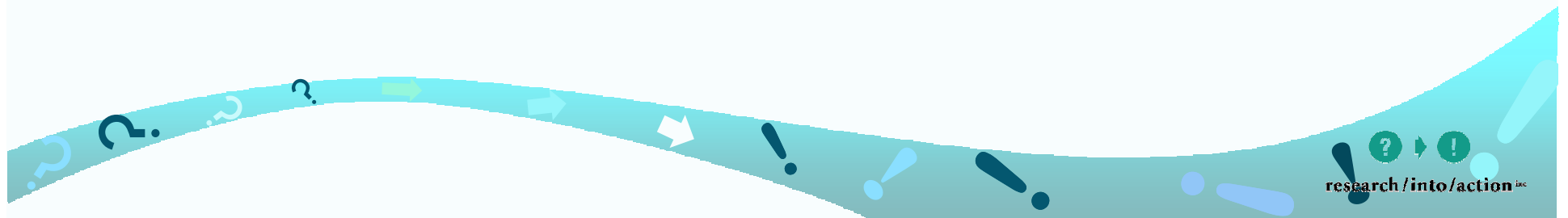# Sample Size Selection in Energy Efficiency Research and Evaluation
## The Use and Abuse of the Coefficient of Variation

Jane Peters, Ph.D., & Andie Baker, M.S.

Research Into Action

# The Funder's Stipulations for Sampling:

- 95/5 Level of Confidence and Precision must be met for the overall study

- Sampling for each domain must also attain the 95/5 level of rigor

- There are 11 domains - 4 states and 6 Public Power domains

- The funder estimates a sample of 1200 (this is what they can afford)

**The Misuse of the Coefficient of Variation**

research/into/action inc

**The Misuse of the Coefficient of Variation**

research/into/action inc
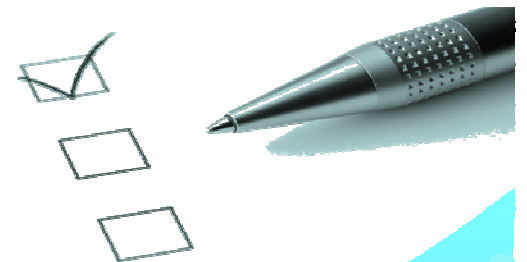
# Cochran's Formula

## For the Proportional Sampling of Large Populations

$$n = \frac{Z^2 \, \sigma^2}{e^2}$$
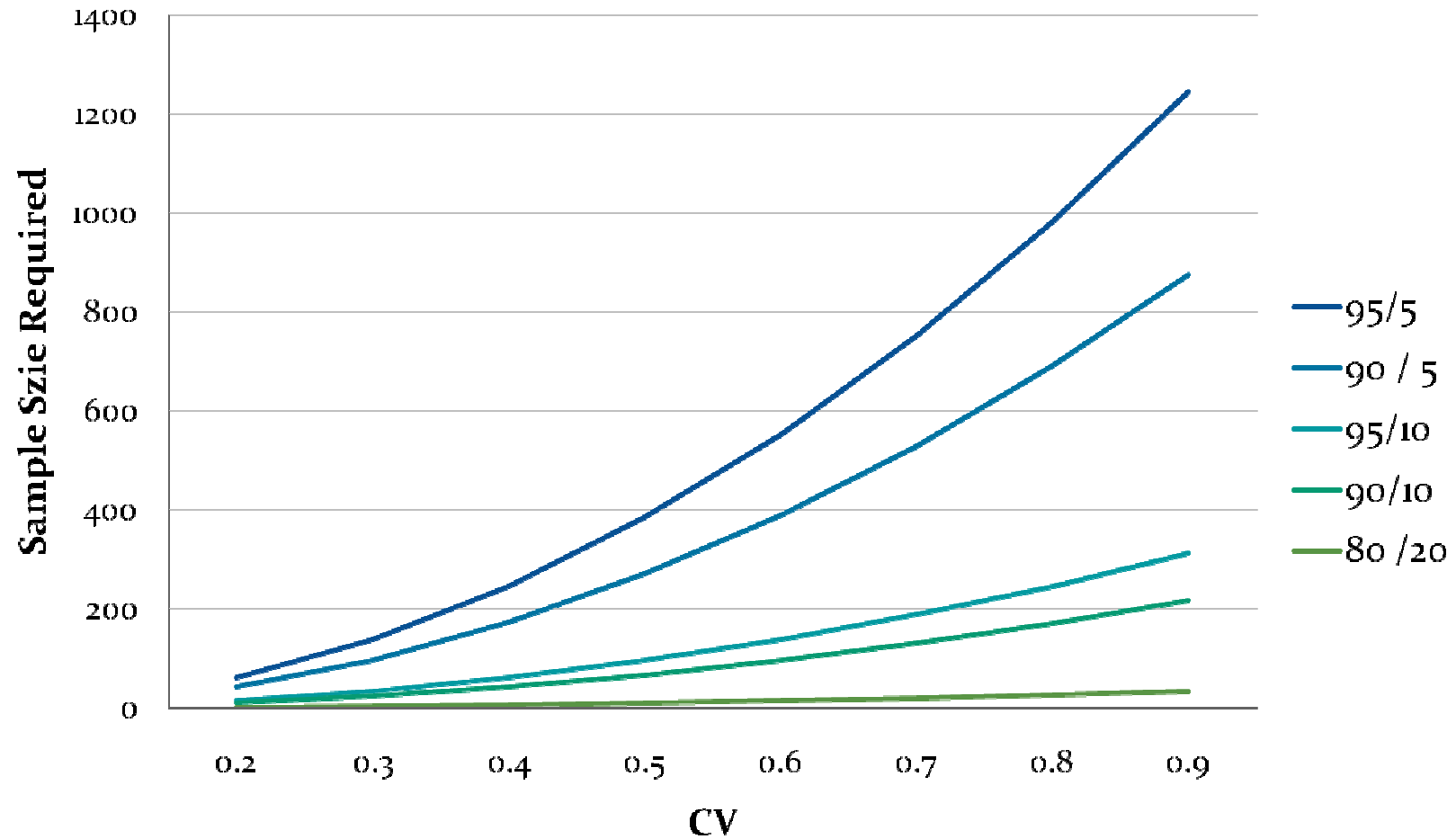
$Z^2$ is the desired confidence level

$e^2$ is the desired level of precision

$\sigma^2$ is the variance within the population, expressed as the coefficient of variation (CV)

research /into/action inc

# Coefficient of Variation

$$CV = \frac{\text{Standard Deviation}}{\text{Mean}}$$

research/into/action inc

# Relationship Between Sample Size and Variability

research/into/action inc

The Northwestern US covers ~3.79 million square miles and encompasses 5 climate zones. It is populated by multiple cultures living in residences of widely varying sizes, styles, and characteristics.

**The Misuse of the Coefficient of Variation**

research/into/action inc

The Misuse of the Coefficient of Variation

research/into/action inc

# Samples Proposed by Four Firms:

- 1400-1500 min, stratified, and reduce precision to 10% (95/10) at CV=0.5

- 1536 min, stratified, and reduce precision to 7% (95/7) at CV=0.5

- 1400-1500 min, stratified, or reduce to 95/10 (CV=0.5)

- 1192, no stratification specified, CVs ranging from 0.25 to 0.4 (95/5) – The Winner
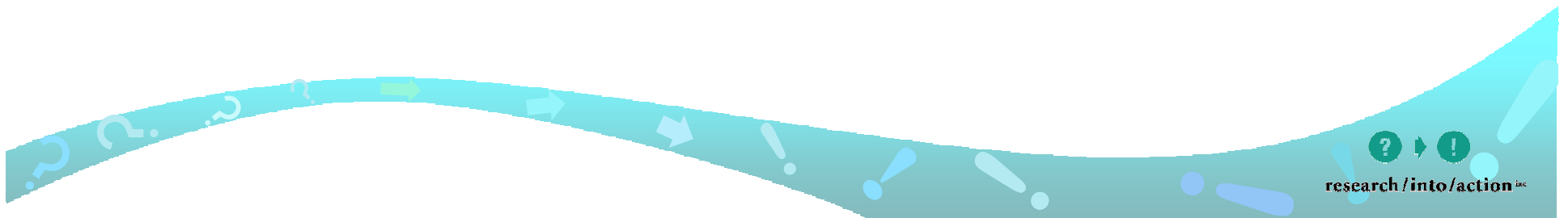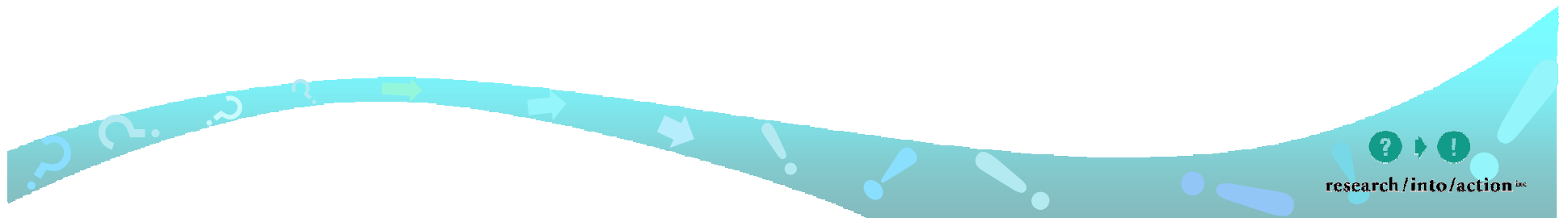
# Winning Consultant's Sample Design

| Housing Type | Total Sample |
|---|---|
| Single-Family | 745 |
| Multi-family | 347 |
| Manufactured | 100 |
| Totals | 1,192 |

research/into/action inc

# CVs used to derive sample:

| Domain Name | House Type | | |
|---|---|---|---|
| | Single-Family | Multi-family | Manufactured |
| Regional | 0.40 | 0.3 | 0.25 |
| State | 0.3 | 0.3 | 0.25 |
| Sub-Region | 0.3 | 0.2 | 0.25 |

research/into/action inc

# Results

| Domain | Sample Size | Variability in Square Footage (CV) | | Mean Square Footage |
|---|---|---|---|---|
| | | Expected | Actual | |
| **NWR Total Sample** | 1,056 | 0.45 | 0.47 | 2,054 |
| | | | | |
| **WA** | 42 | 0.3 | 0.41 | 2,071 |
| **OR** | 283 | 0.3 | 0.45 | 2,043 |
| **ID** | 182 | 0.3 | 0.48 | 2,232 |
| **MT** | 166 | 0.3 | 0.55 | 2,295 |
| | | | | |
| **Western WA** | 137 | 0.3 | 0.41 | 1,863 |
| **Puget Sound** | 178 | 0.3 | 0.43 | 1,979 |
| **Western OR** | 227 | 0.3 | 0.46 | 1,883 |
| **Eastern WA** | 110 | 0.3 | 0.38 | 2,245 |
| **Eastern OR** | 56 | 0.3 | 0.39 | 1,962 |
| **ID/MT** | ** | ** | ** | ** |

research/into/action inc
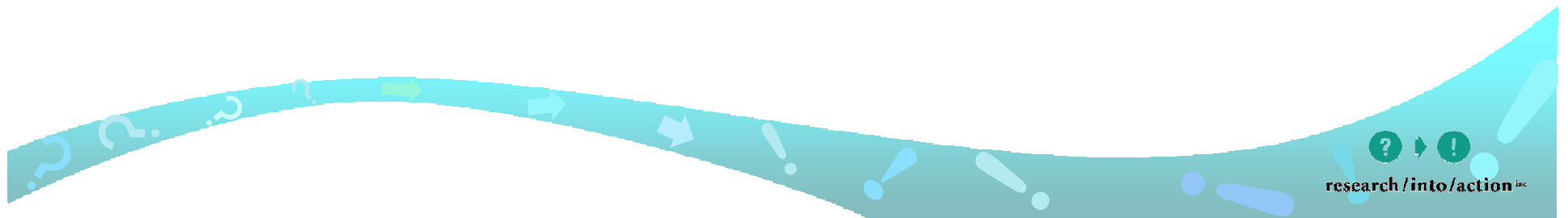
# Lessons Learned

- Include oversight committees

- Consider alternatives to 95/5

- Involve statisticians early and often

- Even seasoned energy professionals and statisticians tend to underestimate chance variability in a population based on experience and intuition.  Use Available Guides.

- Always risk asking the obvious question.
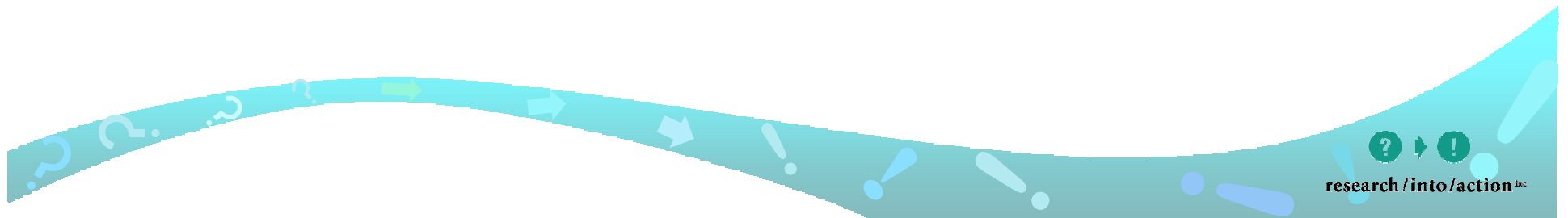
- Broader question: for studies on which targets and potential assessments will be based, what degree of rigor is required?

- How many variables should be considered when calculating these samples?

research/into/action inc

# Relationship Between Sample Size and Variability

research/into/action inc
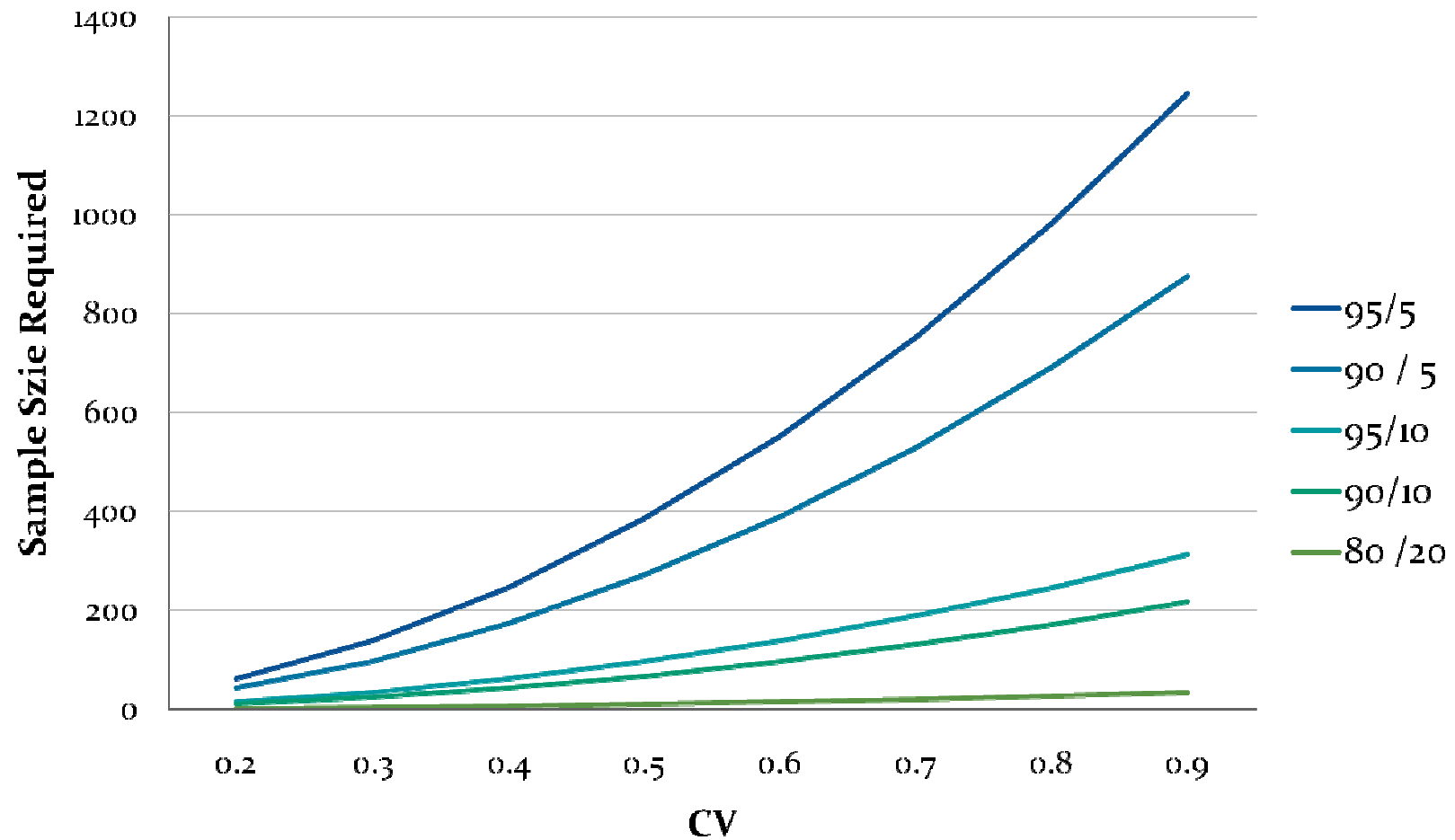
# Grazie!

**Jane S. Peters, Ph.D.**
janep@researchintoaction.com

**Andie Baker, M.S.**
andie@researchintoaction.com

research/into/action inc