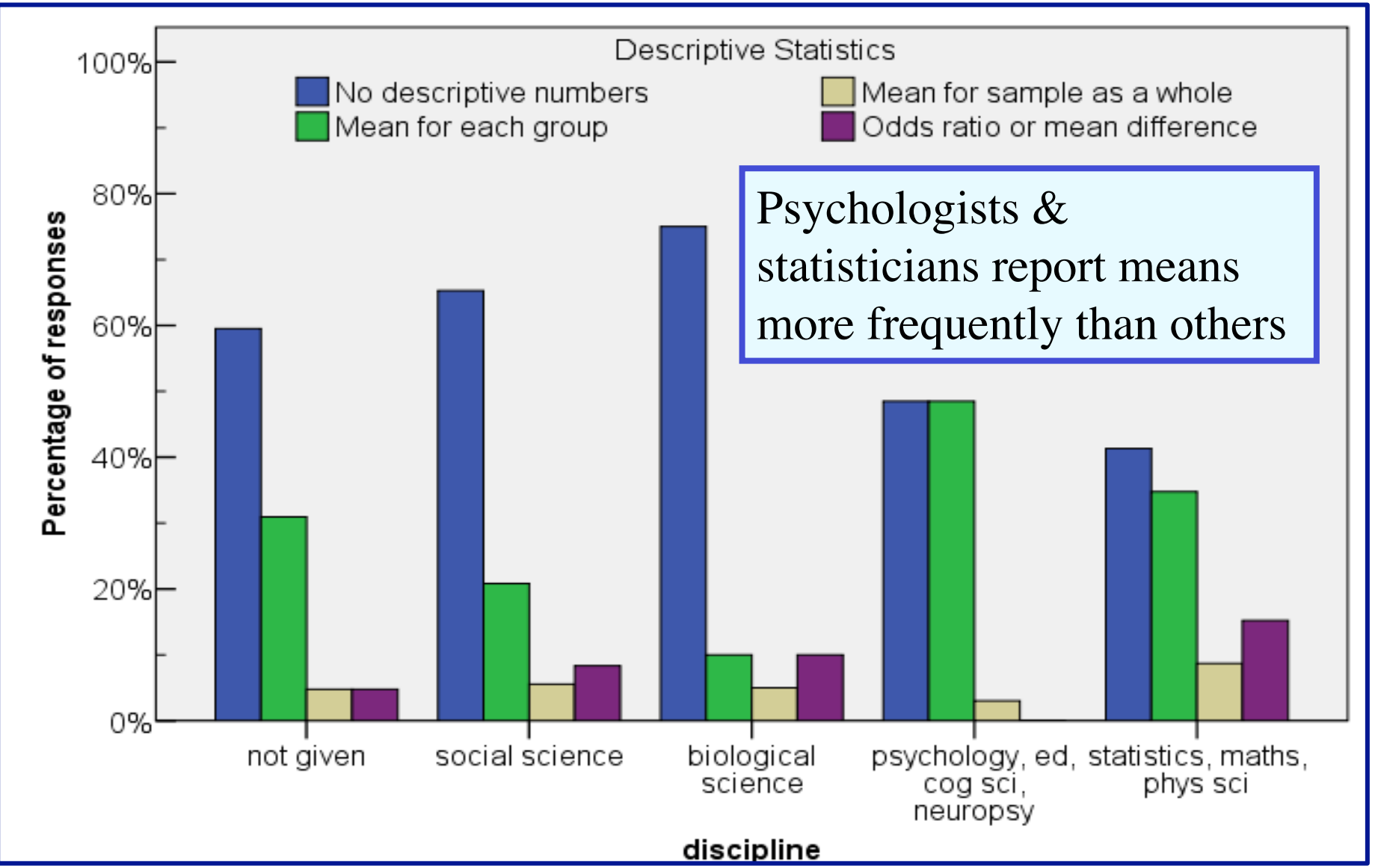
Summary: All Participants

- Inferential Statistics: Hypothesis Tests
  - Majority DO report "no significant" effect (85%)
- Descriptive Statistics
  - Majority do NOT report anything (73%)
- Sample size, N
  - Most do NOT report (76%)
- Sufficiency for Replication/Meta-analysis
  - Most do NOT give sufficient information (86%)
- Predictions of future: NO CONSENSUS
  - 38% predict replication as combined group
  - 32% predict replication as separate groups
  - 30% say no predictions possible

Free Form Scientific Report Expertise Effects

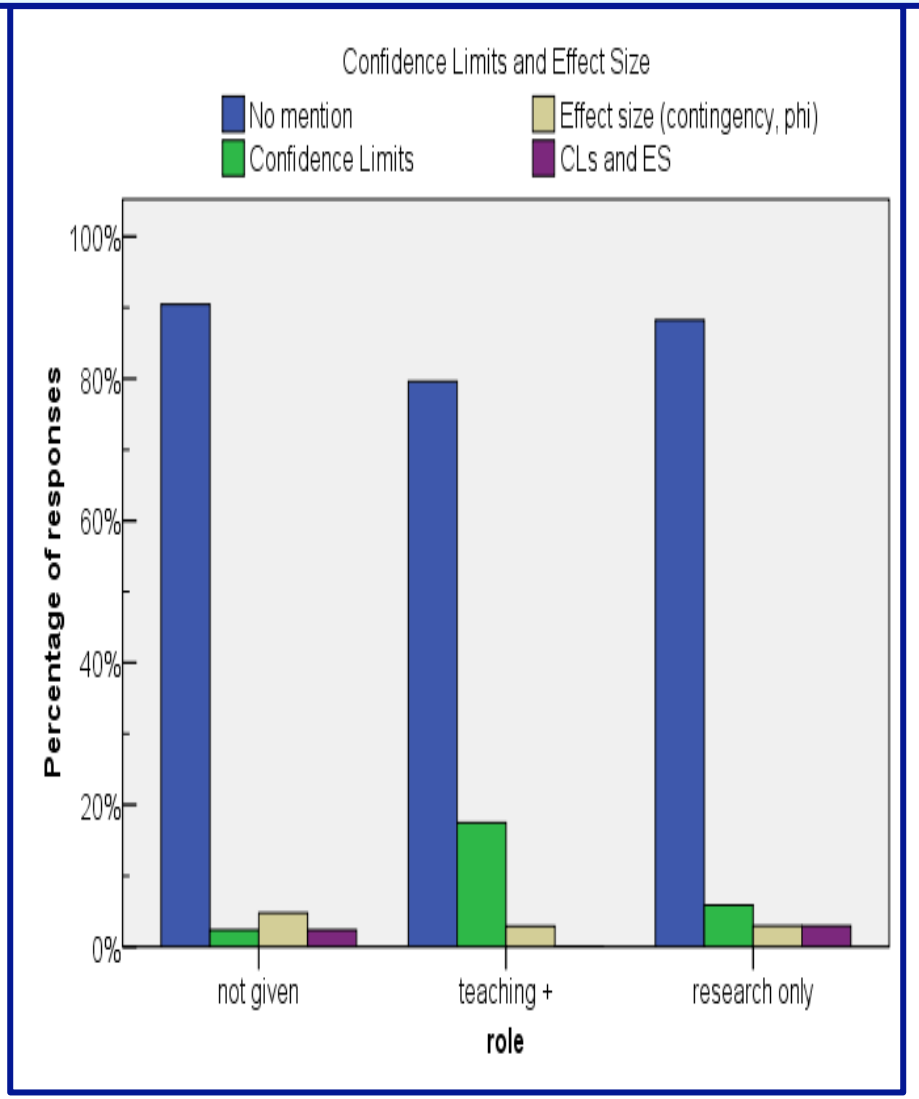
- **Descriptive statistics**
  - **Psychologists & statisticians report more often than other disciplines**
- **Confidence Levels**
  - **Those with role that includes teaching report more frequently than other roles**
- **Odds Ratios**
  - **Those with role that includes teaching report more frequently than other roles**

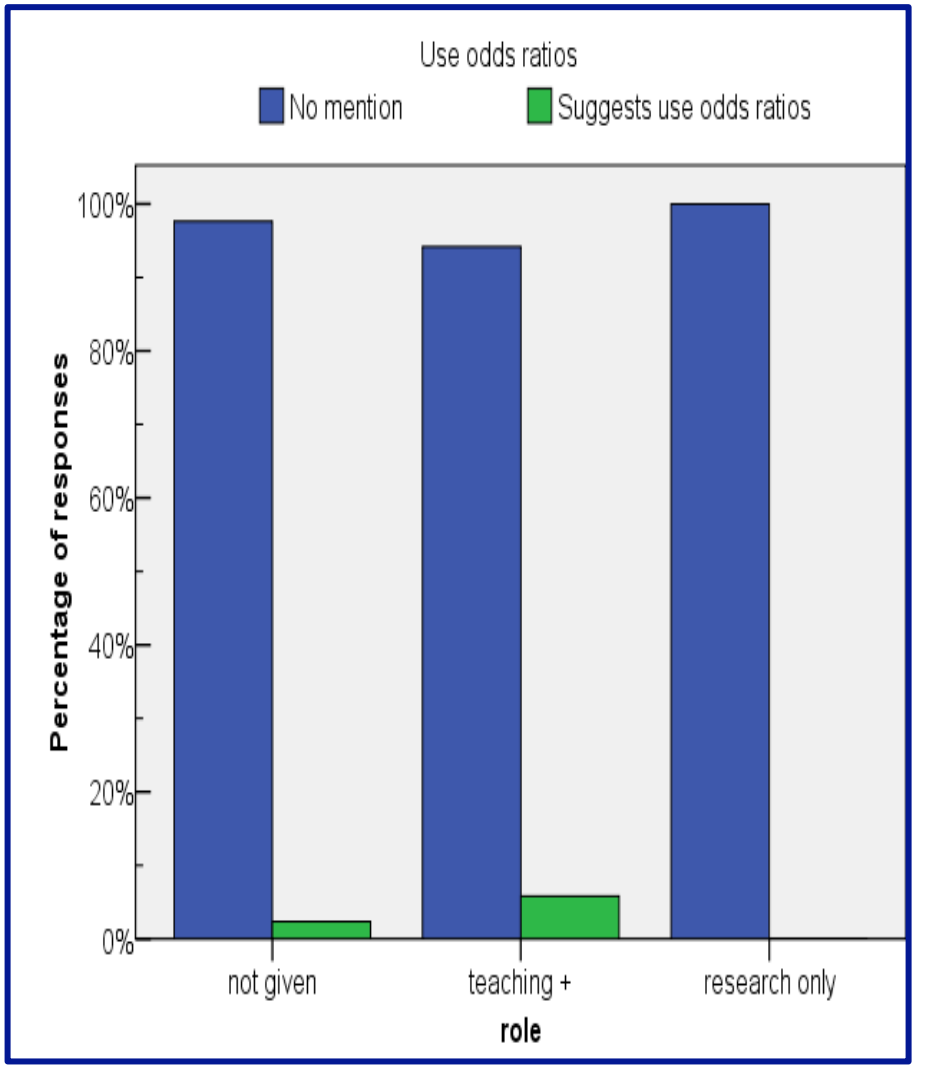
Free Form Scientific Report
Descriptive Statistics by Discipline



Free Form Scientific Report
Role & effect size/confidence levels (left); odds ratio (right)

Those with teaching+ role more frequently report both CL and OR





Summary: Expertise Effects

- **Descriptive Statistics**
  - Statisticians, 59%, & psychologists, 51%,more frequently reported some descriptive statistic than biologists, 25%, or social scientists, 35%.
- **Confidence Intervals**
  - Those with teaching+ role were more likely to recommend CLs, 20%, than others,10%.
- **Odds Ratio**
  - Those with teaching+ role were more likely to recommend ORs, 6%, than others, 1%.

Main Negative Criticism of Design

- **Design used convenience sample**
  - So no inference possible for *general* population
- **Hindsight… use more specific story**

  Community health team considering Health Clinic
  Intention: identify people 'at risk'
  So, estimate the potential demand for men & women
- **Highlight well known issues**

  1.Purpose needed to interpret results

  2.If no purpose, people make assumptions

Other Criticisms of Scenario Design

- Dichotomizing continuous variable
  - True, but wanted a 'simple problem
- Low power
  - Not *that* low.
  - Power for 10% difference at.05 level > .88

    With community health scenario

    10% difference justifies separate materials for men & women

Dissemination Suggestions

- Hard Problem!
  - Guidelines for significant effects widely flouted!
- Journals
  - Guidelines: Clearly insufficient
  - Policy: Needs to be enforced
- Grant awarding bodies
  - Some, require a priori power analyses
  - Modest carry over into reporting results
- Scientific and statistical education
  - Awkward, if flouted by recommended reading!
  - Library of good and bad examples?

Summary

- Scientists do NOT ‘naturally’ follow guidelines
- Inference results *still* reported without any descriptive statistics
- Power is not a salient issue for NS reporting
- No consensus on future predictions

Implications for NS Guidelines

- **Guidelines are needed**
  - Always include N
  - Support replication/meta analysis
- **Power is *always* crucial**
  - Numeric scientific or practical importance (e.g. number of lives saved) is highly desirable
  - Otherwise, give power for statistically specified effect sizes. Conventional ‘large’, ‘medium’, ‘small’ useful