ExxonMobil

March 17, 2017

RISKS AND BENEFITS OF REDUCING OZONE EXPOSURE AFTER RE-EVALUATING A RECENT CHAMBER STUDY

Energy lives here™

R. Jeffrey Lewis, PhD, MBA
Richard B. Belzer, PhD, MS, MPP

Society for Benefit-Cost Analysis Annual Meeting
March 16-17, 2017
Washington, DC

Background

- NAAQS are set based on small percentage differences in pulmonary function tests conducted in chamber studies
- All such tests assume that data are fixed, with no within-person variability.
  - Within-person inter-test variability is a known phenomenon and is sometimes subjected to statistical control
  - Within-person intra-test variability is a known phenomenon and is ignored
- Failure to account for within-person variability in established test protocols may generate measurement error
- Measurement error is potentially large relative to changes described as statistically significant and deemed biologically meaningful

Background….link to benefit-cost analysis

- The O_3 NAAQS benefit-cost analysis is based on epidemiological endpoints for which reduced FEV_1 are essential prerequisites
  - e.g., asthma hospitalizations, asthma exacerbation
- The key evidence for FEV_1 reduction is a human chamber study by Schelegle et al. 2009
  - 8-hour average 70 ppb O_3 reported to induce statistically significant FEV_1 decrements in healthy young adults
- Estimated benefits from avoiding similar O_3 concentrations require that observed epidemiological endpoints be caused by O_3 exposure
  - This causal nexus requires at a minimum that FEV_1 decrements be statistically significant

Key Lessons About Measurement Error from Previous Spirometry Simulation

1. Neither inter- nor intra-test variability are accounted for in chamber studies or observational epidemiology used to define risk and estimate benefits

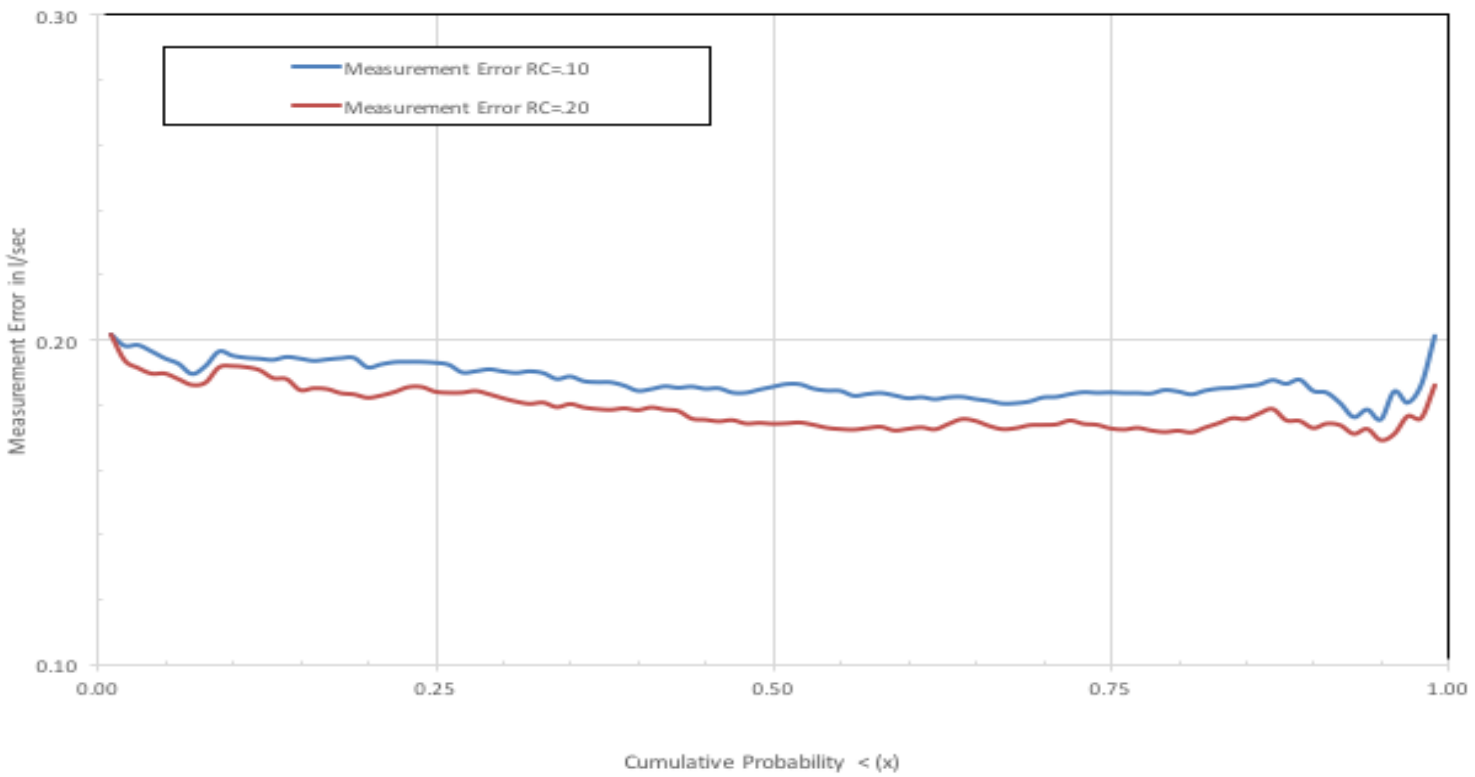
2. ATS* protocol with repeatability criterion and early test termination
   a. Designed for clinical application (not research)
   b. Does not account at all for inter-test variability
   c. Prevents collection of data needed to account for intra-test variability
   d. Discards valid data containing intra-test variability

3. Choice of repeatability criterion has little effect on measurement error

4. Differences are incorrectly characterized as statistically significant

*ATS=American Thoracic Society (1979, 1987, 1994)

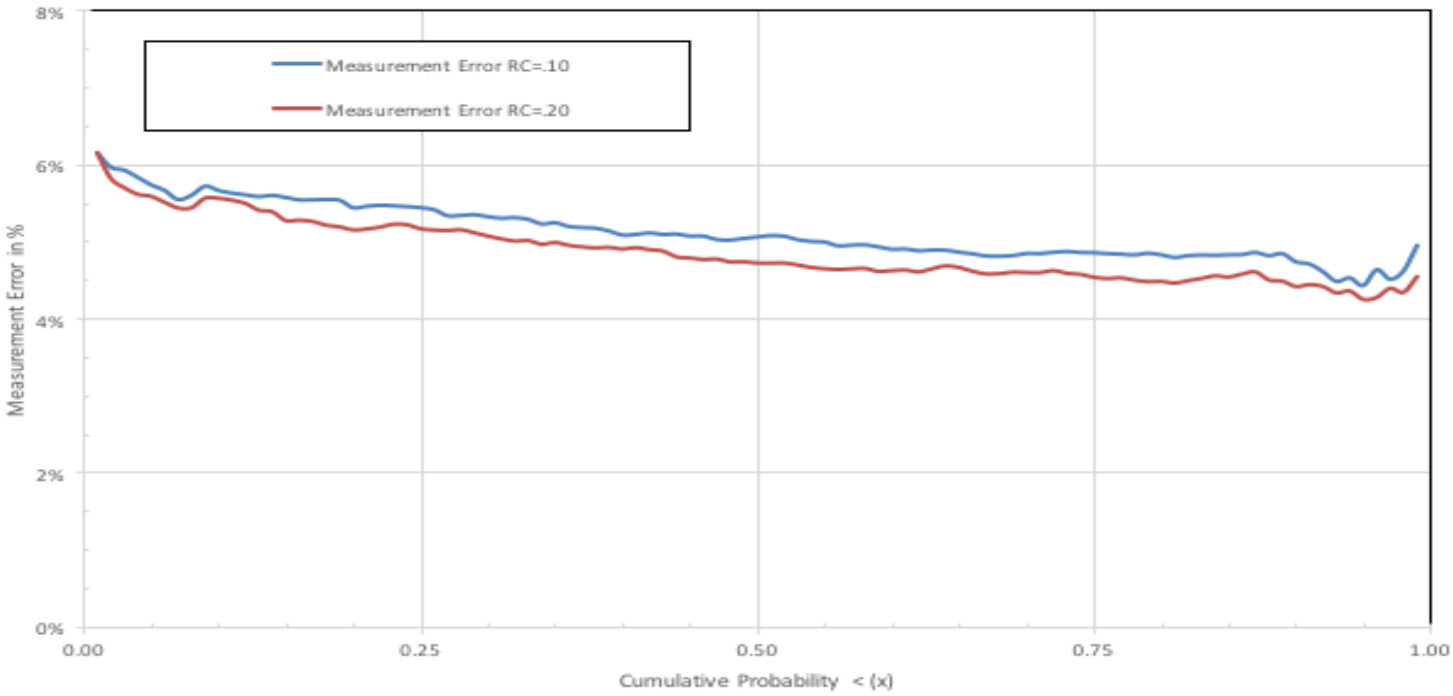
Measurement Error for Range of ATS* Reproducibility Criteria, l/sec



ExxonMobil

*American Thoracic Society (ATS) 1979, 1987, 1994

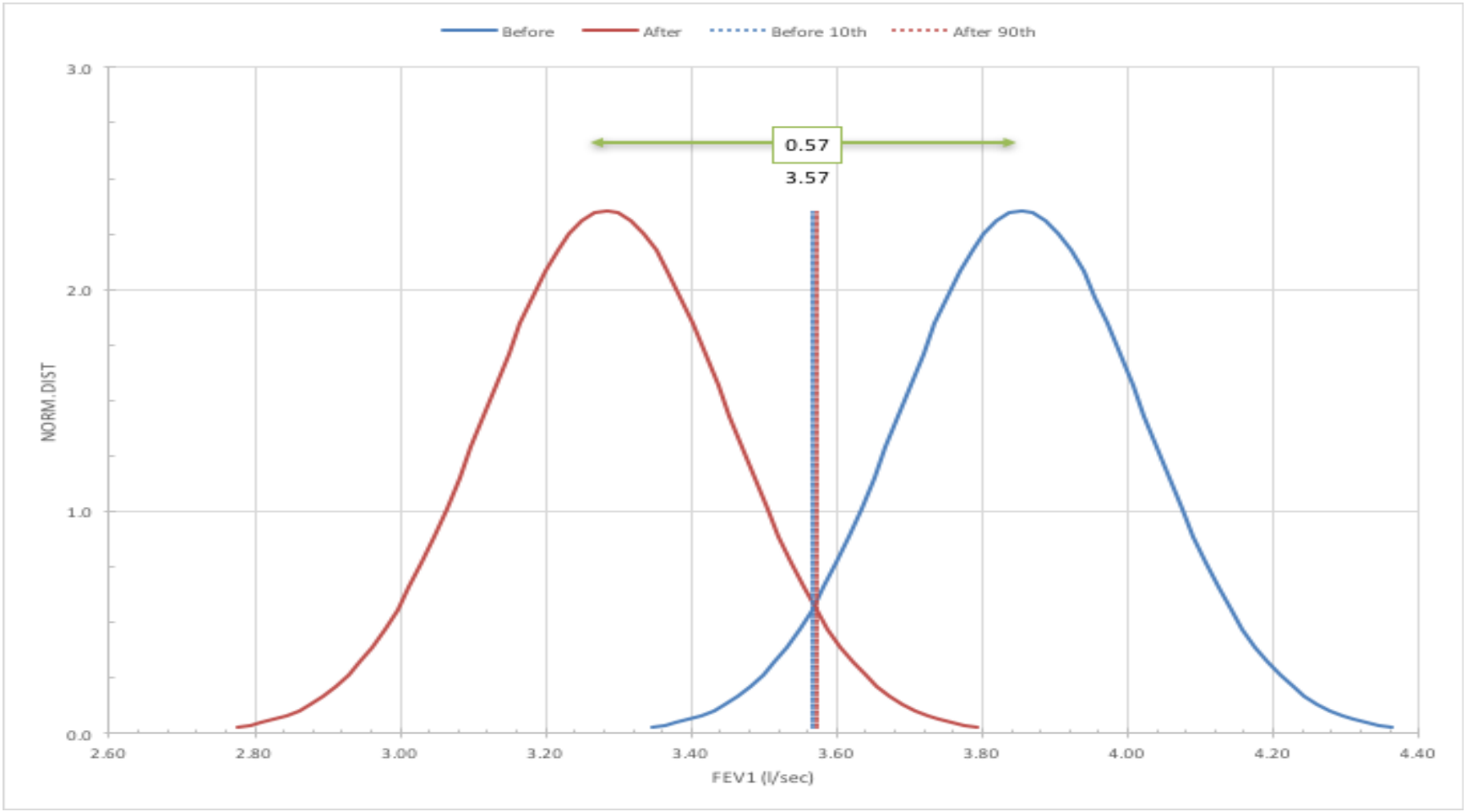
Measurement Error for Range of ATS * Reproducibility Criteria, %



ExxonMobil

*American Thoracic Society (ATS) 1979, 1987, 1994

Reduction in FEV1 must > 16% to be statistically significant



ExxonMobil

Are differences reported in a recent chamber study actually statistically significant?

Schelegle ES, Morales CA, Walby WF, Marion S, Allen RP. 6.6-hour inhalation of ozone concentrations from 60 to 87 parts per billion in healthy humans. *American Journal of Respiratory and Critical Care Medicine.* 2009;180(3):265-272.

Females and males are treated as if they are the same but are demonstrably different

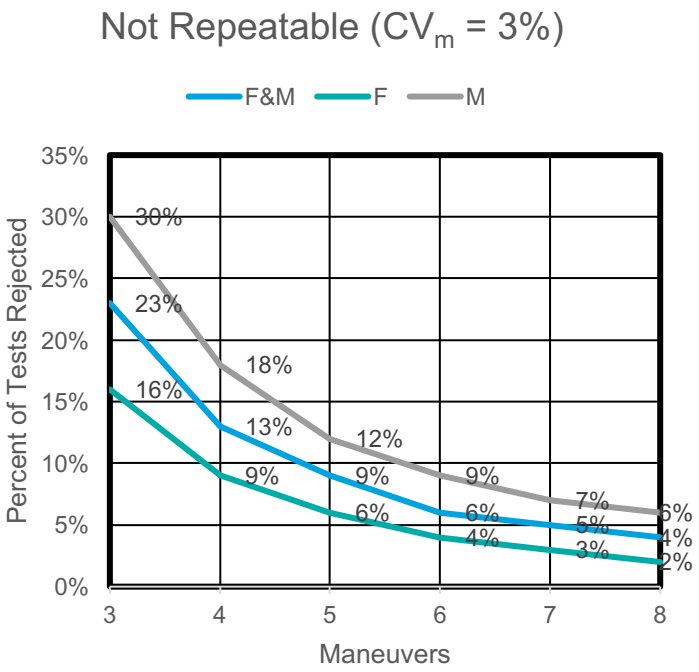
Paired differences are highly statistically significant, even in baseline

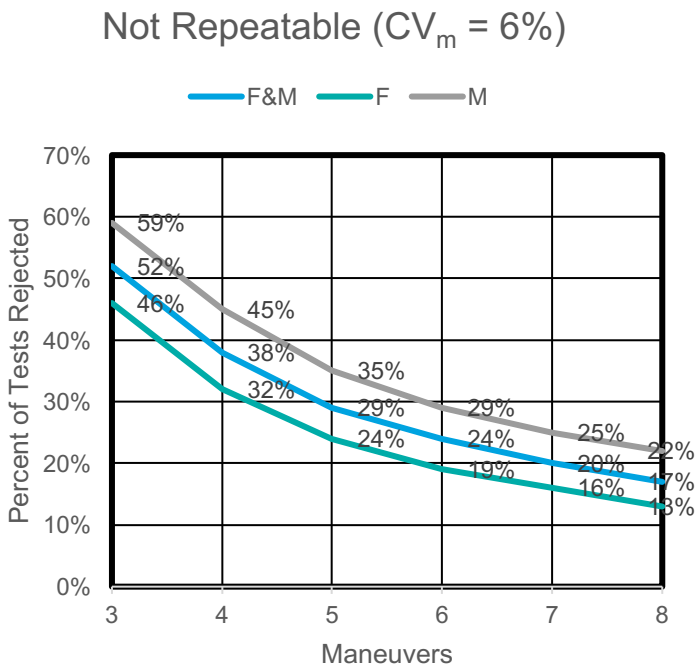
FEV1 (L/sec)						
Statistic	All		Female Only		Male Only	
Scenario	*5 Scens*	*FA Scen*	*5 Scens*	*FA Scen*	*5 Scens*	*FA Scen*
Mean	3.90	3.94	3.32	3.20	4.51	4.59
Std Dev	.635	.686	.472	.539	.594	.594
CV	.163	.174	.142	.168	.132	.129
T-test (p)	**< .001**		**.023**		**.024**	

FA Scen = Filtered Air scenario
5 Scens = Filtered Air, 60 ppb, 70 ppb, 80 ppb and 87 ppb scenarios
t-test for equality of means in paired samples: significant differences in **bold** ($p<.05$)

Published results are inconsistent with having followed the ATS protocol, and actual protocol followed cannot be discerned

Repeatable FEV_1 cannot be obtained using 2-4 maneuvers





CV_m = coefficient of variation across maneuvers within a single test.

Observed differences are uncertain; intra-test variability is ignored

Use simulation to understand the effects of inter- and intra-test variability

- Variability can be represented by coefficients of variation across tests *t* (CV^t) and maneuvers *m* (CV_m)
- Coefficient of variation for *inter-test* variability
  - Substantial literature with results ranging from 3-13%
- Coefficient of variation for *intra-test* variability
  - Virtually no literature estimating CV_m
  - Repeatability criterion combined with early test termination deters collection of enough maneuver data to estimate it
- $CV^t = 0$ and $CV_m = 0$ are implicitly assumed in air pollution studies
- Effects of CV^t and CV_m can be estimated through simulation

Simulation

Monte Carlo simulation model and parameters

Model

- Single-test individual FEV_1 is assumed to be normally distributed
- 10,000 tests simulated with 8 maneuvers per test, per ATS protocol

Parameters

- Baseline CV^t is averaged from 5 scenarios (F: 3.3%, M: 2.7%)
- 70 ppb scenario CV^t averaged across 8 tests, 1 day (F: 3.6%, M: 2.8%)
- Default $CV_m = CV^t$, then CV_m is doubled in sensitivity analysis
  - Chamber study subjects were all physically fit and young
  - Any representative sample of the population, or sample of a subpopulation of interest (e.g., COPD, asthma), would have higher sample CV_m

Simulation methods

- Calculate difference between 70 ppb and Filtered Air scenarios, separately by sex and exposure duration, for hypothetical average subjects

Questions to answer:

- What proportion of tests yield no repeatable maximum FEV_1 after *M = 3?
- What is average measurement error resulting from the repeatability criterion with early termination after *M = 3?
- How large are average test differences compared to average measurement error?

*A minimum of 3 acceptable maneuvers is required per the American Thoracic Society testing protocol

Lack of repeatability is substantial; what to do about it is unclear

	Percent of Maneuvers that are Not Repeatable After M = 3 (Average [SD] over 9 tests)		
Sex	Filtered Air	70 ppb	Combined
Female	21 [0.8]	20 [0.9]	64 [0.9]
Male	25 [2.1]	24 [1.3]	56 [0.7]

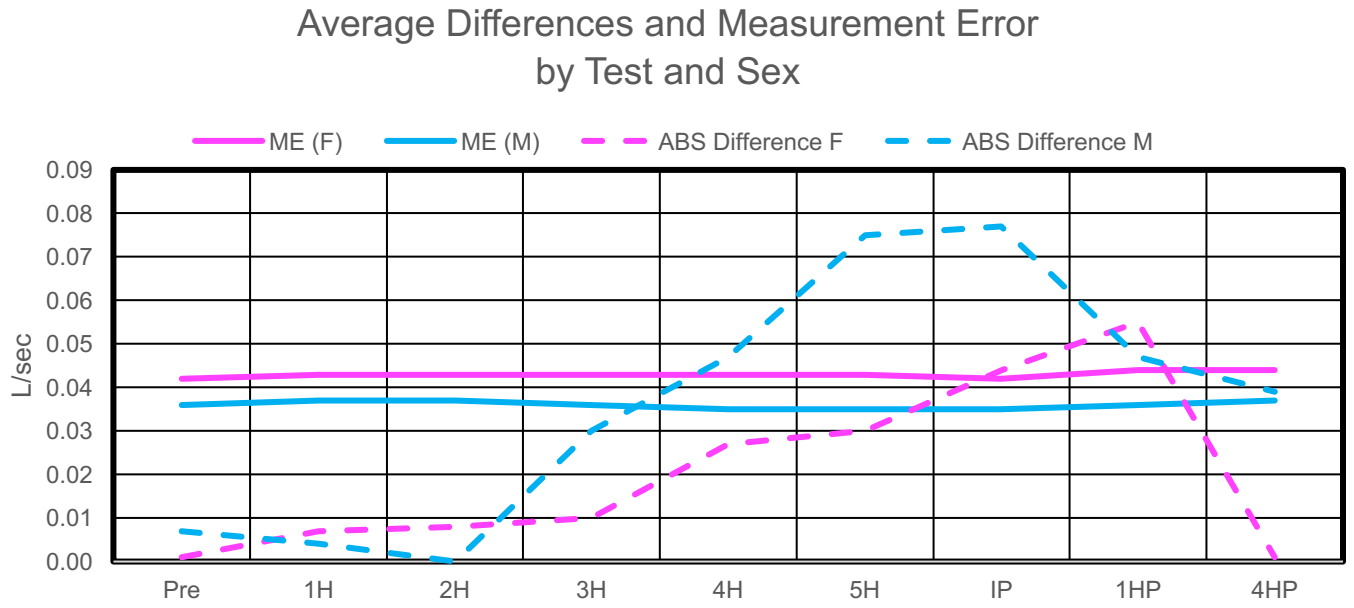
CV^t: F: 3.3%, M: 2.7% (from chamber study data)
CV_m: F: 3.3%, M: 2.7% (default assumption)

Estimated average measurement error is large relative to reported mean differences

Sex	Reported Difference at 70 ppb (% of baseline [SD], 8 tests)	Average Measurement Error (% of baseline [SD], 8 tests)
Female	-2.0 [2.0]	4.3 [0.1]
Male	-3.4 [3.2]	3.6 [0.1]

CV^t: F: 3.3%, M: 2.7% (from chamber study data)
CV_m: F: 3.3%, M: 2.7% (default assumption)

Average differences rarely exceed measurement error



CV^t: F: 3.3%, M: 2.7% (from chamber study data)
CV_m: F: 3.3%, M: 2.7% (default assumption)

20

Sensitivity analysis

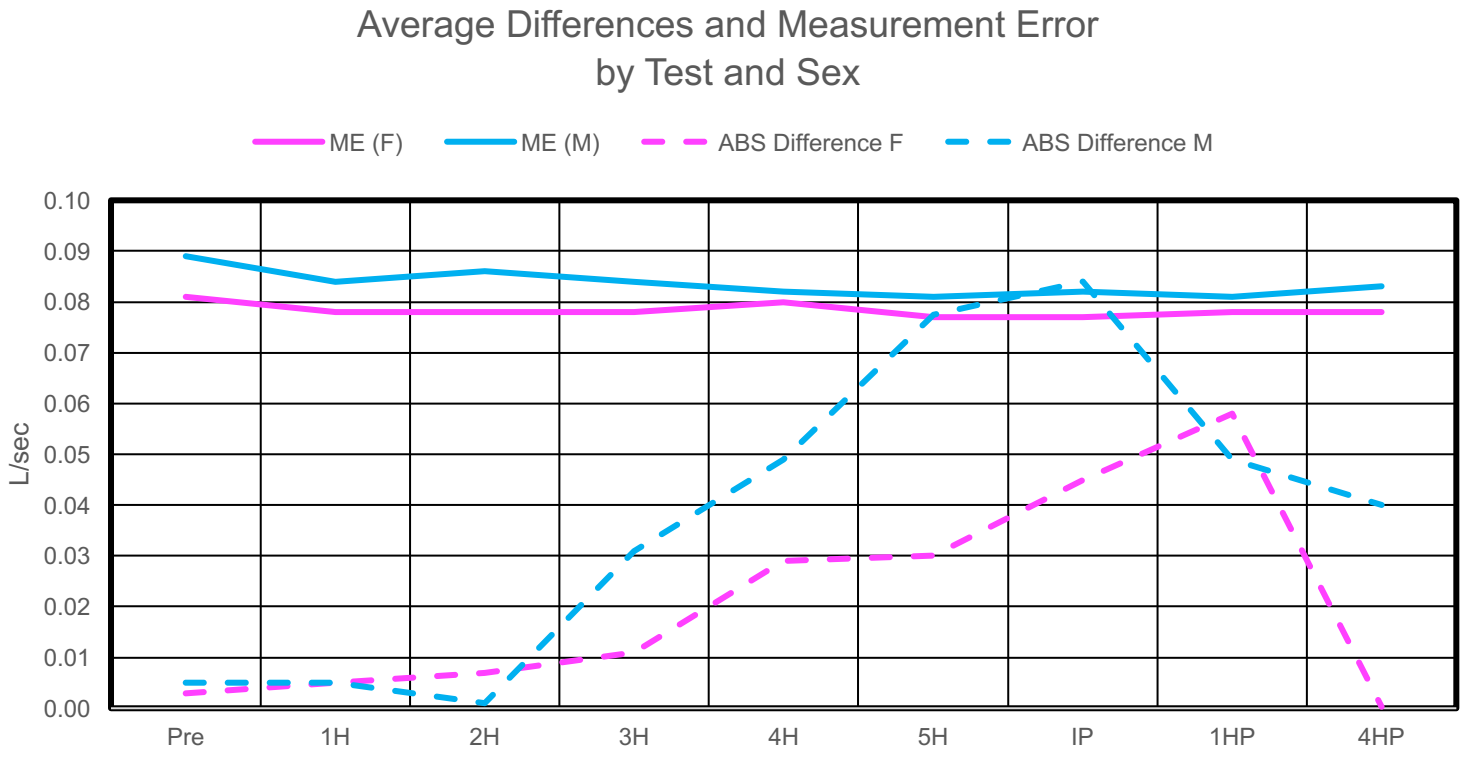
Estimated average measurement error is large relative to reported mean differences

Sex	Reported Difference at 70 ppb (% of baseline [SD], 9 tests)	Average Measurement Error (% of baseline [SD], 9 tests)
Female	−1.4 [2.6]	7.8 [0.1]
Male	−3.8 [3.1]	8.4 [0.3]

CV^t: F: 3.3%, M: 2.7% (from chamber study data)
CV_m: F: 6.6%, M: 5.4% (default assumption X 2)

ExxonMobil

Sensitivity analysis: $CV_m = 2 \times CV^t$



CV^t: F: 3.3%, M: 2.7% (from chamber study data)
CV_m: F: 6.6%, M: 5.4% (default assumption X 2)

Summary and next steps

- Results show effects of measurement error on hypothetical average subject, using both the average CV^t and 2 x the average CV^t
  - For hypothetical average subjects, measurement error is $\geq$ average FEV_1 differences after exposure under the 70-ppb scenario
  - Differences are not statistically significant once measurement error resulting from inter- and intra-test variability are accounted for
- However, analyses only considered hypothetical average subject
- Refined analysis via simulations for all 31 subjects using each subject's own CV^t is next step
  - Determine which subjects (if any) show statistically significant effects
  - Test for statistical significance for the sample (recognizing that the sample is not representative in any case)

References

- American Thoracic Society. (1979). ATS Statement-Snowbird Workshop on Standardization of Spirometry. *Am Rev Respir Dis* 119:831-838.
- American Thoracic Society. (1987). Standardization of spirometry: 1987 update. *Am Rev Respir Dis* 136:1285-1298.
- American Thoracic Society. (1995). Standardization of spirometry, 1994 update. *Am J Respir Crit Care Med* 152:1107–1136.
- Belzer RB, Lewis RJ. (in preparation). The practical significance of measurement error in pulmonary function testing conducted in research settings.
- Schelegle ES, Morales CA, Walby WF, Marion S, Allen RP. (2009). 6.6-hour inhalation of ozone concentrations from 60 to 87 parts per billion in healthy humans. *Am J Respir Crit Care Med* 180:265-272.
- U.S. Environmental Protection Agency. National Ambient Air Quality Standards for Ozone; Final Rule. (2015). *Federal Register* 80(206):65292-65468.
- U.S. Environmental Protection Agency. Regulatory Impact Analysis of the Final Revisions to the National Ambient Air Quality Standards for Ground-Level Ozone. EPA-452/R-15-007, September 2015

Questions?