ExxonMobil

March 18, 2015

Implications for Benefit Estimation of Uncontrolled Variance in Spirometric Studies of Air Pollution

Energy lives here™

Richard B. Belzer
R. Jeffrey Lewis
Ammie M. Bachman

This presentation includes forward-looking statements. Actual future conditions (including economic conditions, energy demand, and energy supply) could differ materially due to changes in technology, the development of new supply sources, political events, demographic changes, and other factors discussed herein (and in Item 1A of ExxonMobil's latest report on Form 10-K or information set forth under "factors affecting future results" on the "investors" page of our website at www.exxonmobil.com). This material is not to be reproduced without the permission of Exxon Mobil Corporation.

Personal Spirometer Manufacturer's Protocol for Interpreting Spirometric Results

Traffic Light Scheme for Classifying PEF Abnormalities	
PEF as Percent of Best	*Traffic Light and Interpretation*
> 80%	Green 'OK'
50 – 80%	Yellow 'Caution'
< 50%	Red 'Danger'

Source: Manufacturer's Manual and Instructions

ATS Protocol for Interpreting Spirometric Results

Severity of Spirometric Abnormality Based on FEV_1		
Degree of Severity	*FEV_1 as Percent of Predicted*	*FEV_1 60-year old Male 5'8" (3.9 L)*
Mild	> 70%	> 2.7 L
Moderate	60-69%	> 2.4 L
Moderately severe	50-59%	> 2.0 L
Severe	35-49%	> 1.4 L
Very severe	< 35%	< 1.4 L

Source: Pellegrino et al. 2005 (p. 957)

Spirometry Example

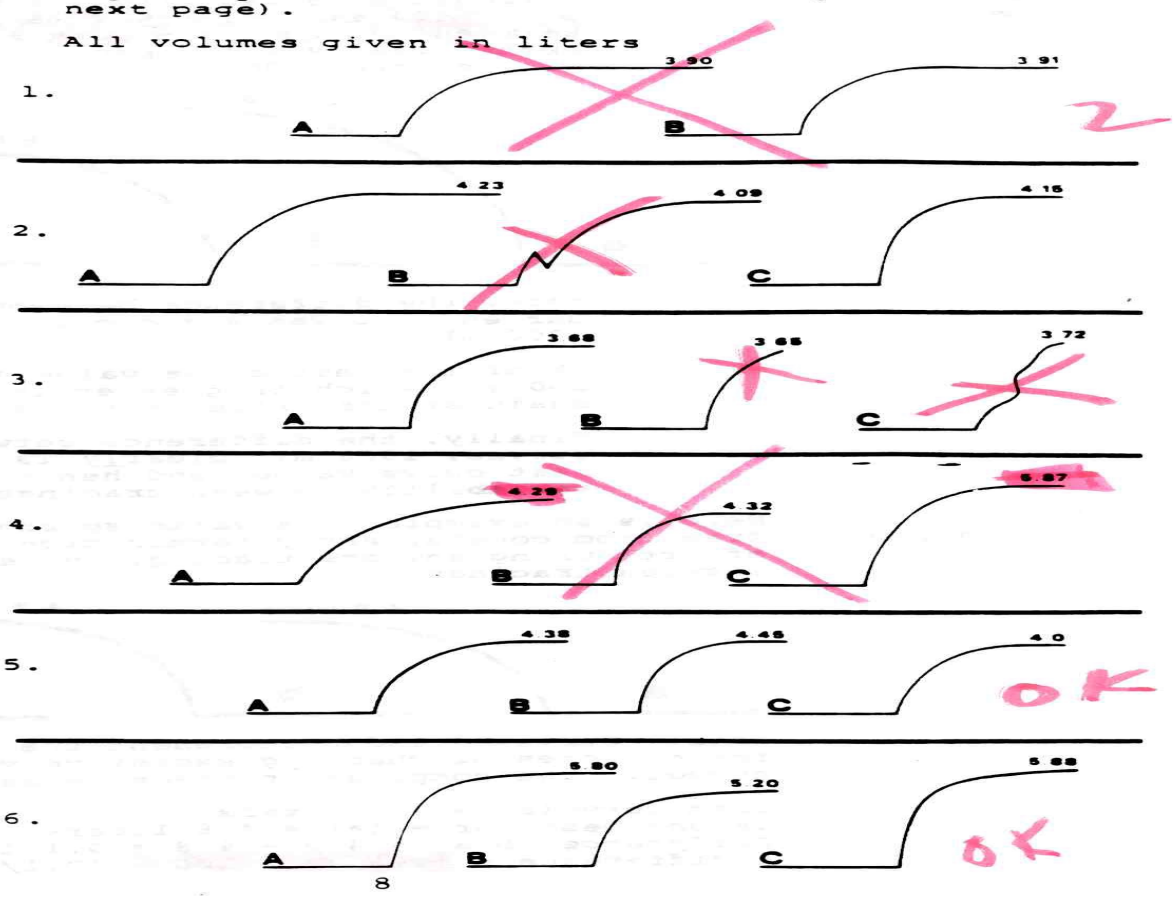
UNIT I: SPIROMETRIC TECHNIQUE

EXERCISE #1

Determining validity of spirograms.

Which of the following represent valid spirograms? Explain your choices. (Feedback is provided on the next page).

All volumes given in liters

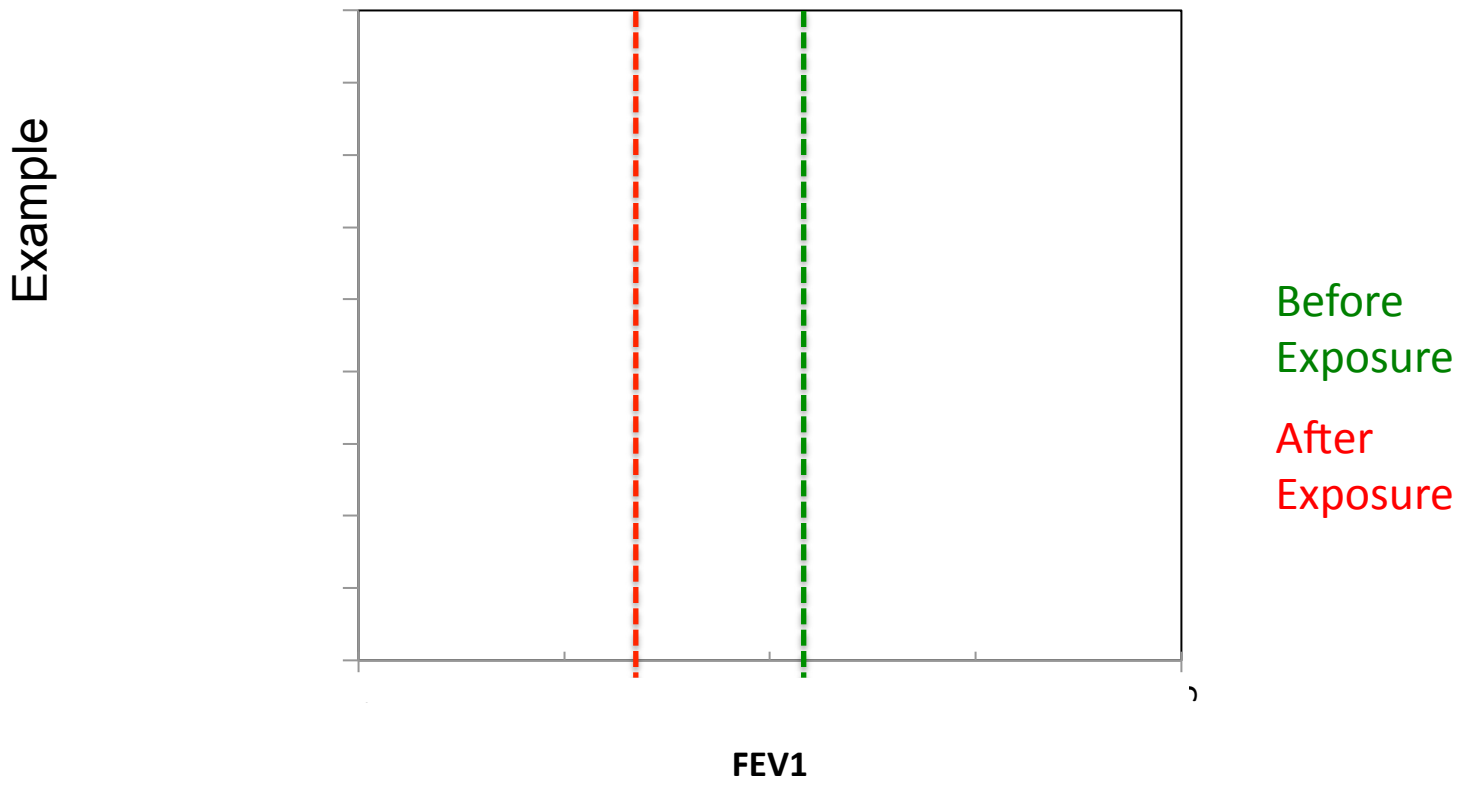


8

Source: NIOSH, 1989

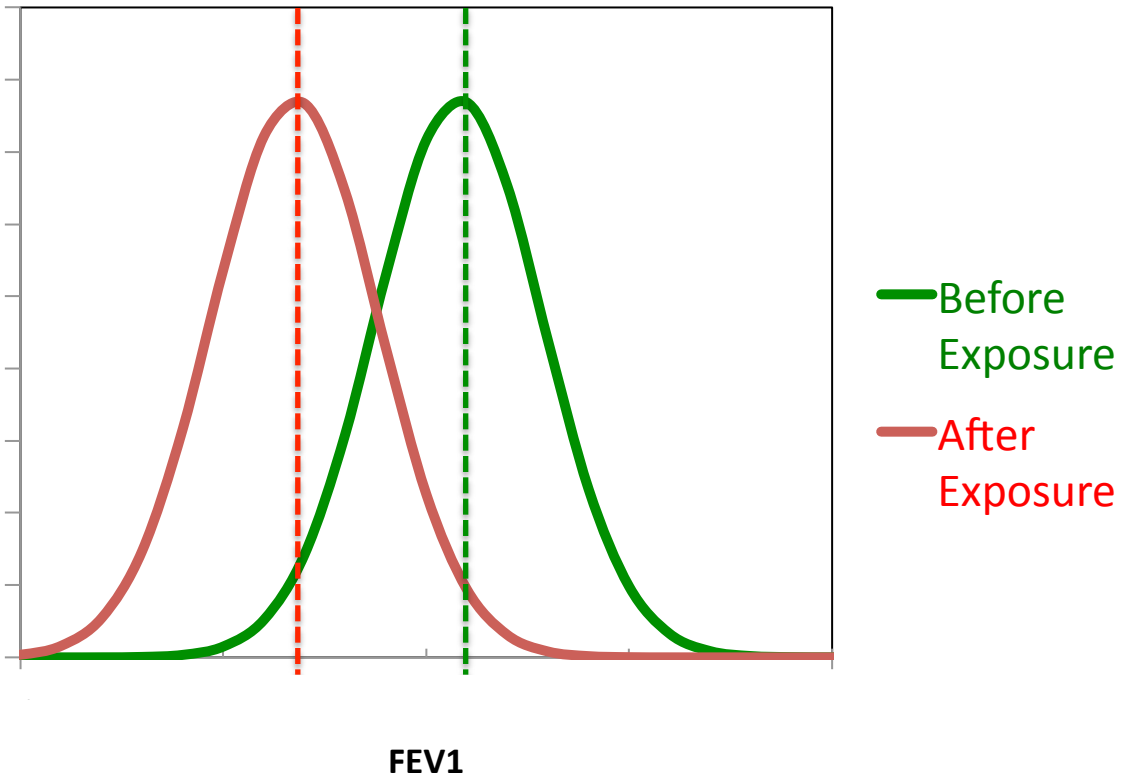
ExxonMobil

Air Pollution Study Y Shows *x*% Average FEV_1 Reduction After Exposure

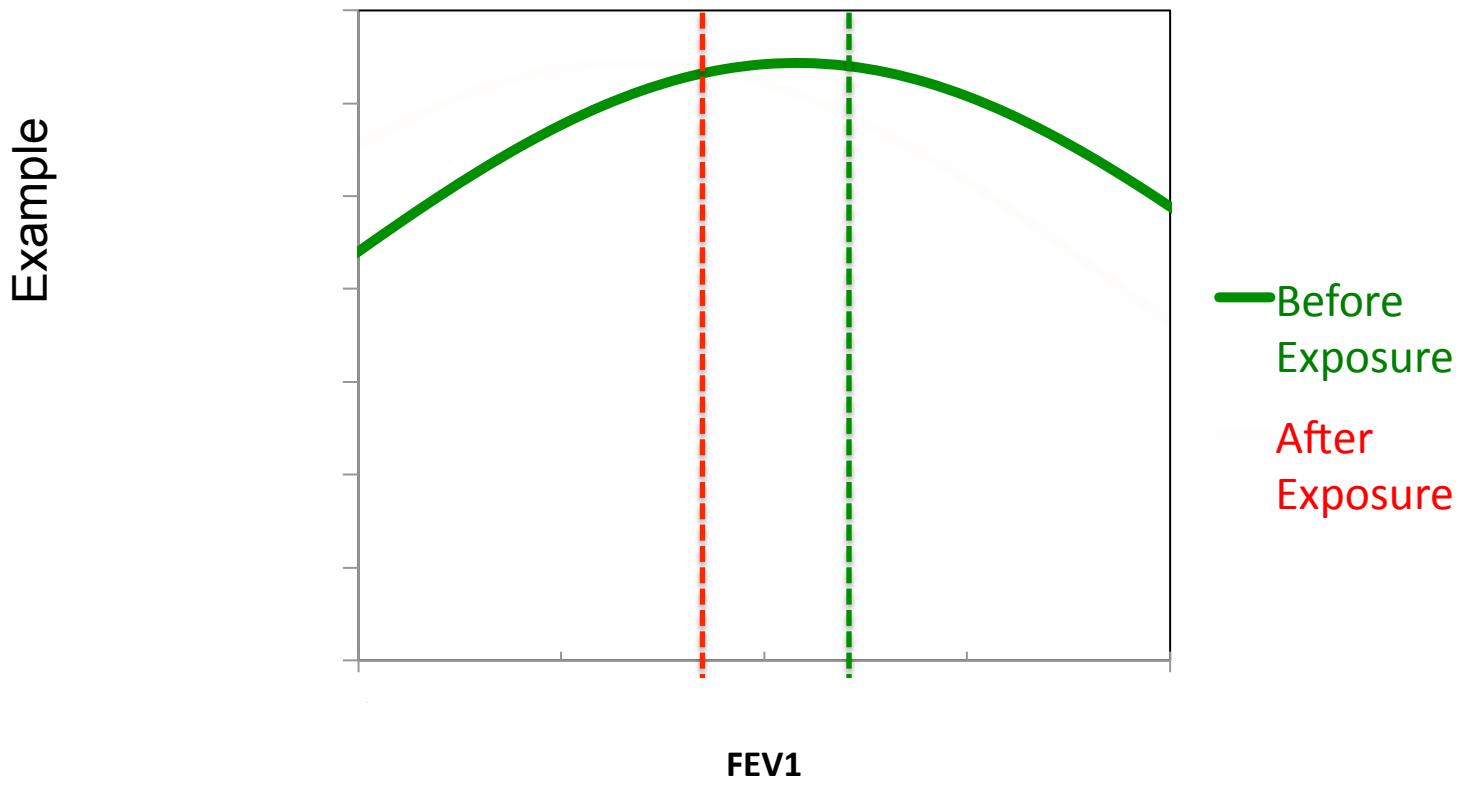


Difference Is Assumed to Reflect *x*% Leftward Shift in FEV_1 Distribution

Example

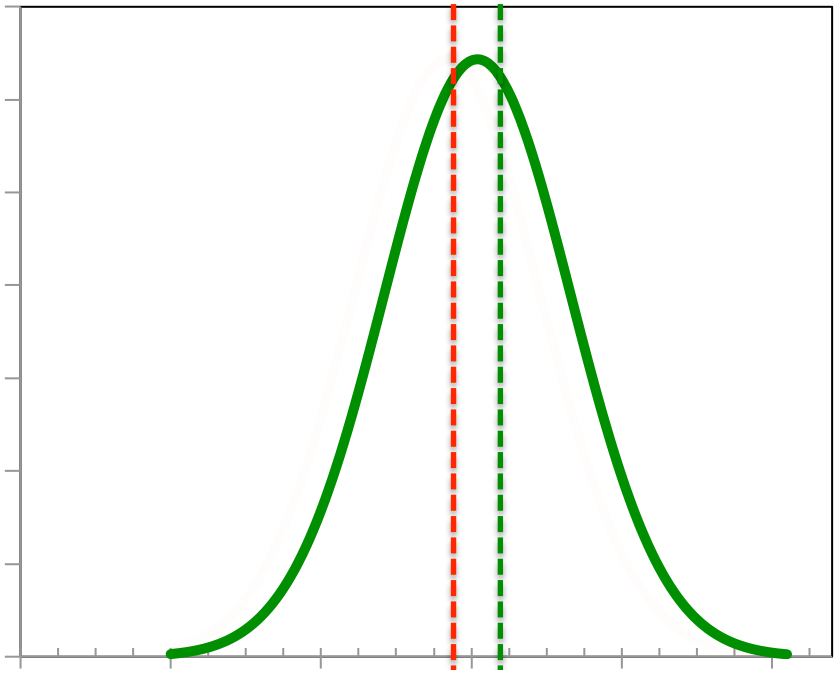


Averages *x*% Apart Could Be Random Differences from a Single, More Variable Distribution



In Which Case They Could Look Like This

Example



FEV1

Before Exposure

After Exposure

Sources of Potential Bias in Spirometry

Source	Comments
Non-representative samples	Accidental, convenience or ‘grab’ samples generally are unrepresentative
Non-response bias	Non-responders and dropouts usually differ from responders
Censoring	Rules for data acceptance may exclude valid data asymmetrically
Strategic behavior	Examiner and/or subject may know the exposure and/or research hypothesis of interest, and this knowledge may affect conduct

Sources of Inter-Subject Variability in Spirometry

Source	Comments
Protocol effects	Protocol may differ or be unequally followed
Device effects	Spirometers vary in performance
Examiner effects	Variability in skill, experience, professional judgment, adherence to external protocols
Subject effects	• Personal characteristics (age, sex, height) • Variability in skill, experience, effort • Interaction with devices
Setting effects	Setting (lab, physician's office, field, home) may vary
Subject/device interaction	• Subjects vary in capacity to use technology • Some technologies produce unreliable data
Exposure to allergens and other confounders	Co-exposures may be important
Test failure	Failure to complete test may provide important data

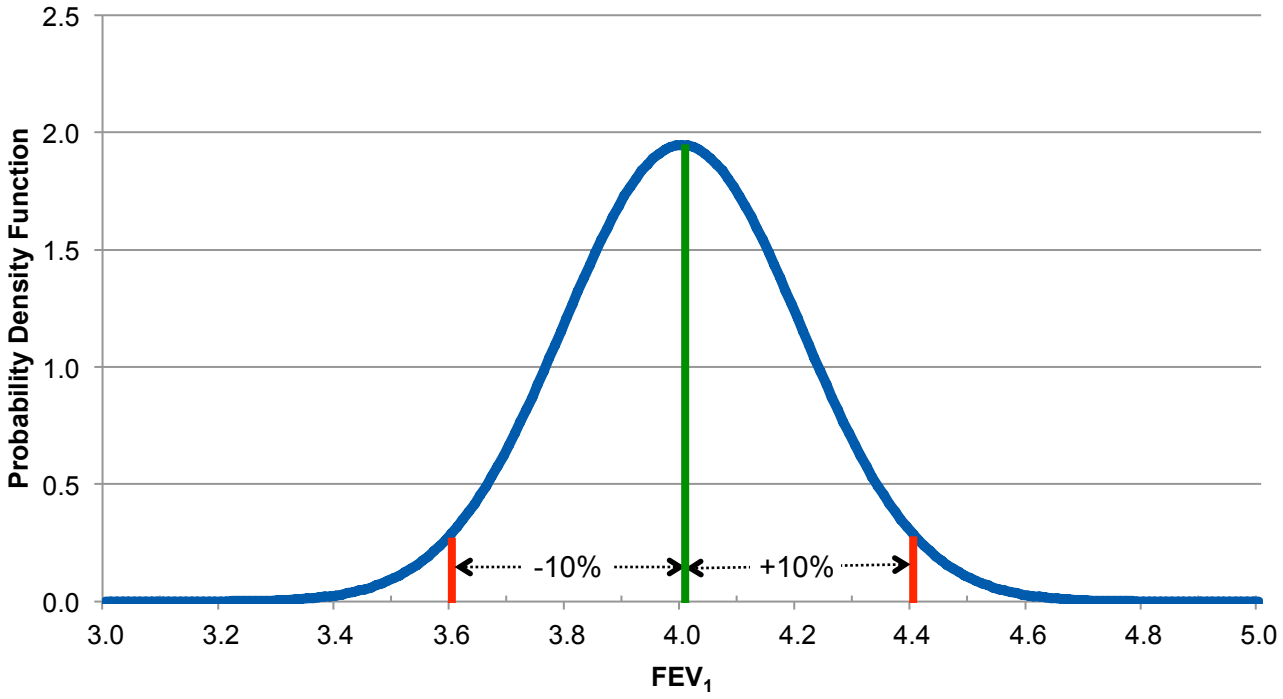
Sources of Intra-Subject Variability in Spirometry

Source	Comments
Age, height and sex	• Proportional effects of pollutants are controlled when before and after data are obtained from same subjects • Effects may not be proportional
Test order	• Spirometric tests may induce bronchospasm • Other pulmonary function tests may influence spirometry
Time of day	Peak function around noon, declines thereafter
Duration of testing	Some studies collect data over many months
Seasonal effects	Winter/summer, dry/humid, ozone/non-ozone
Number of maneuvers per test	• Chamber studies (typically 2-4) • Observational epidemiology (2-3) • Clinical setting (3-8)
Inter-maneuver variability	Mean or maximum value per test is recorded; other data are discarded

Examples Showing Why Unaccounted Variability Is a Serious Problem in Air Pollution Studies

- Device accuracy
  - Air pollution studies assume no variability.
  - This is wrong because
    - Personal devices typically are accurate ± 10% (ATS 1994 protocol)
    - Professional devices may be accurate ± 3% (ATS 2005 protocol)
- Inter-maneuver variability
  - Air pollution studies assume no inter-maneuver variability
  - This is wrong because
    - 2-8 maneuvers are performed
    - Keeping the average or maximum of 2-8 maneuvers censors data, discards variance

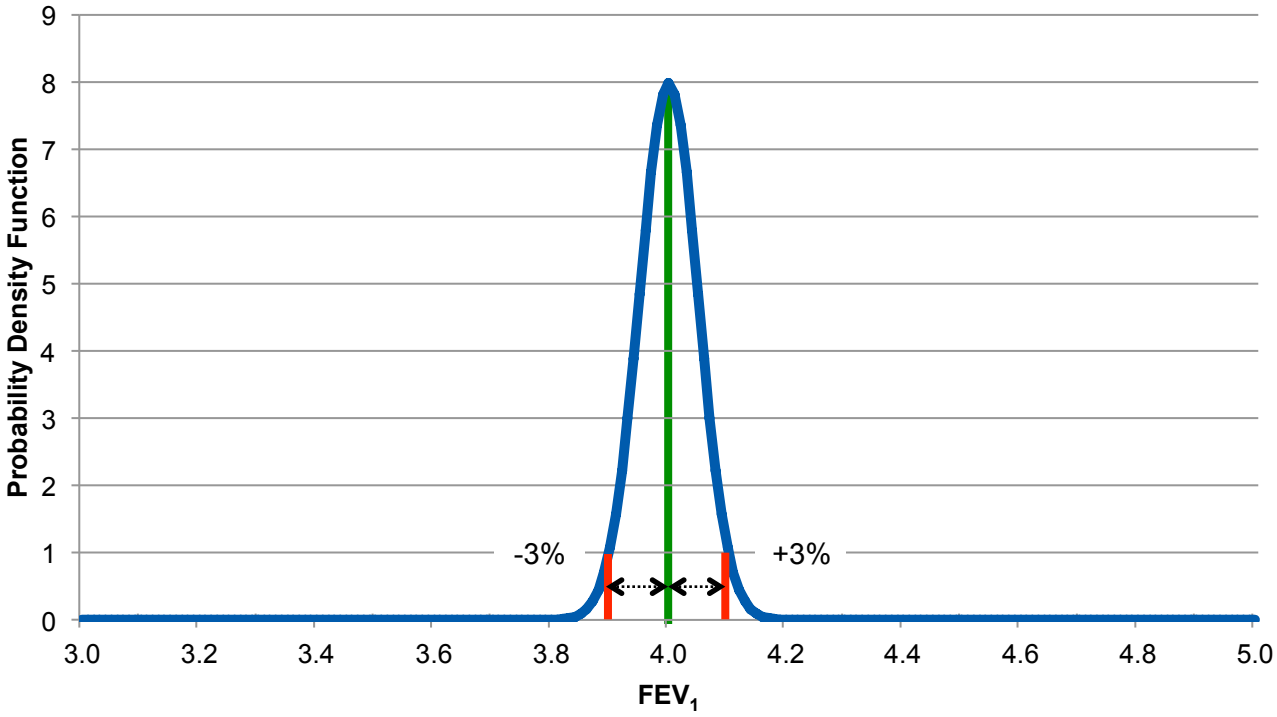
Spirometric Device Accuracy ± 10%



ATS 1994

ExxonMobil

Spirometric Device Accuracy ± 3%



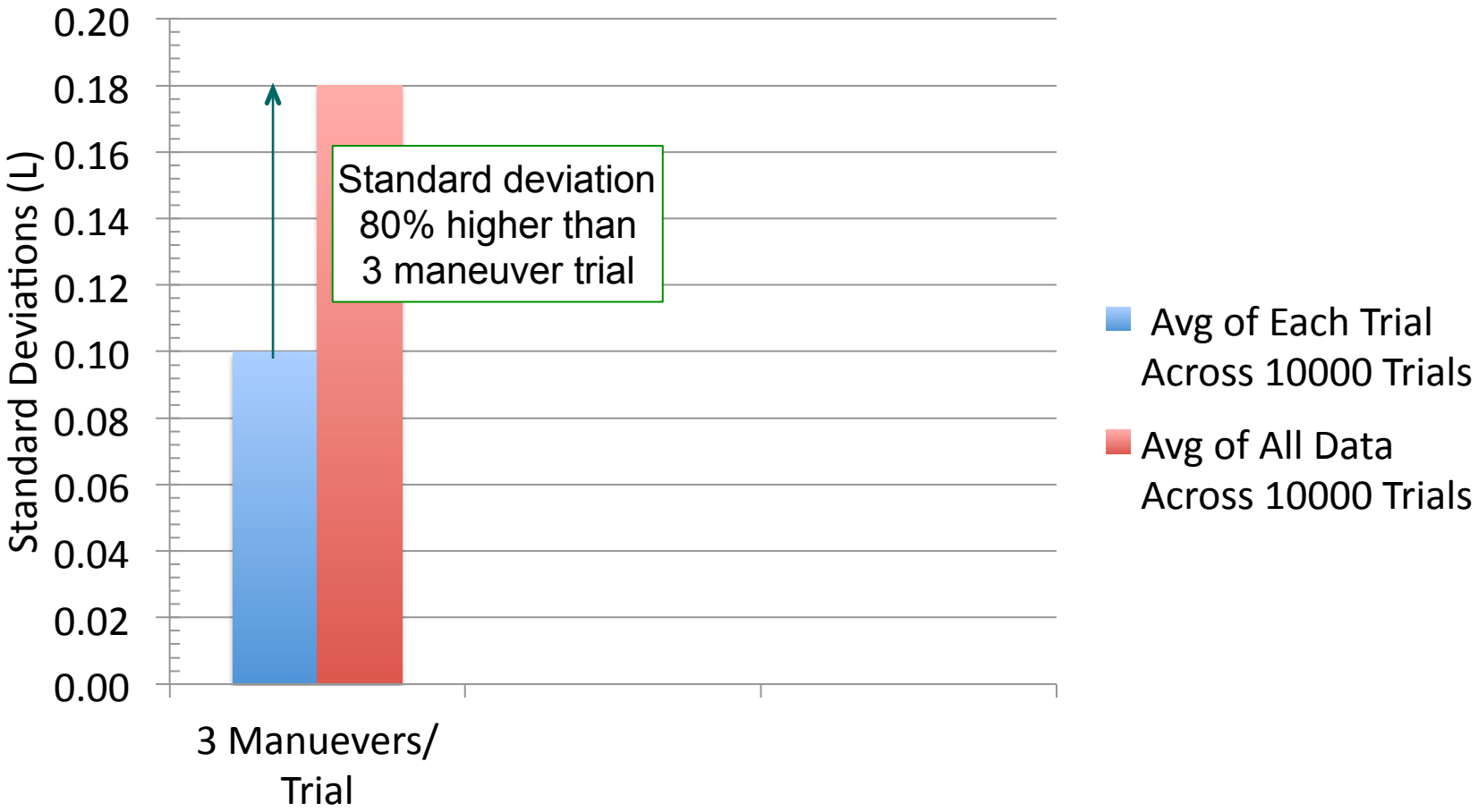
ATS 2005

ExxonMobil

How Large is Inter-Maneuver Variability?

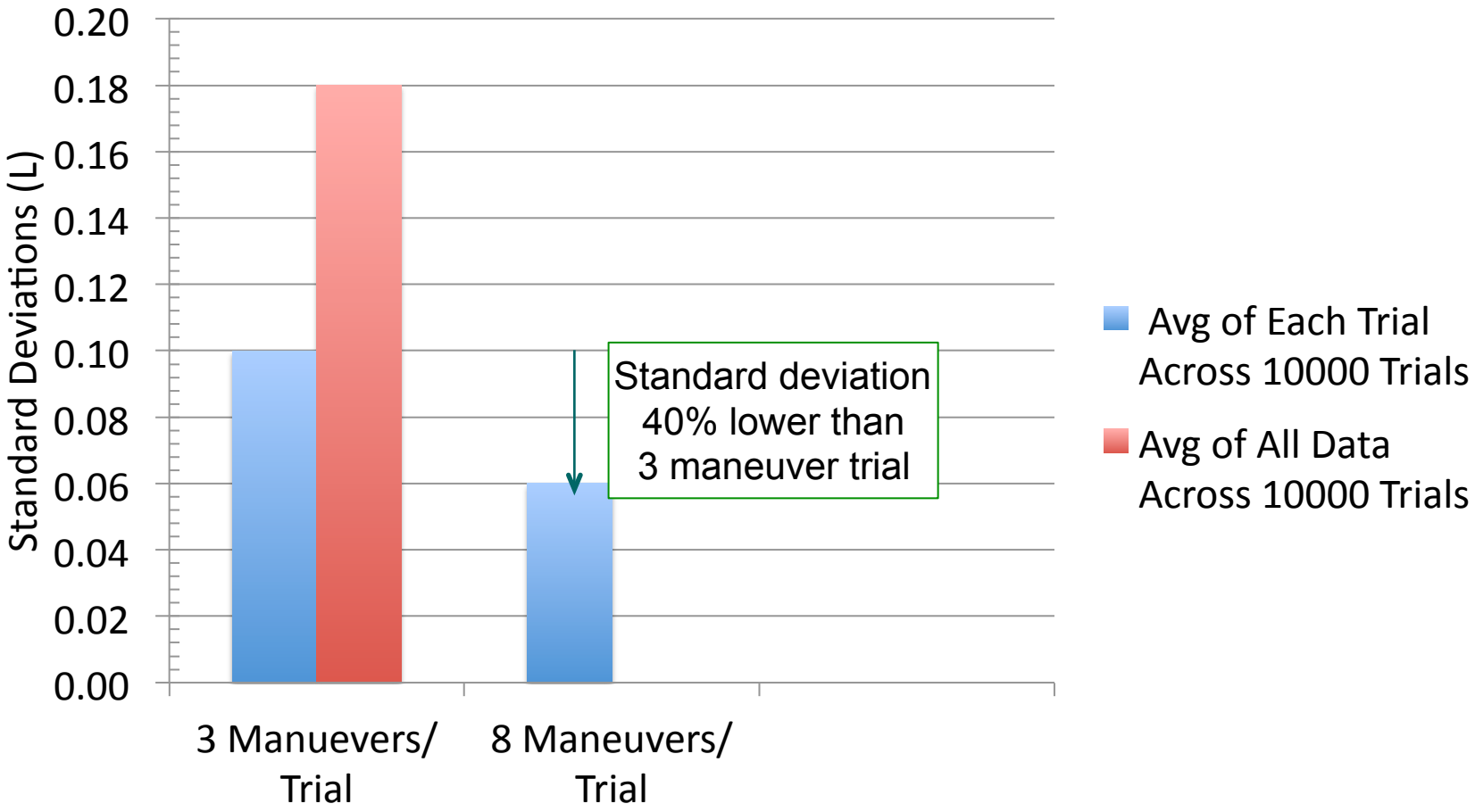
- An exploratory data analysis using co-author as guinea pig
  - Perform 16 tests, each with 8 maneuvers
  - Calculate grand mean (2.49 L) and standard deviation (0.18 L)
  - Perform simulation of 10,000 trials using these parameters
- Scenario #1: Three maneuvers
  - Standard deviation of 10,000 means of 3 maneuvers = 0.10 L
  - Standard deviation of 30,000 maneuvers = 0.18 L (80% higher)
- Scenario #2: Eight maneuvers
  - Standard deviation of 10,000 means of 8 maneuvers = 0.06 L
  - Standard deviation of 80,000 maneuvers = 0.18 L (200% higher)
  - What does it mean?
    - Using the mean to represent each test censors data and discards inter-maneuver variance
    - Statistical significance may incorrectly be inferred

Using All 30,000 Maneuvers Yields Standard Deviation 80% Greater than Using Only Means



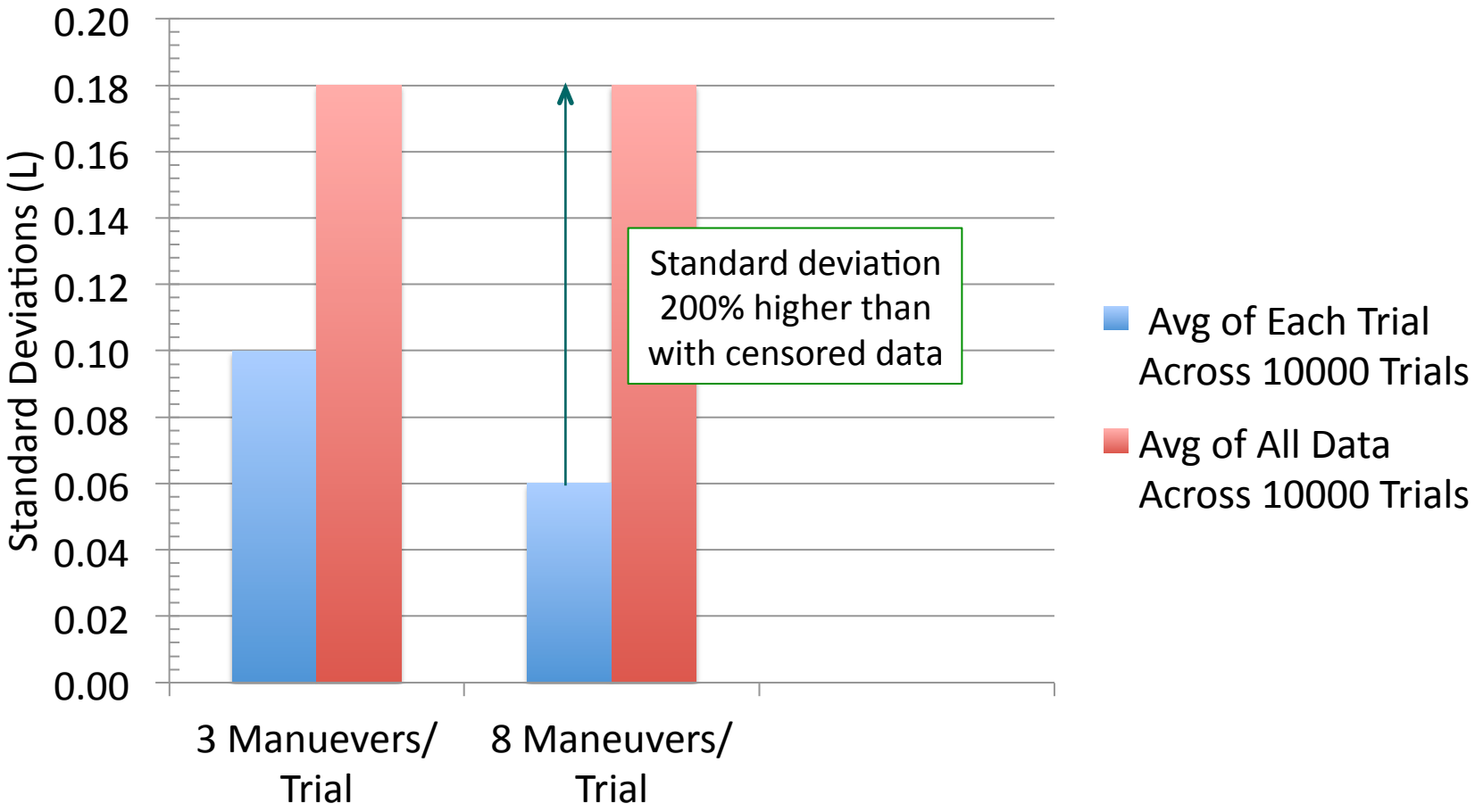
ExxonMobil

More Maneuvers Reduces the Standard Deviation Even More When Using Only Means



ExxonMobil

Using All 80,000 Maneuvers Yields Standard Deviation 200% Greater than Using Only Means



Take Home Lessons About Spirometric Studies of Air Pollution

- There are many sources of variability, few of which appear to be controlled.
- Differences may be incorrectly characterized as statistically significant because variability was discarded.
- Benefit estimates derived from air pollution studies using spirometric data are similarly dubious.

Key Research Questions

- If device accuracy or inter-maneuver variance is taken into account:
  - How likely is it that a 5% decrement in FEV1 after exposure would occur by chance?
  - What about a 10% decrement?
  - More generally, how large does a decrement have to be before it can be statistically significant?
- How reliable are the results of air pollution research that fails to take account of known sources of variance?

Is Unaccounted Variance a New Problem?

Three Little Pigs, 1933. Walt Disney Company.

References

- Hayes GB, Christiani DC. 1993. Measures of small airways disease as predictors of chronic obstructive pulmonary disease. Occup Med: State Art Revs 8(2): 375-395.
- Miller MR, Hankinson J, Brusasco V, et al. 2005. Standardisation of spirometry. *Eur Respir J*, *26*(2), 319-38.
- National Institute of Occupational Safety and Health (NIOSH). 1989. NIOSH spirometry training guide. NIOSH, Morgantown, WV.

  Pellegrino R, Viegi G, Brusasco V, et al. 2005. Interpretative strategies for lung function tests. *Eur Respir J* 26(2): 948-68.

Questions?