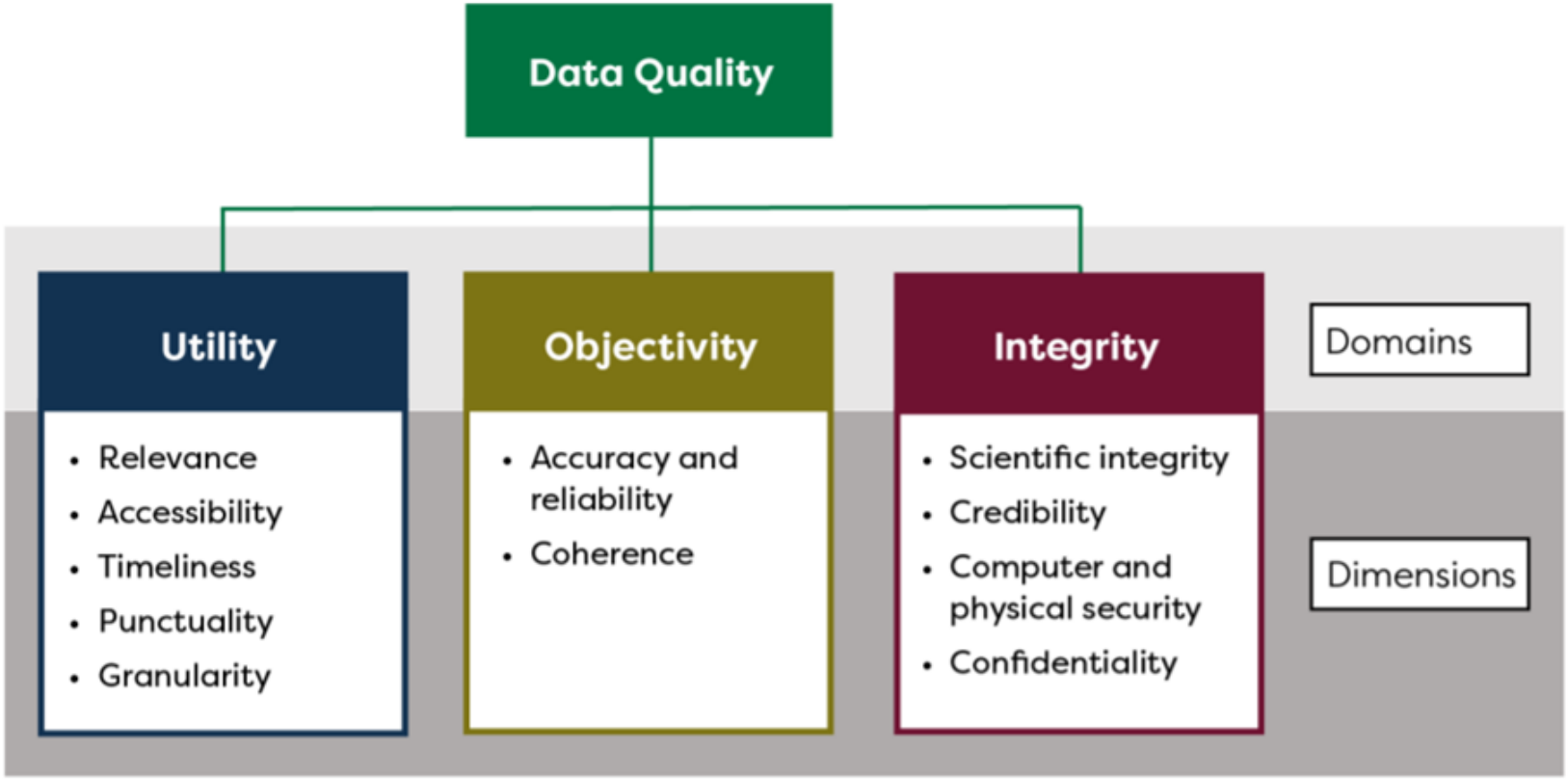
# More Fully Capturing Uncertainty Associated with Official Statistics

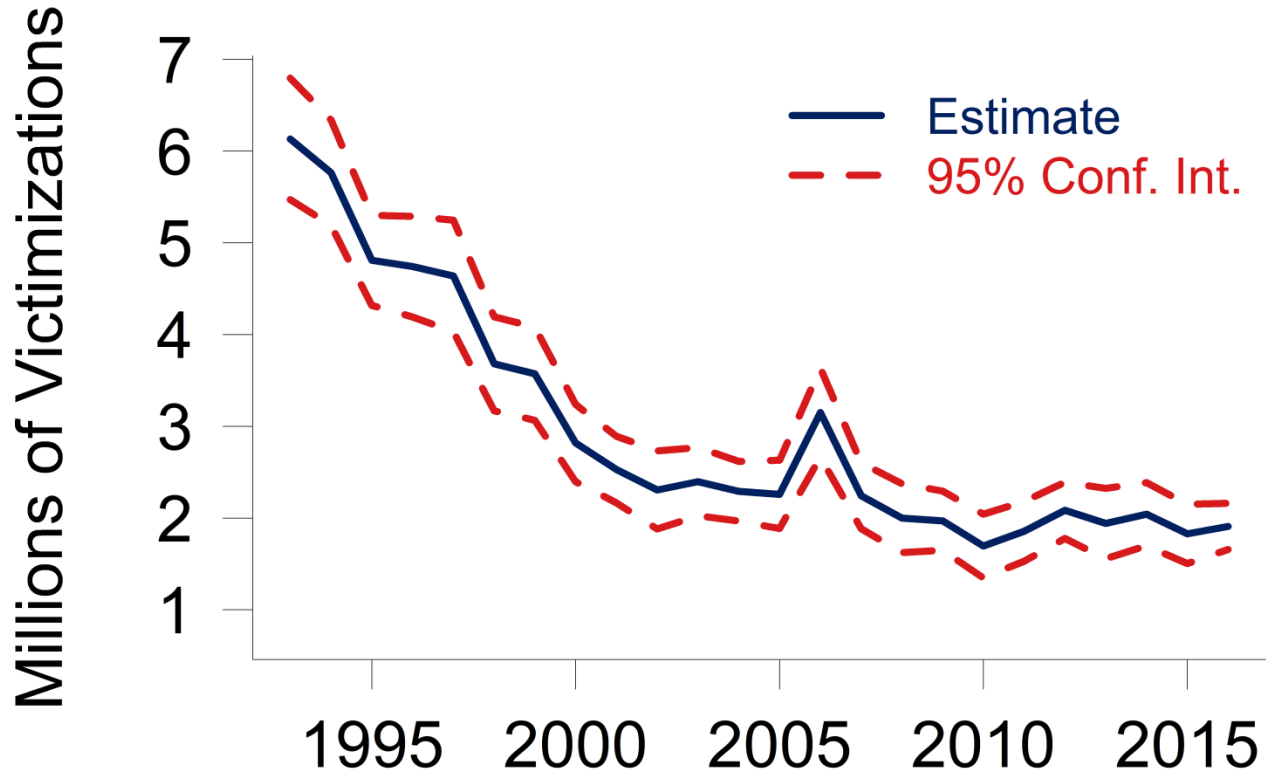
Linda J. Young
USDA National Agricultural Statistics Service
November 2, 2021

**The findings and conclusions in this presentation are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy.**

# FCSM Framework for Data Quality



# NCVS Rape, Robbery, Aggravated Assault




Courtesy of Sharon Lohr

# Is It Really a Break in Series?

- Breaks in series may be reported when simple changes are made
  - Change in question ordering
  - Change in design
  - Change in level of interviewer training
- Are these breaks or a failure to fully reflect the uncertainty in the estimates?

# Census of Agriculture

- Conducted every 5 years (ending 2 & 7)
- Count of all U.S. agricultural operations ($1,000 or more in sales or potential sales)
- Uses a list frame, the Census Mailing List
  - Census Mailing List is not complete
- An expanded June Area Survey sample drawn from the NASS area frame to assess undercoverage
  - 2009 study found substantial misclassification in the June Area Survey

# Adjusting the Census Weights

- Sampling for nonresponse follow-up (new in 2017)
- Undercoverage modeled in a capture-recapture framework
- Nonresponse modeled
- Misclassification modeled
  - Misclassification of farms as non-farms
  - Misclassification of non-farms as farms
- Calibration of adjusted weights
- Rounded the calibrated weights to integers

# 2012 Standard Errors

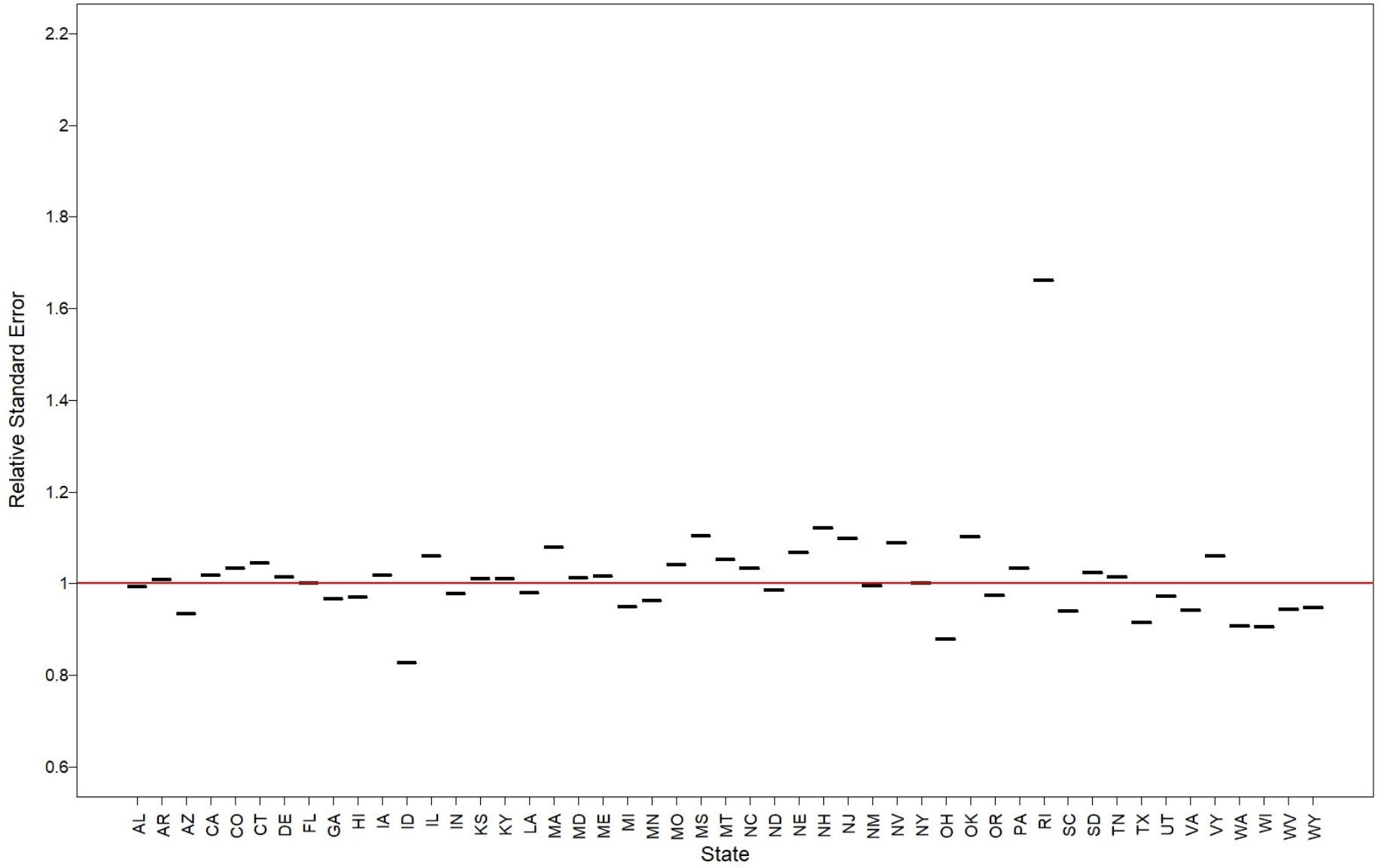
- Capture-Recapture weight for record *i*: $\widehat{CR}_i$
- Group jackknife (10 groups)
- Refit the models selected from the full dataset to obtain the weight of record *i* in jackknife group *j*: $\widehat{CR}_i^{(j)}$
  - Undercoverage
  - Nonresponse
  - Misclassification (both types)
- **Obtained the jackknife estimate of standard error of the capture-recapture weights**
- Calibrated and rounded to obtain the final adjusted weight for record *i*: $\widehat{w}_i$
- Calibrated and rounded the weights of each record *i* in jackknife group *j*: $\widehat{w}_i^{(j)}$
- **Obtained the jackknife estimate of standard error of the calibrated, rounded weights**
- **Used the larger of the two jackknife estimates as the standard error.**
- Estimate population total *T*
  - Estimate using the integer, calibrated weights $\widehat{w}_i$
  - Reported the larger of the jackknife estimate of standard errors of the fitted weights and the jackknife estimate of standard errors of the calibrated weights
- Did not account for uncertainty due to model selection
- Accounted for uncertainty associated with model estimates

**2017 Goal: Properly Account for the Variation Due to Calibration and Rounding**

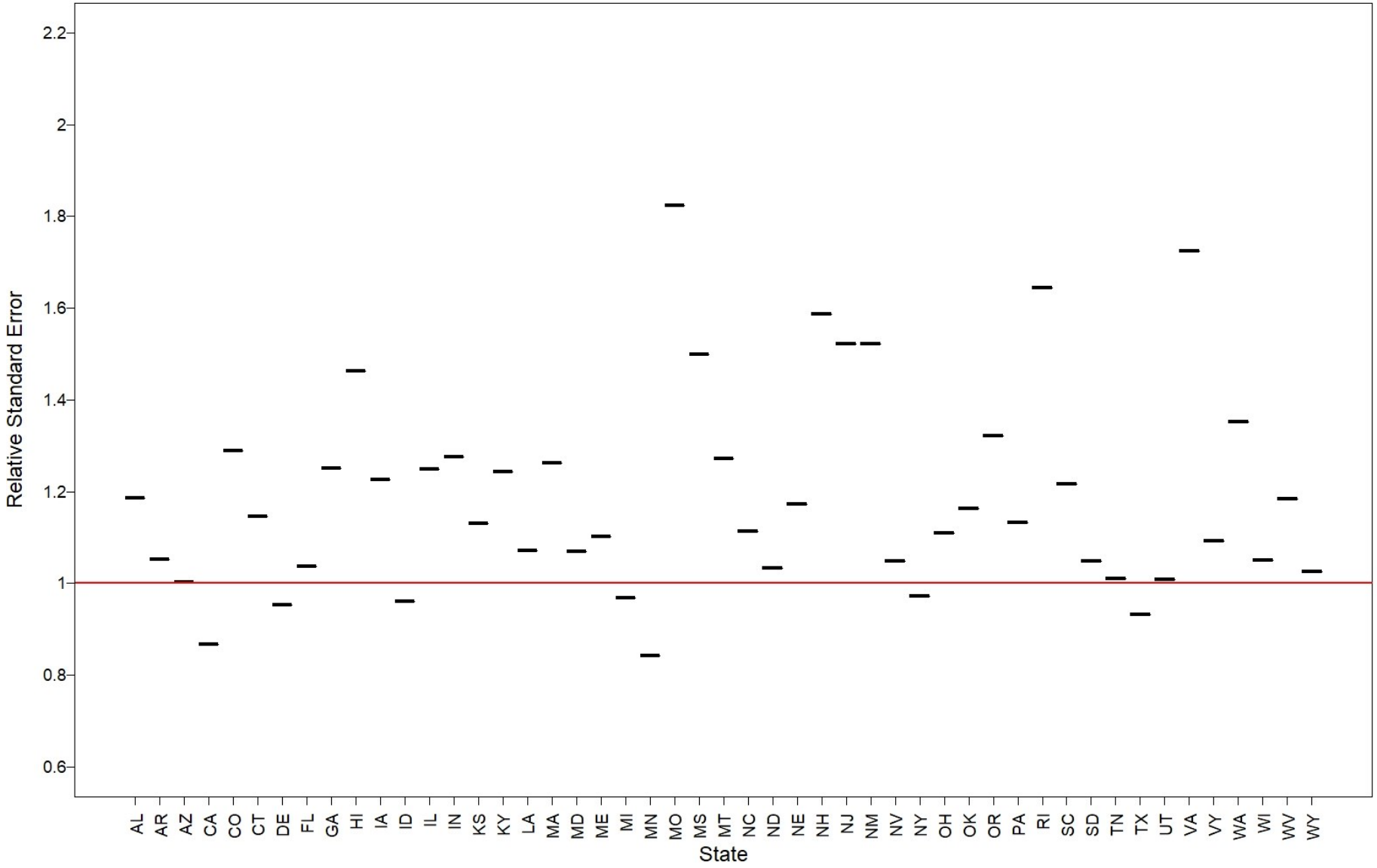
# 2017 Standard Errors

- Capture-recapture weight for record *i*: $\widehat{CR}_i$
- Integer, calibrated capture-recapture weight for record *i*: $\hat{w}_i$
- Calibration-adjusted bootstrap weight of record *i* can be approximated by: $\hat{w}_i^* = a_i\hat{w}_i$ where $a_i \sim N(1, (\hat{w}_i - 1)/\hat{w}_i)$
- Using 10 jackknife groups, refit models selected from the full dataset
  - Obtain weight for record *i* in jackknife group *j*: $\widehat{CR}_i^{(j)}$
  - Determine the integer, calibrated weight for record *i* in jackknife group *j*: $\hat{w}_i^{(j)}$
- Approximate bootstrapped-adjusted weight of record *i* in jackknife group *j*: $\hat{w}_i^{(j)*} = a_i^{(j)}\hat{w}_i^{(j)}$ where $a_i^{(j)} \sim N\left(1, (\hat{w}_i^{(j)} - 1)/\hat{w}_i^{(j)}\right)$
- Estimate the standard error of the integer, calibrated capture-recapture weights by jackknifing the approximate bootstrap-adjusted weights
- Estimate population total *T*
  - Estimate using the integer, calibrated weights $\hat{w}_i$
  - Estimate of standard error using bootstrap-adjusted jackknife weights

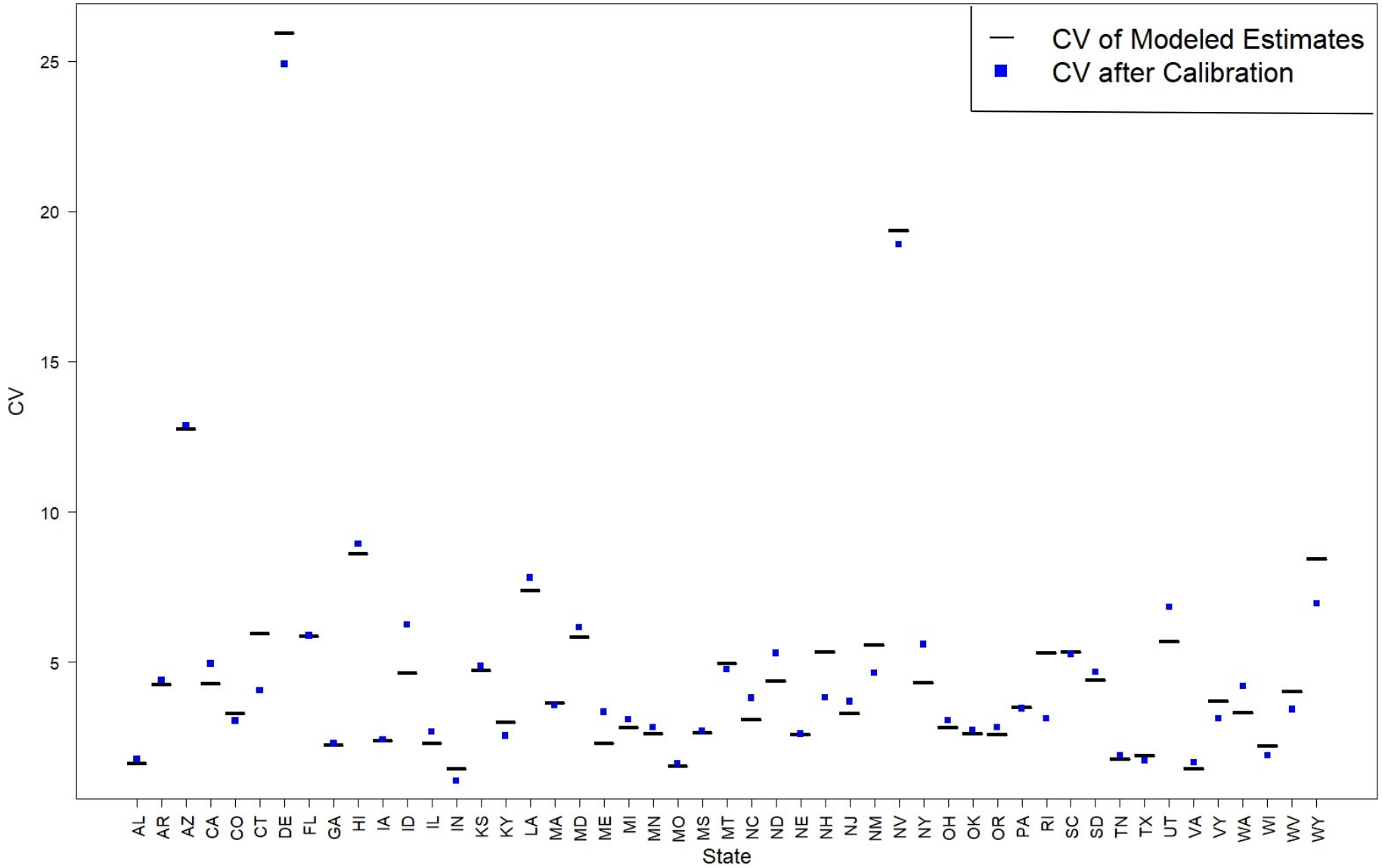
# Relative Standard Errors: 2012 Number of Farms



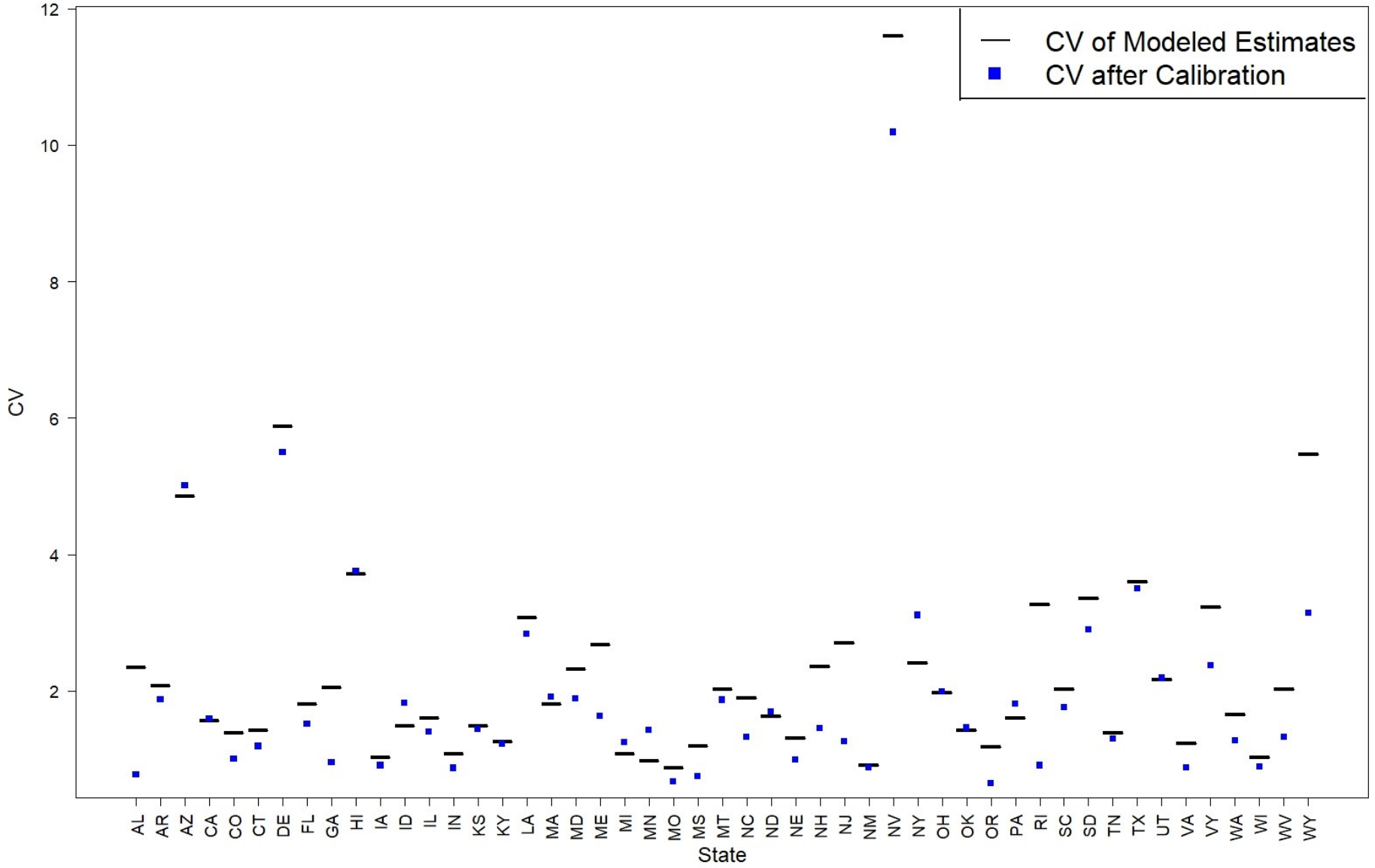
# Relative Standard Errors: 2012 Total Value of Production



# Impact of Rounding and Calibration on CVs: 2012 Number of Farms



# Impact of Rounding Calibration on CVs: 2012 Total Value of Production



# Final Thoughts

- Accounted for variation due to sampling and modeled adjustments to the weights
  - Sampling error (2017 only)
  - Undercoverage
  - Nonresponse
  - Misclassification
- Accounted for variation due to rounding and calibration
  - Approximate, conservative approach in 2012
  - Fully in 2017

# Final Thoughts

- Did not account for other sources of variation
  - Record linkage of Census Mailing List to NASS area frame
  - Editing
  - Imputation
  - Model selection
- Opportunity to develop more accurate standard errors

# Thank You!

Linda.J.Young@usda.gov