Annual Update on Activities of the FSSB Statistics Task Group

William Guthrie

June 5, 2019

Background: Statistics Task Group (STG)

- Made up of statisticians embedded in the SACs or Subcommittees, at-large affiliates, and observers
  - Embedded statisticians and at-large affiliates are voting members
  - Embedded statisticians cover the FSSB, 4/5 SACs, and 11/25 Subcommittees
- Resource for OSAC statisticians to leverage expertise and foster cross fertilization of ideas
- Now beginning to operate like a Resource Committee for OSAC Units with no embedded statistician as well
- (17+6)=23 current members + 3 observers + 2 Kavi Liaisons

Members

- Madeleine Ausdemore, South Dakota State University
- David Banks, Duke University
- Fred Bieber, Harvard Medical School
- Georgiy Bobashev, RTI International
- Alicia Carriquiry, Iowa State University
- James Curran, University of Auckland
- Simone Gittelson, Univ. of Technology Sydney
- William Guthrie, NIST Statistical Engineering
- David Kaye, Penn State Law
- Mark Lancaster, Northern Kentucky University
- Steven Lund, NIST Statistical Engineering
- Abhyuday Mandal, University of Georgia
- Max Morris, Iowa State University
- Cedric Neumann, South Dakota State Univ.
- Brent Ostrum, Canada Border Services Agency
- Mark Ruefenacht, NIST Weights & Measures
- Chris Saunders, South Dakota State Univ.
- Michael A. Smith, FBI
- Hal Stern, University of California Irvine
- Bill Thompson, Univ. of California, Irvine
- Haonan Wang, Colorado State University
- Margaret Warner, National Center for Health Statistics
- Bruce Weir, University of Washington

Individual Self-Directed Tasks

- Review and comment on documents as part of SAC and Subcommittees, prior to sending documents to SDO's, and during open comment

- Membership in Task Groups drafting or revising documents
  - Revision of *ASTM E2764 Standard Practice for Uncertainty Assessment in the Context of Seized-Drug Analysis*
  - Drafting of *Standard Practices for Evaluating Measurement Uncertainty of Quantitative Measurements in Forensic Toxicology*
  - Drafting of conclusions language for the comparison of questioned documents

- Short-term statistical consultation to SACs and Subs in support of document development
  - Comparison of methods for dental age estimation

Example – Task Group Work

Step 1 • ***Specify the measurement process***

The statement defining the measurand can be a written statement, a visual diagram and/or a mathematical expression. To be clear about what measurement process the estimation is being calculated for, it is important to be as specific as possible when defining the measurand. To distinguish one measurement process from another in a laboratory it may be necessary to include a reference to a specific type of equipment used, a specific procedure, etc. in the statement defining the measurand.

EXAMPLE Step 1: *Specify the measurement process*

The Concentration of Ethanol in Ante-Mortem Blood using the validated laboratory procedure.

The measurement process can be shown by the following mathematical expression:

$$C_{measurand} = C_{calibrators} x \frac{I_{measurand}}{I_{calibrators}} + \mathrm{b}$$

Where,
I is the instrument (GC or GCMS) response
C is the concentration
b is the bias

Each part of the measuring process will have uncertainty components that will be considered.

Example – Task Group Work

Measurement Process Reproducibility data: is in the correct measurement unit
is expressed as one standard deviation

In this example, the test item is sampled in duplicate and the laboratory procedure for the reported ethanol concentration is to average the two results. Repeat measurements of the test items provide more information and more confidence that the reported result is the best estimate of the true value. When multiple measurements are made of the test items, the average is reported, and the measurement process reproducibility data is based on single measurements of quality control samples, the standard deviation of the measurement process is divided by the square root of the number of measurements. This statistic is the standard deviation of the mean. If a single measurement result for the test items is selected to be reported (e.g., the lowest value), then the standard deviation of the mean calculation is not applicable. If the laboratory makes an equal number of multiple measurements of the quality control sample as it does of the test items and averages the results to evaluate the acceptability of the quality control sample, then the standard deviation of the mean calculation is not applicable.

The relative standard deviation of the reproducibility data in this example = 3.4084 %

The mathematical expression for relative standard deviation of the mean:

$$RSD_{mean} = \frac{s}{\sqrt{n}}$$

$$RSD_{mean} = \frac{3.4084\%}{\sqrt{2}}$$

$$RSD_{mean} = 2.4101\%$$

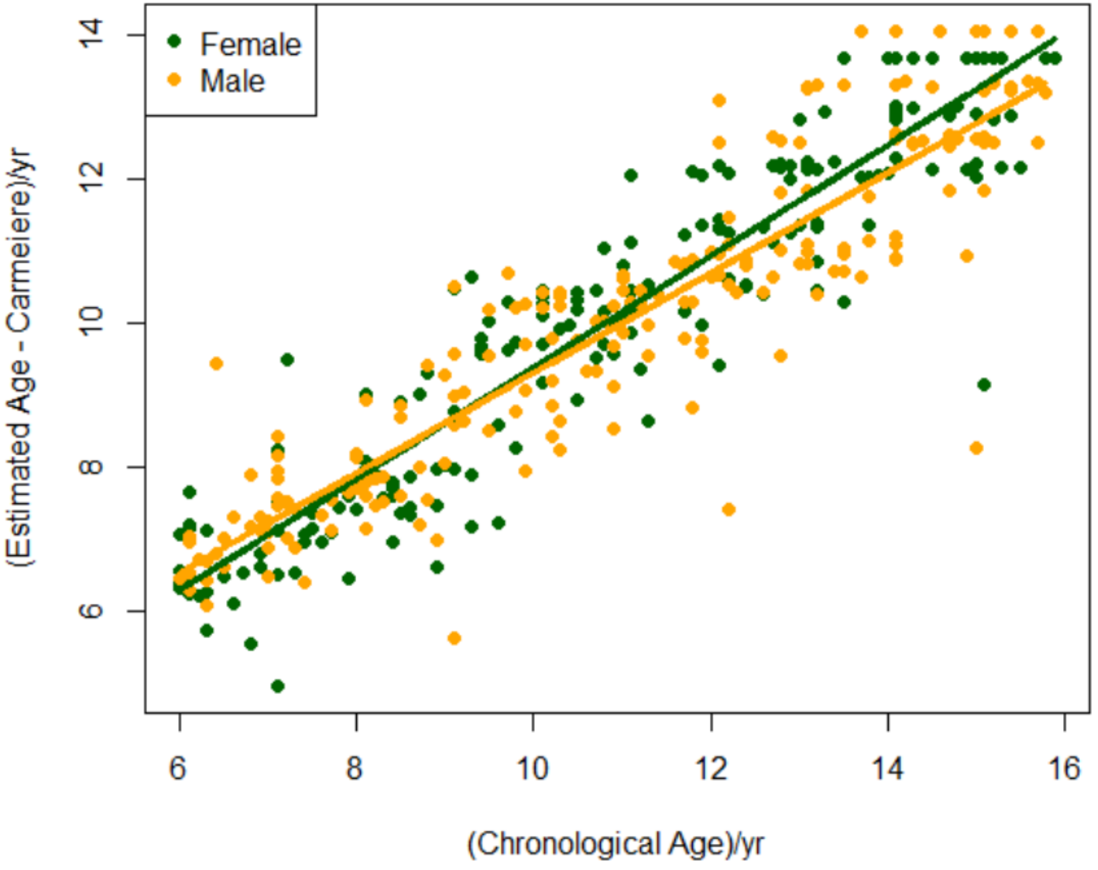
Example – Stat Consulting

calibration-analysis.R

```r
while("data"%in%search()) detach(data)
data <- read.table("data.txt",header=T)
attach(data)

library(EnvStats)
cF <- "darkgreen"
cM <- "orange"

plot(ChronologicalAge,EstimatedAgeCameriere,type="n",
     xlab="(Chronological Age)/yr",ylab="(Estimated Age - Carmeiere)/yr")
points(ChronologicalAge[Gender=="F"],EstimatedAgeCameriere[Gender=="F"],col=cF,pch
points(ChronologicalAge[Gender=="M"],EstimatedAgeCameriere[Gender=="M"],col=cM,pch
legend("topleft",c("Female","Male"),pch=19,col=c(cF,cM))

iM <- as.numeric(Gender=="M")
lf <- lm(EstimatedAgeCameriere~ChronologicalAge+iM+ChronologicalAge*iM)

plot(ChronologicalAge,EstimatedAgeCameriere,type="n",
     xlab="(Chronological Age)/yr",ylab="(Estimated Age - Carmeiere)/yr")
points(ChronologicalAge[Gender=="F"],EstimatedAgeCameriere[Gender=="F"],col=cF,pch
```

~/bluegrass/OSAC/Odontology/Age Assessment/

```
   obs.y     pred.x       lpl.x     upl.x   minus.U   plus.U
1      6   5.209570   2.455681  7.598266 -2.753889 2.388696
2      7   6.653993   4.017686  8.969268 -2.636307 2.315275
3      8   8.098417   5.575279 10.370357 -2.523137 2.271940
4      9   9.542840   7.124640 11.822230 -2.418200 2.279390
5     10  10.987264   8.657297 13.325926 -2.329967 2.338663
6     11  12.431687  10.156016 14.861149 -2.275671 2.429462
7     12  13.876111  11.601481 16.411745 -2.274629 2.535634
8     13  15.320534  12.998076 17.969977 -2.322458 2.649443
9     14  16.764957  14.366838 19.532344 -2.398120 2.767386
10    15  18.209381  15.722141 21.097149 -2.487240 2.887768
11    16  19.653804  17.070665 22.663495 -2.583140 3.009690
```



Centrally Organized Tasks

- Training Sessions
  - Two plenary training sessions offered at Nov./Dec. 2018 All-Hands Meetings
  - Stats 101 for Forensic Science – Interactive Session on Control Charts
  - Forensic Statistics and the Probative Value of Evidence

- Membership in Interdisciplinary Task Groups
  - Two STG members participating in FSSB Technical Issues Task Group with other members of Trace Subcommittee, Chemistry SAC, and affiliates
  - One goal to address concerns over definitions for differences between trace evidence items when compared using expert judgement
  - Another goal to investigate use of statistical methods for comparison rather than expert judgement – with a focus on comparison of automotive paint by FTIR

Definitions: Significant and Meaningful Differences

- **Old:** *significant difference* - a difference between two samples that indicates that they do not share a common origin.

- **Current:** *meaningful difference* - a feature or property of a sample that does not fall within the variation exhibited by the comparison sample, considering the limitations of the sample or technique, and therefore indicates the two samples do not share a common origin.
  - *Discussion* - The use of this term does not imply the formal application of statistics.

Proposed Definition: Meaningful Difference

- *meaningful difference* - a difference in a feature or property of items compared, based on expert judgment by a qualified analyst, deemed substantial enough to conclude that the two items do not share a common origin.
- Note 1: The finding of a meaningful difference must be based on comparison methods that account for (or consider) all necessary secondary characteristics (e.g., environmental exposure).
- Note 2: Determination of a meaningful difference must consider and account for limitations of the technique(s) used and the items being compared.
- Note 3: This term shall be used when the determination of the difference is based on expert judgment and not on the formal use of statistical methods. As a result, meaningful difference does not imply statistical significance.

Proposed Definition: Statistically Significant Difference

- *statistically significant difference* - a difference in a feature or property of items compared, based on a statistical analysis, that quantifies that the difference is extreme enough to be unlikely to be observed when similar same-source evidence is analyzed under similar conditions.
- Note 1: The finding of a statistically significant difference must be based on comparison methods that account for (or consider) all necessary secondary characteristics (e.g., environmental exposures).
- Note 2: Determination of a statistically significant difference must consider and account for limitations of the technique(s) used and the items being compared.
- Note 3: The determination of a statistically significant difference requires that all known or suspected sources of uncertainty are investigated and accounted for. This includes uncertainty arising from both random and systematic sources of measurement error that could affect the results.

Summary

- STG members have made many different types of statistical contributions to OSAC work this year

- Many difficult topics that still require substantial statistical input
  - Methods for comparing complex responses (e.g. FTIR of paint)
  - Optimization of analytical methods (e.g. GC-MS for fire debris analysis)
  - Methods for summarizing results and reporting conclusions
  - Many others

- Need more embedded statisticians to cover all OSAC units with relevant standards in development