

# Communicating Meaningful Results: One Statistical Agency's Explorations

Marilyn M Seastrom  
Chief Statistician  
National Center for  
Education Statistics (NCES)  
Marilyn,Seastrom@ed.gov

Disclaimer: This paper is intended to promote the exchange of information. The views expressed do not necessarily reflect the position of the U.S. Department of Education.

## Overview

- Background
- Current State—Where Are We Now
- ASA Statement on  $p$ -Values
- NISS Expert Panels on Citing Significance in NCES Data Reporting
- Potential Steps Forward

## Background

NCES has struggled with this issue over the last 35 years. From calls in the mid-1980's for products to have both utility and quality guided by a written set of Statistical Standards, to specific parameters placed on the reporting of statistical findings in NCES Statistical Standards published in 1987, 1992, 2003, and 2012; NCES has struggled with the topic of how best to use the field of statistics to support the description and analysis of NCES data. Over the years, different approaches have been proposed, discussed, and tried--to no avail. The recent discussions within the profession over whether or not we should continue our devotion to significance tests with a p value of .05 has opened up a new platform for NCES to use in identifying a way forward. There are hints of movement in this direction on the part of NCES as early as the 2002 NCES Statistical Standards. More recently NCES held a seminar on the role of a .05 p-value and convened two NISS panels to review NCES practices and provide ideas for a path forward. Our early efforts in this direction, the recent panel results, and possible steps forward will be discussed in this presentation.

[Back](#)

## Current State—Where Are We Now

### TIMELINE:

#### **2002 NCES Statistical Standards and Guidelines underwent major revisions**

- 50% of NCES staff actively participated
- Input was sought from contractors
- Reviewed by NISS Expert Panel
- 60-day public comment period under the Information Quality Guidelines Act

#### **2006 OMB issued Standards and Guidelines for Statistical Surveys**

#### **2012 NCES Statistical Standards and Guidelines underwent minor revisions to harmonize with OMB Standards**

# NCES Current Practices

- Several NCES Standards are relevant to the ASA Statement and the deliberations of the NISS Expert Panels.
  - **5-1: STATISTICAL ANALYSIS, INFERENCE, AND COMPARISON NCES**
  - **5-1-1:** Statistical analyses must be approached from an analysis plan . . .
  - **5-1-3:** The criterion for judging statistical significance in all reported hypothesis tests should not exceed  $\alpha = 0.05$  (0.95 for confidence intervals). Reports will indicate an observed difference as statistically significant when an appropriate hypothesis test rejects the null hypothesis at  $\alpha = 0.05$  or less.
  - **5-1-4:** Reported analyses must focus on differences that are substantively important
  - **5-1-5:** Failure to reject the null hypothesis does not imply acceptance of the null hypothesis. When the null hypothesis is not rejected, reporting options are available:
- Despite the consensus process used and the staff support for the Standards when they were introduced, the success in the implementation of individual Standards and Guidelines has been mixed.

# ASA Statement on p-Values

**“ASA believes that the scientific Community could benefit from a formal statement clarifying widely agreed upon principles regarding the proper use and interpretation of *p*-values.”**

**ASA concluded that “No single index should substitute for scientific reasoning.” Good statistical practice, and good scientific practice emphasize the use of:**

- A. Good study design and conduct
- B. A variety of numerical and graphical summaries of data
- C. An understanding of the phenomenon under study
- D. An interpretation of the results in context
- E. Complete reporting
- F. Proper logical and quantitative understanding of what data summaries mean

# **NISS Expert Panels on the Use of Significance Testing in NCES Reports**

**Panel 1: Significance and Non-Significance in NCES Reports**

**Panel 2: Citing Significance in NCES Data Reporting**

**The importance ASA places on moving away from a heavy dependence on the bright line of a dichotomous significant vs not significant decision based on  $p \leq 0.05$  is an essential component of the NISS panels' work.**

# Potential Steps Forward

Some of the conclusions and recommendations from the ASA Statement and the NISS Expert Panels are addressed in the NCES Statistical Standards

- The requirements specified in Standards 1-1, 2-1-1, 5-1 and 5-3-1 are consistent with the conclusions and recommendations
- The Standard 5-3-1 requirement to test all hypotheses for statistical significance using  $p \leq 0.05$  should be revisited considering the conclusions and recommendations.

Other conclusions and recommendations provide topics for consideration when revising the NCES Statistical Standards and Guidelines.



# Continue to Follow Existing NCES Standards and Guidelines

- Specify analytic goals during the study design ([Standard 2-1-1](#), part 1)
  - Include a detailed discussion of the goals and objectives of the survey or survey system, identify the information needs that will be met, describe the content areas, target population(s), and analytic goals.
- Develop an analysis plan as part of the study design ([Standards 1-1](#), part 5 and [2-1-1](#), part 4)
  - Identify analysis issues, objectives, key variables (including definitions), minimum substantively significant effect sizes (for the total population and for key reporting variables by reporting domains) and proposed statistical techniques.
- Ensure that statistical analyses, comparisons, and inferences included in NCES products focus on substantively important findings and are based on appropriate statistical procedures ([Standard 5-1](#))

# Explore Recommendations for Change – Slide 1

- Revisit requirement for [Standard 5-3-1](#) for judging statistical significance in all reported hypothesis tests using  $p \leq 0.05$ ; use confidence intervals; consider importance and relevance of findings (NISS Panels and ASA Statement)
- Pose questions of **how** (e.g., how big, how different, how much), **what** (e.g., what distinguishes one group from another), **which** (e.g., which factors are important) rather than questions of is and does that elicit binary yes/no question (NISS Expert Panel 2).
- Focus findings on magnitude and uncertainty (NISS Expert Panels)
  - Magnitude
    - Defined as what is meaningful and “worthy of serious consideration.”
    - Estimates, differences, relative contributions of factors to outcomes or model.
  - Uncertainty
    - Essential to understanding magnitude
    - Multiple possible presentations depending on the methodology used to estimate magnitude –standard error, coefficient of variation, log odds, likelihood ratio,  $R^2$ , Bayes’ posterior distribution

# Explore Recommendations for Change – Slide 2

- Analysis and reporting should be driven by analysis plans, not by testing many comparisons and “fishing” for significant tests (NISS Expert Panels)
  - Avoid cherry-picking—aka “p-hacking”
- Implement complete reporting of ALL planned analyses, regardless of outcomes—report every test conducted (initial planned analysis, secondary analysis, and exploratory analysis (ASA statement and NISS Expert Panels)
- Use a variety of numerical and graphic summaries of data (ASA statement and NISS Expert Panels)
  - Well chosen data visualizations can help readers understand findings
  - Revise NCES Standard on Tabular and Graphic displays—update with recent data visualization concepts

# Explore Recommendations for Change – Slide 3

- Expand recent steps toward multi-tiered reports (NISS Expert Panels)
  - Accessible reporting that is credible at multiple technical levels from non-technical for the general public to technically clear for researchers in academia and outside
  - Take advantage of electronic reporting to allow readers to drill down to more detail
- Revisit adjustments for multiple comparisons; explore using the False Discovery Rate (FDR) as an alternative to Bonferroni. (NISS Expert Panel 1)
- Use appropriate tests for trends, not sequential *t*-tests. (NISS Expert Panel 1)
- Explore and present analysis of interactions between variables. (NISS Expert Panel 1)
- Explore the use of multivariate analysis. (NISS Expert Panel 1)

[Back](#)

Thank You!

Marilyn Seastrom, Ph.D.

Chief Statistician, NCES

Director, Statistical Standards and Data Confidentiality Staff

[Marilyn.Seastrom@ed.gov](mailto:Marilyn.Seastrom@ed.gov)

