The National Ballistics Imaging Comparison Projects

James Yen

Statistical Engineering Division, National Institute of Standards & Technology (NIST)

Ted Vorburger, Alan Zheng, Brian Renegar, John Song, Hans Soons, Robert Thompson, Wei Chu, NIST; Martin Ols, ATF

American Academy of Forensic Sciences Meeting

Seattle, WA

February 20, 2014

Disclaimers

- Certain commercial equipment, instruments, or materials may identified in this report in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.
- This talk mentions NIST Standard Bullets and Standard Cartridge Cases sold by NIST, a U.S. government entity
- This talk mentions the National Integrated Ballistic Information Network (NIBIN), which utilizes systems manufactured by a private company

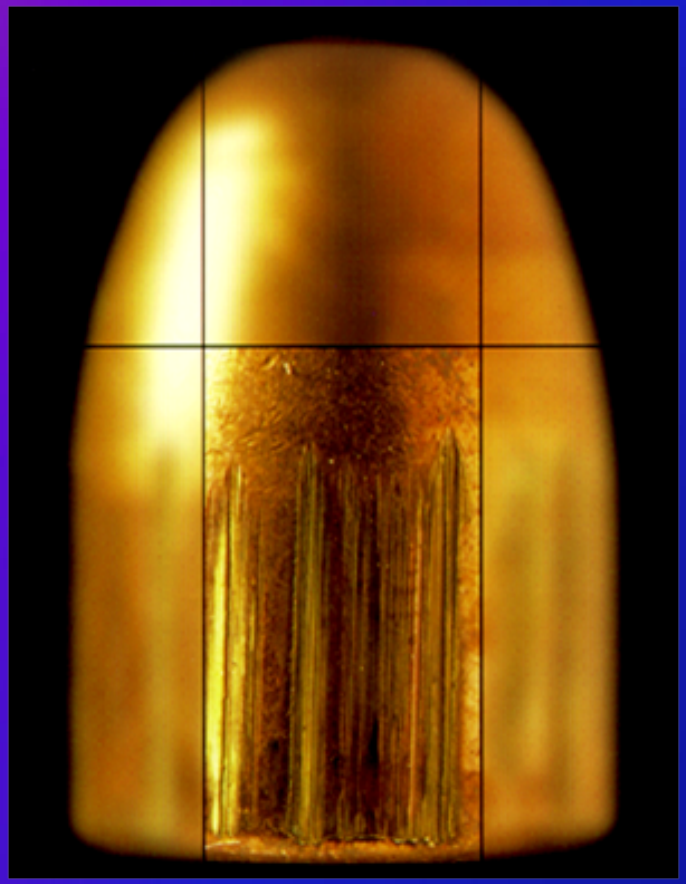
Acknowledgements

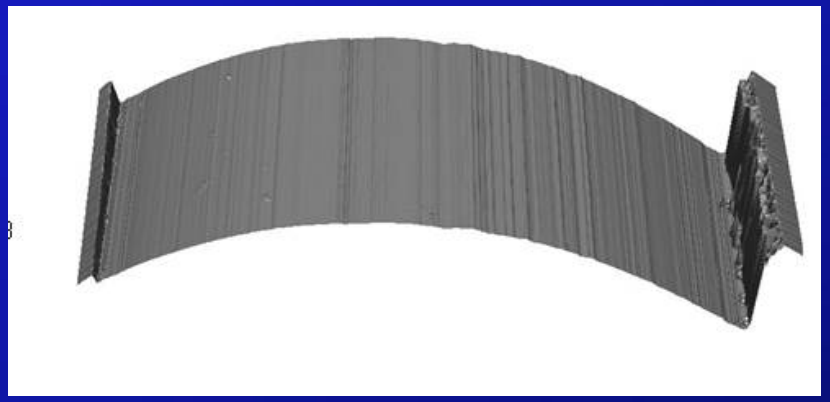
- The funding for this research was partly provided by the U.S. National Institute of Justice (NIJ) through the Law Enforcement Standards Office (OLES) at NIST.

Outline

1. Primer on Firearms markings
2. National Integrated Ballistic Information Network (NIBIN)
3. Standard Bullets and Cartridge Cases
4. National Ballistic Identification Comparison (NBIC) studies and their results
5. Conclusions

Bullet striations

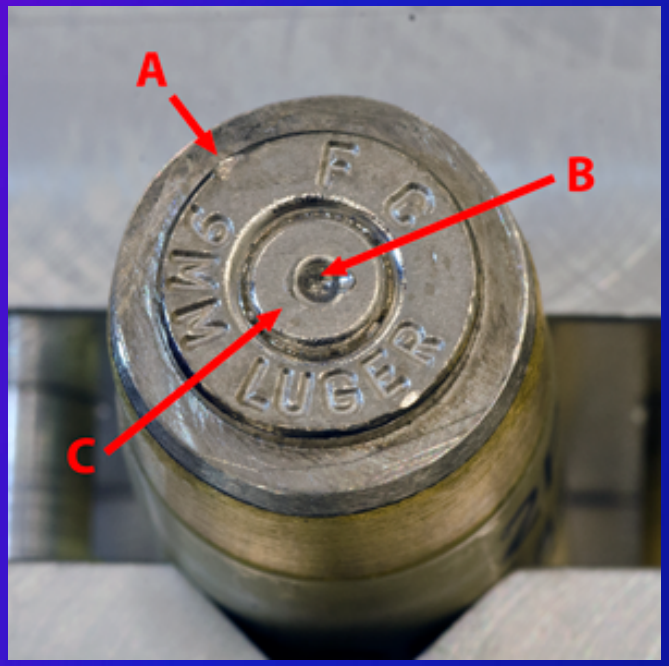


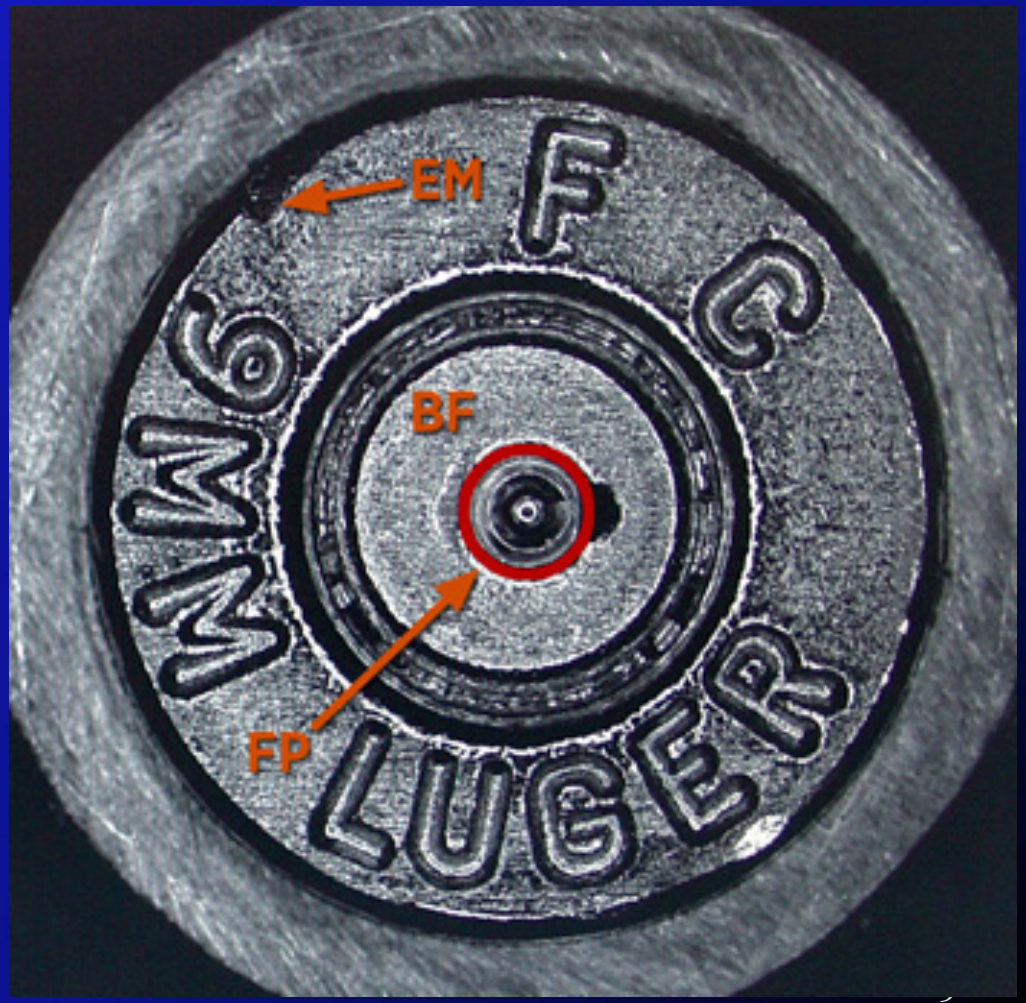


- Pictures from Alan Zheng, R.M. Thompson, Brian Renegar, NIST

Cartridge Casing marks

- FP- Firing Pin
- BF- Breech face
- EM- Ejector Mark





NIBIN

- National Integrated Ballistic Information Network (NIBIN) administered by ATF since 1999
- Over 180 NIBIN stations in U.S.
- Images of evidence casings/bullets are acquired/ measured at station
- Database of previously entered test fires can be searched to find the “most similar” casings/bullets present
  - Typical output is “Top Ten” lists with “correlation” scores

NIBIN (cont'd)

- Quality of acquired images a factor in usefulness of NIBIN database
- Individual NIBIN stations run their own check standards
- ATF approached NIST about development of a more universal reference standard to aid quality control in firearm identification in general and NIBIN in particular

SRM 2460 Standard Bullet & SRM 2461 Standard Cartridge Case

NBIC Studies :

- Master images (“Golden Images”) of Standard Casing and Bullet were entered into part of NIBIN database
- National Ballistic Identification Comparison (NBIC [-1]) involved 19 NIBIN operators from across the country
  - They took repeated measurements on standard casings and standard bullets (as if they were evidence) over time and recorded their “correlation” scores with respect to the Golden Images

NBIC-2

- NIBIN began transitioning to a newer version of the instrument for casing image acquisitions, so ….
- NBIC-2 involved 14 NIBIN operators (3 from NBIC-1) measuring just casings on the newer version (with new Golden Images acquired on the newer version)

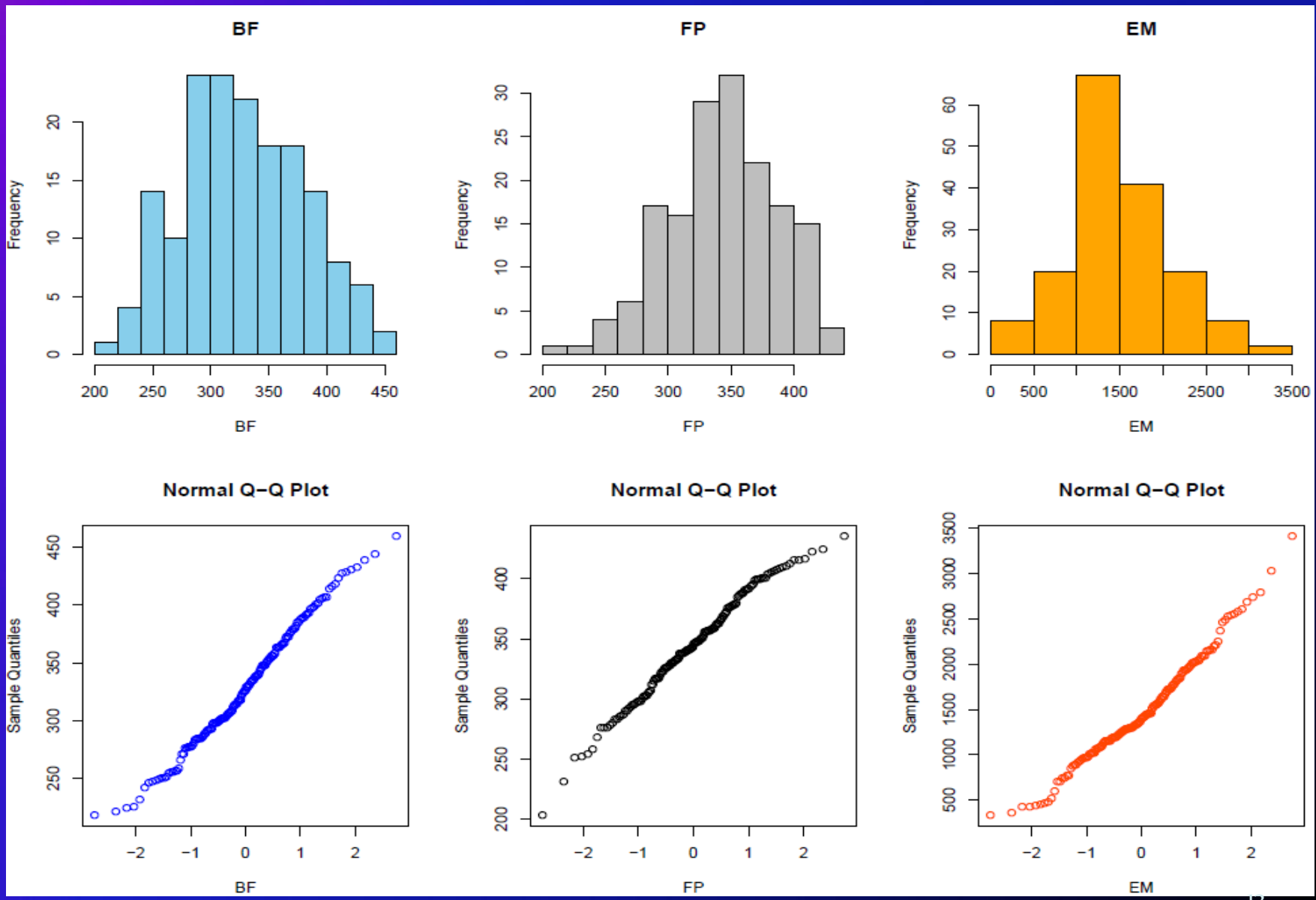
NBIC study goals:

- Not a Gallup poll of operator practice
  - NBIC participants not randomly chosen, but volunteers (the cream of the crop?)
  - NBIC tasks not part of participants' regular heavy job workloads
- Goal is to capture the "normal" variability of scores that would occur from multiple measurements uing proper procedure
- Intended to form a baseline of best practices, formalized by control limits, with control charts for monitoring individual performance
- NBIC-2 needed to capture effect of newer instrument version

Some caveats

- Unlike some other process control situations (e.g. length of metal bar), there is no single 'correct' similarity score:
  - Correlation scores used for sorting, usually not meant to be interpreted as an estimate of an intrinsic physical quantity
  - Since the acquired Standard Casing or Standard Bullet images should "match" the Golden Image, the scores cannot be "too high"
    - Some scores were really, really high (outliers?)

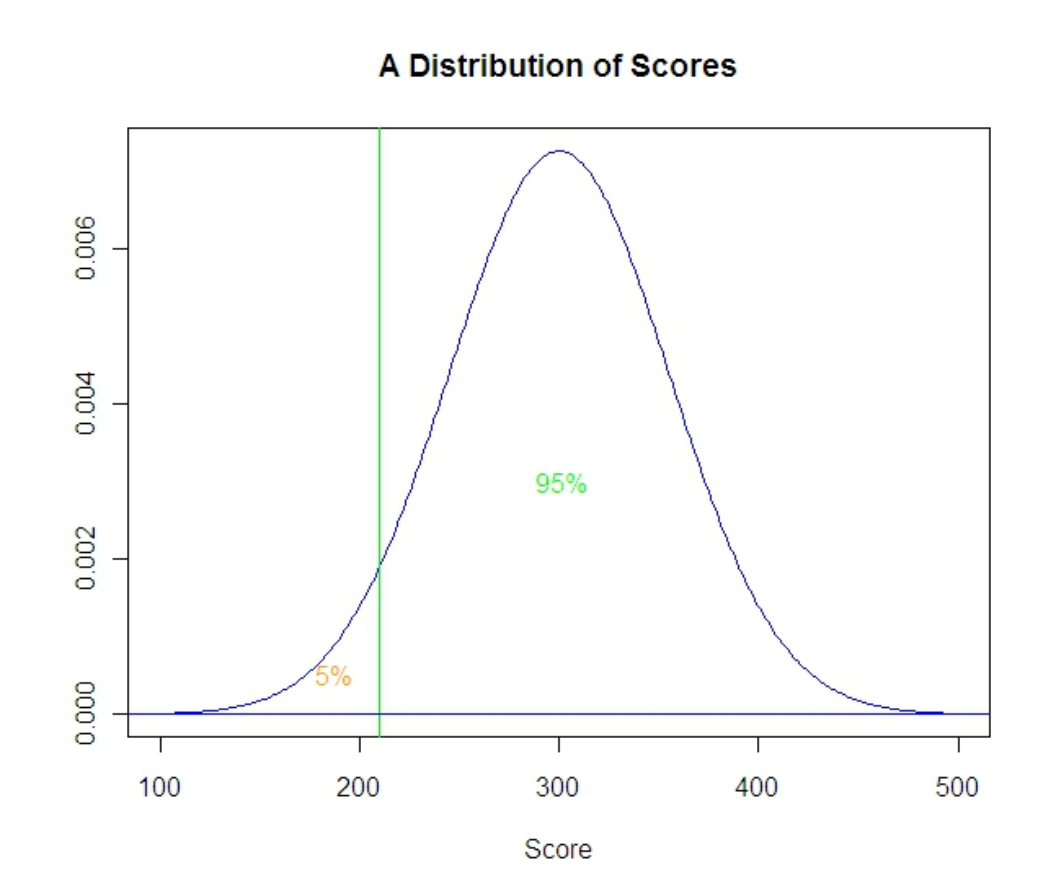
NBIC 2: Histograms and QQ norm plots: Phase 2 scores (with outliers trimmed)



A Theoretical Control Limit

- If the population of scores is normally distributed (Gaussian) and has mean C and standard deviation b, then 95% of scores should be above

  C - 1.645 b



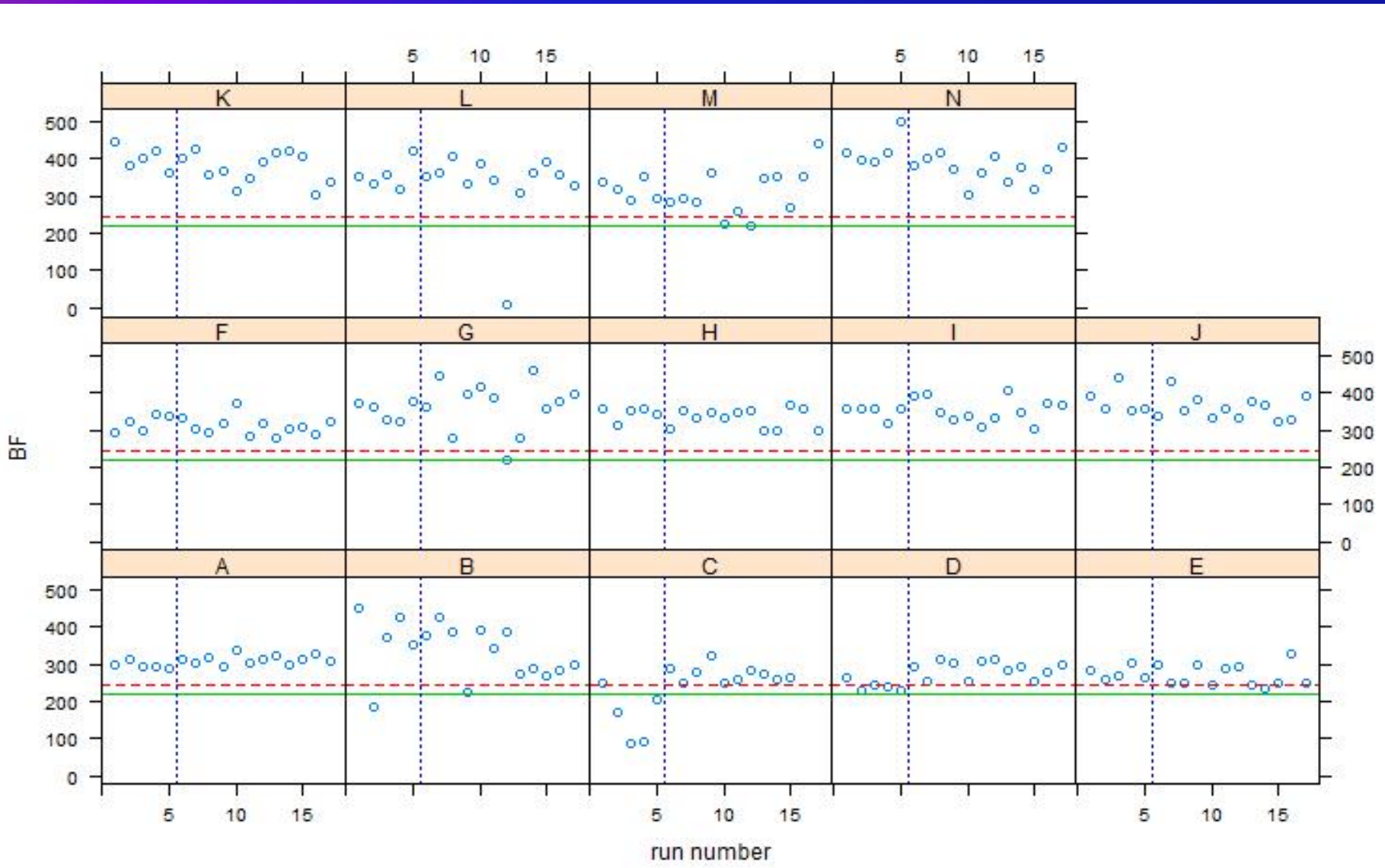
A Clarification on distributions

- The goal to estimate a global distribution that takes into account the variability of the scores from a population of well performing units
  - Various factors can have effects that each contribute to the variability of the global distribution
- It is not implied that the scores of each individual unit will follow this same distribution
  - "Essentially, all models are false but some are useful." - George Box (paraphrased)
- Operators should monitor their individual control charts, and those for their lab, in addition to the global control limits

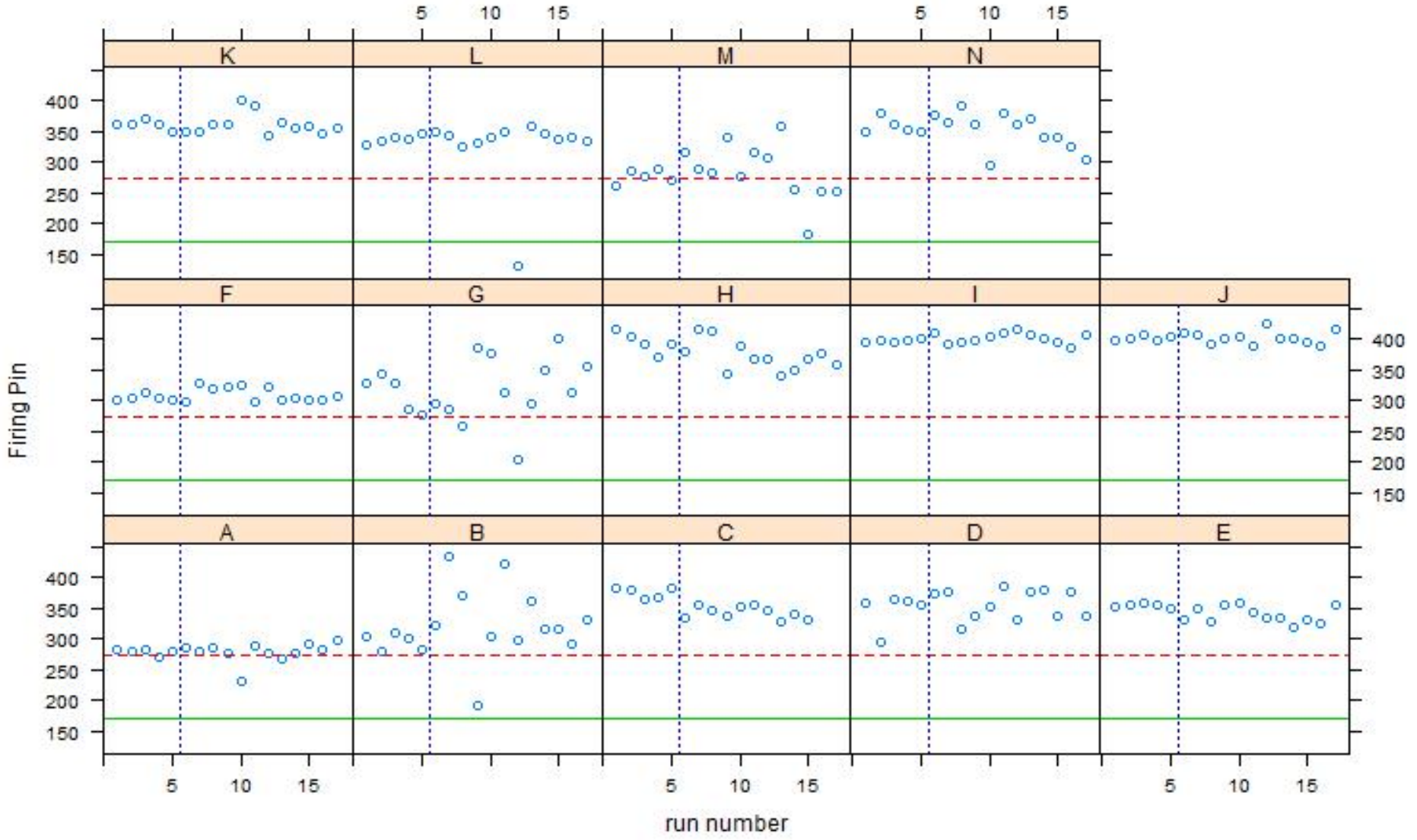
Demonstration of proposed limits

- The following graphs plot the NBIC-2 data by "Run"; the vertical dotted lines divide the runs in the 2 phases
  - Phase 1: 1 - 5
  - Phase 2: 6-17
  - Run is not linear in actual time
- Horizontal lines on the following slides demonstrate the proposed control limits on the past NBIC data:
  - the red dashed upper line is the new proposed NBIC-2 95% control limit
  - The lower dotted green line is the old NBIC-1 95% limit

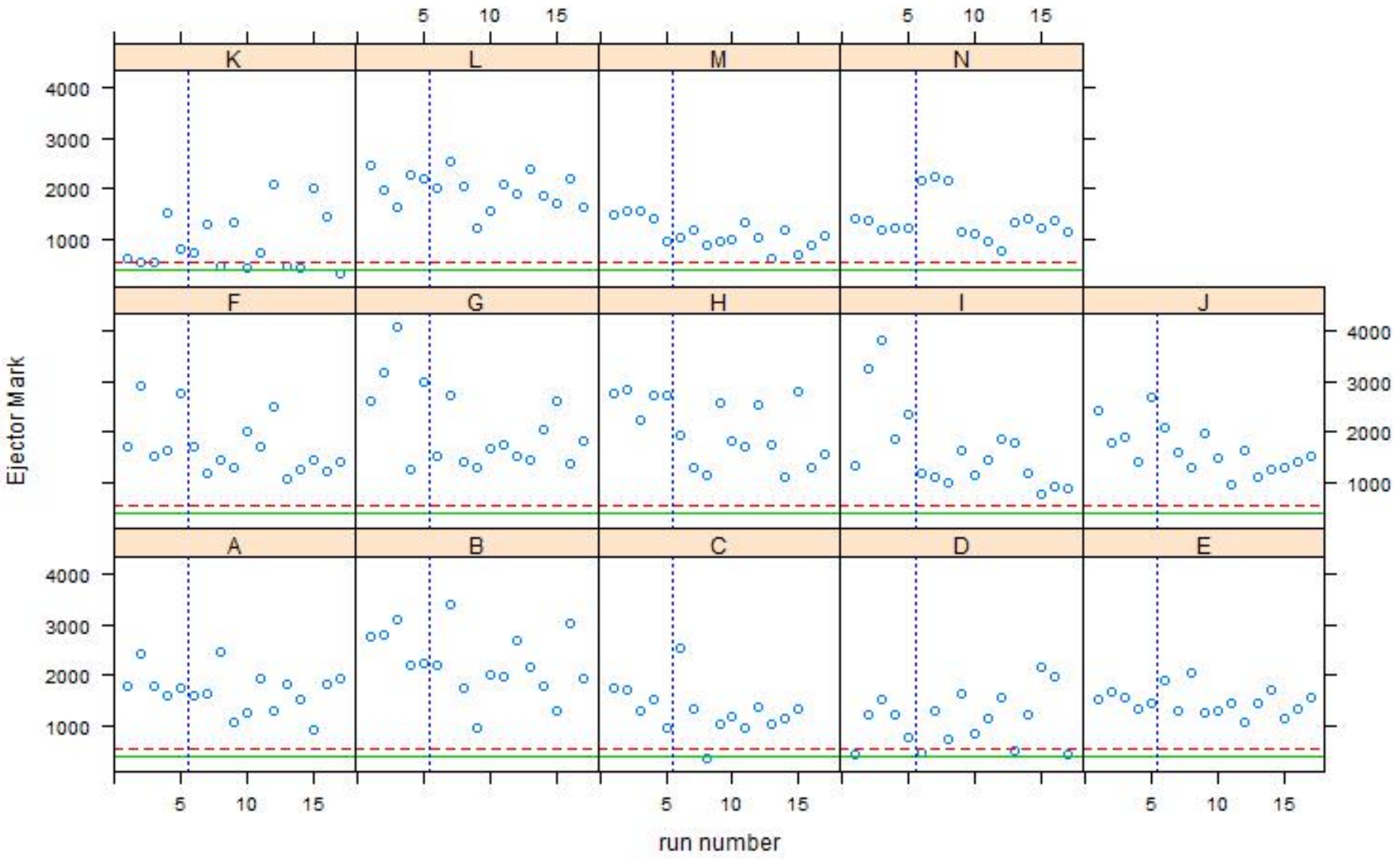
Breech face: proposed limits



Firing pin: proposed limits

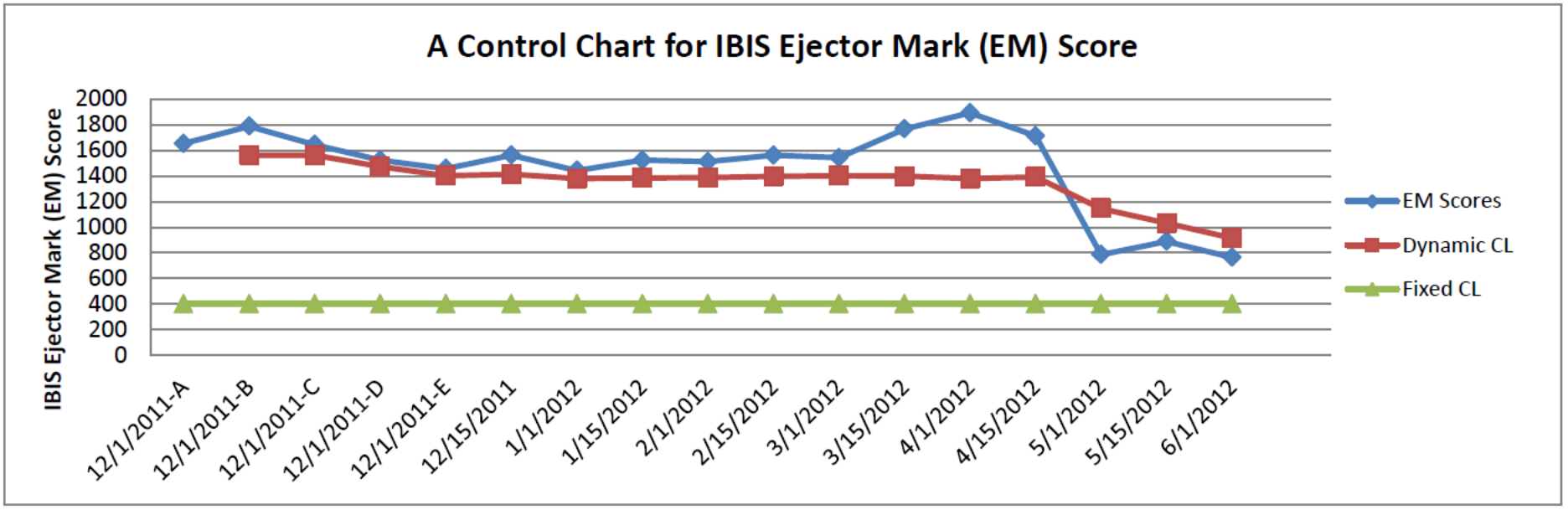


Ejector mark: proposed limits



Know thyself

- Operators should monitor their individual control charts, and those for their lab, in addition to the global control limits
  - a hypothetical control chart below: this situation may need looking into despite keeping above the control limit



Some Early Conclusions and Immediate Benefits

- Deviating even a little from procedure (e.g. angle of lighting, or alignment of lasers) can produce very low scores for bullets
- Doing the outline correctly important for casing marks (especially ejector marks)
- Anomaly in EM scores led to software fix for NIBIN
- Anomaly in BF for one operator led to revised manufacture for standard casing (gluing instead of soldering)

General Conclusions

- The NBIC studies have revealed areas of improvement for those involved in this area
- From NBIC-1 to NBIC-2: Large increase in FP scores, more modest relative increases for BF and EM scores (due both to change in Golden Image and change in system)
- The NBIC studies have enhanced quality assurance for NIBIN operations by describing approximate variability with control limits
- NIBIN stakeholders intend to re-emphasize quality maintenance efforts

Thank you for your attention…… and
Acknowledgements and Thanks to NIBIN operators who participated in the NBIC projects:

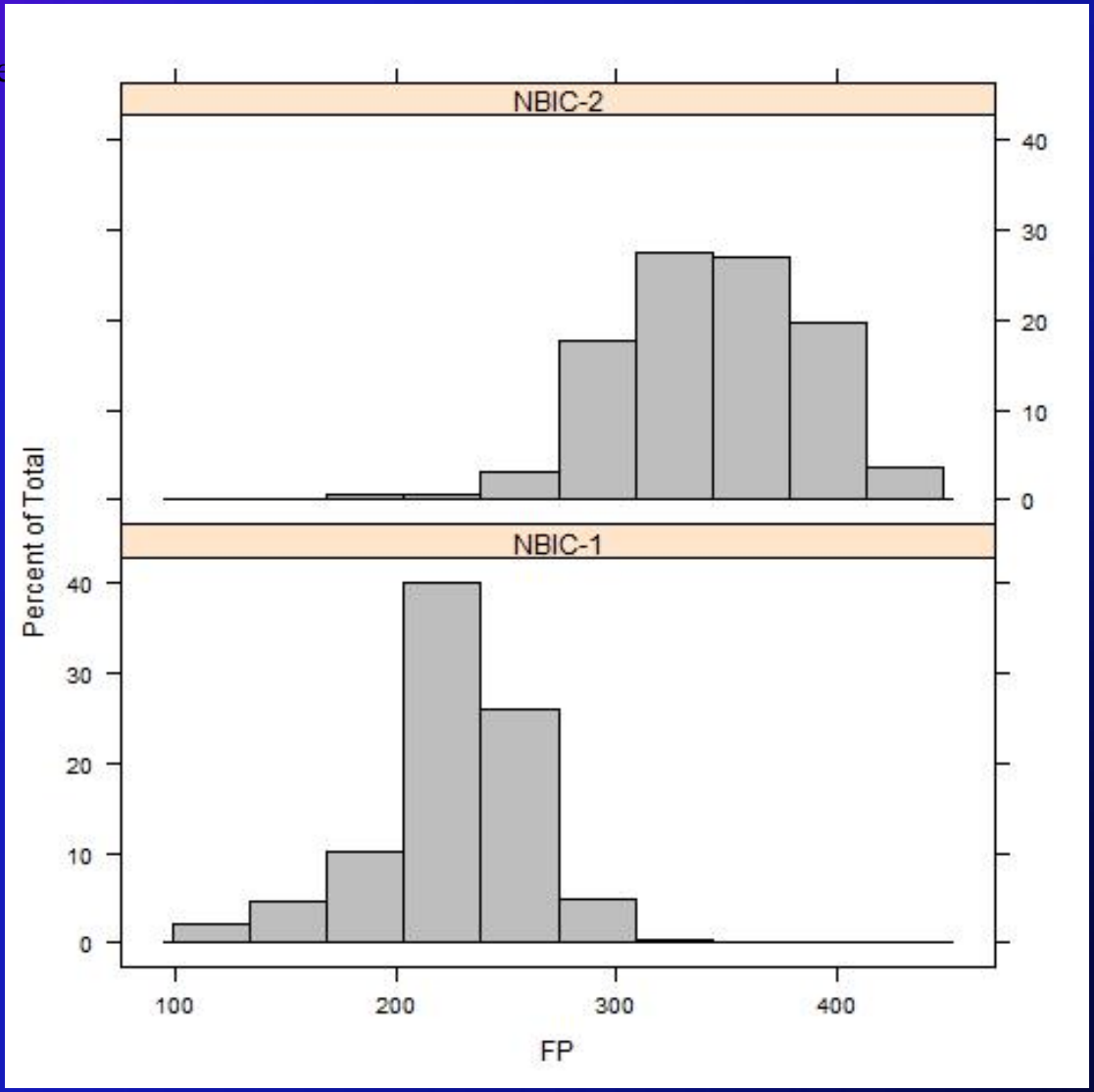
- NBIC-1: Ashleigh Vogel of Michigan State Police-Grand Rapids Laboratory, Edward Jachimowicz of CT Dept. of Public Safety Lab, Caryn Tucker, Daryl Smith and Calvin Box of Illinois State Police Forensic Science Center in Chicago, Jennifer Perry and Steve Hargis of Arkansas State Crime Lab, Jill R. Errickson of Miami-Dade Police Dept., Stephen M. Garten of ATF Forensic Science Lab-Atlanta, Martin Ols of ATF-Maryland, Charles Fancey and Brandy Harrington of Jefferson Co. Sheriff's Dept., Nanette Rudolph and Linzee Willette of FDLE Orlando, FL, Darrell Stein of Houston Police Dept. Crime Lab., Helen Schumacher, Allison Milam and Susan Landen of VA Division of Forensic Science, and Erich Smith of FBI.

- NBIC-2: W. Best, Alleghany County; J. Flaskamp, Illinois State Police; M. Ols, Hine, A. Northrop, ATF-Maryland; S. Garten, ATF-Atlanta; M. Shaw , ATF-Walnut Creek; K. Lancaster, FDLE Tampa ; L. Willette , S. Schoonover , FDLE Orlando; E. Lawton-McWhite , A. Greenspan, Broward Sheriff's Office; E. Osbeck, Columbus Crime Lab; T. Lyons , Houston PD; K. Lancaster, FDLE Tampa.

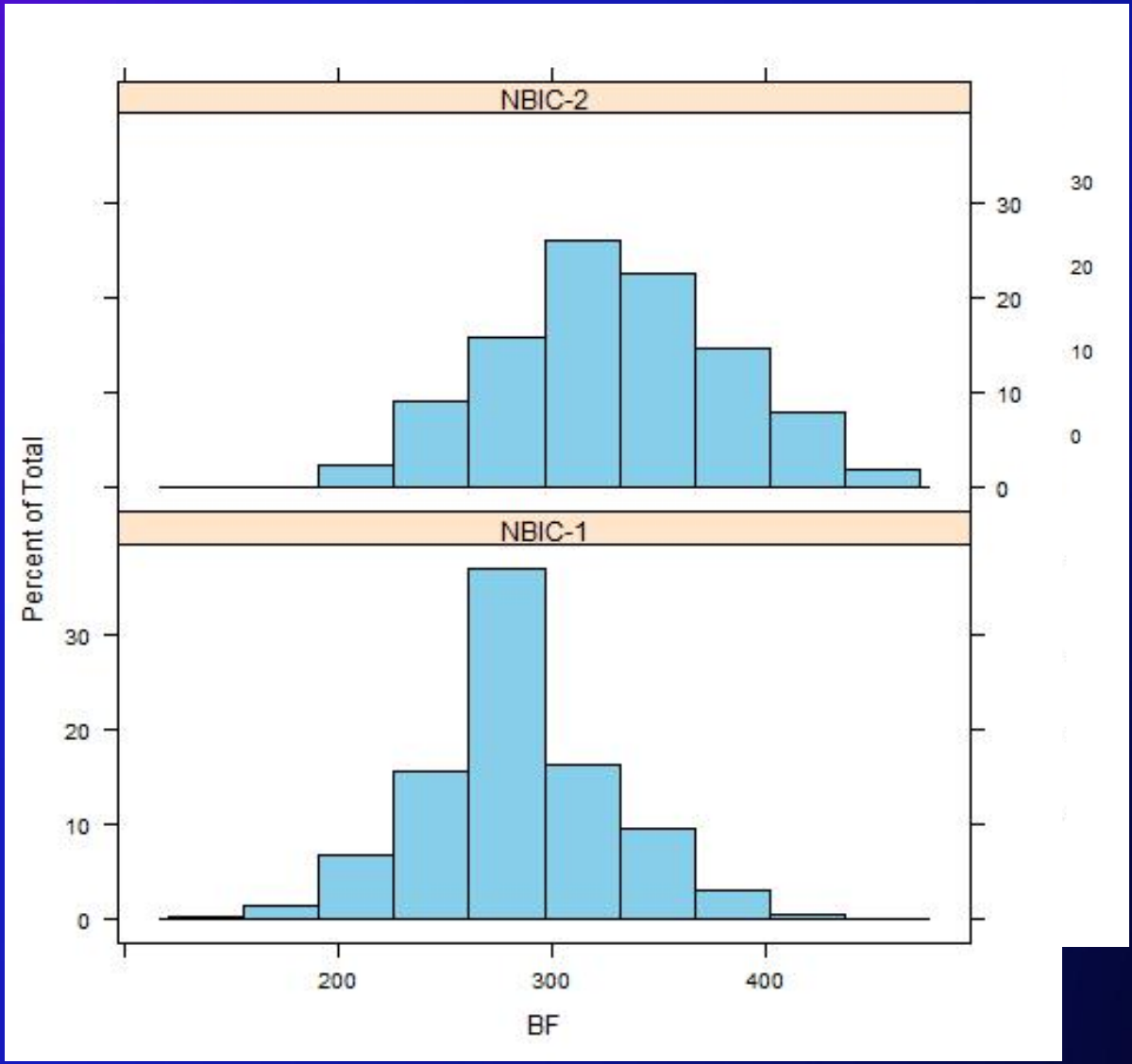
Extra Slides

NBIC-2 Firing pin scores much higher than in NBIC-1

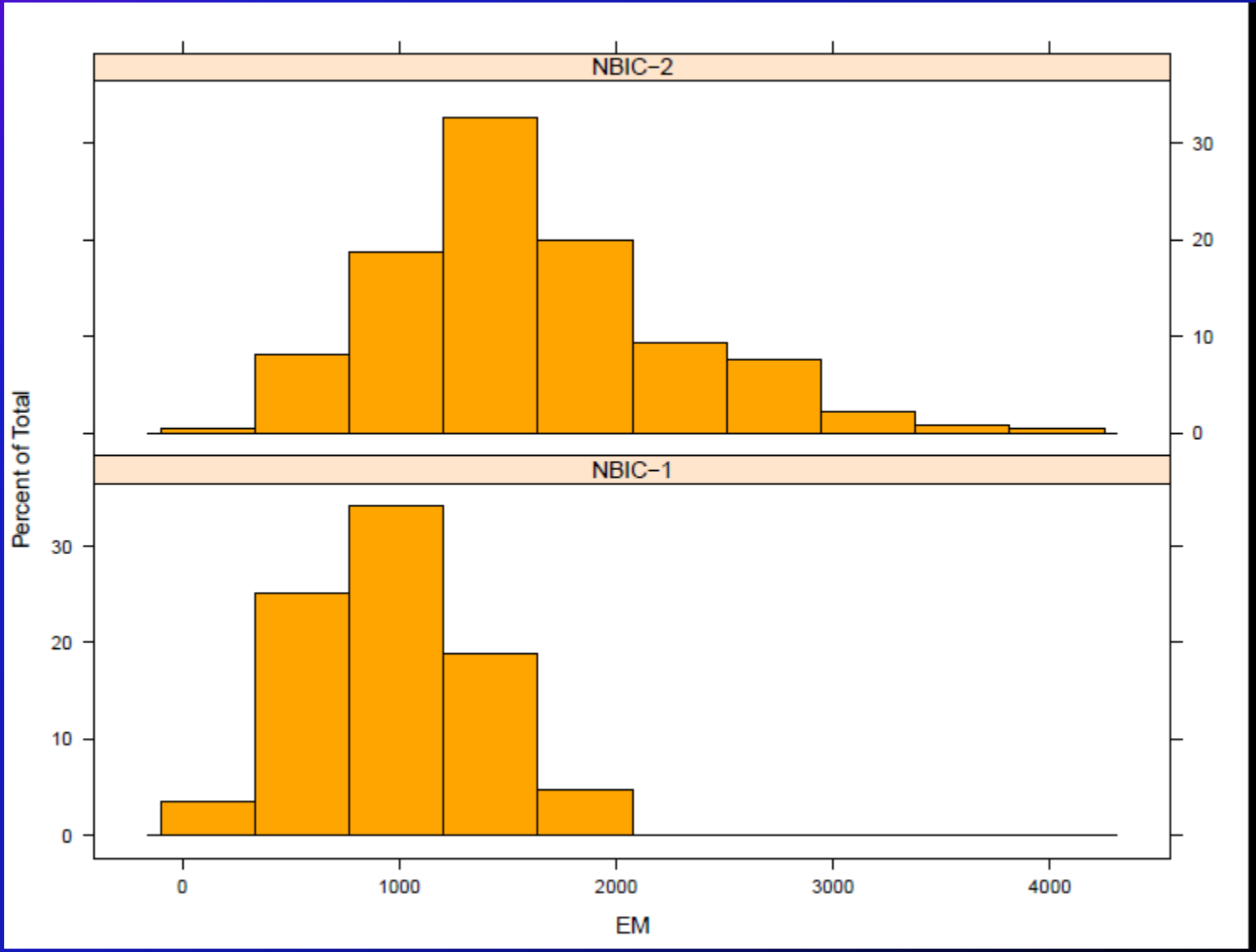
NBIC2 FP score
than NBIC1



NBIC-2 breech face scores similar to, somewhat higher than NBIC-1



NBIC-2 ejector mark scores higher on average, with longer tail (more very high scores) than NBIC-1



Random/Mixed effects models

- A more realistically detailed model would have random or mixed effects:
  - Each Unit's score is randomly distributed around his or her Unit Mean
  - The Unit Means are randomly distributed around a Grand Mean
  - In the fitted model, Within-Unit variation comparable to Between-Unit variation (within-variation larger for Ejector mark, between-variation larger for firing pin)
    - Different units tend to have different variances