STAR Isobar Blind Analysis Method

James L. Drachenberg
for the STAR Collaboration
BNL Isobar Results Seminar
August 31, 2021

OUTLINE
• Considerations
• Data-taking for blind analysis
• Analysis blinding steps
• After unblinding
• Summary
Important Considerations

For STAR Chiral Magnetic Effect (CME) analyses:

• **Critical to account for**
  – *Time-dependent* detector fluctuations
  – Anomalies in the collection of 30-minute “runs” of the data acquisition system

• Do not randomize variables that may severely compromise analysis quality
  – *E.g.*, randomizing the sign of reconstructed charged-particle signals prevents charge-dependent efficiency corrections

• 2018 data-taking used frequent switching of “isobar” species ($^{96}_{44}$Ru + $^{96}_{44}$Ru and $^{96}_{40}$Zr + $^{96}_{40}$Zr)
  – Species expected to have comparable behavior, e.g., luminosity, trigger, energy, vertex distribution, occupancy of tracks
  – *Possible to blind species by interleaving or “mixing” events from two species*

• Certain non-analyst experts need access to un-blind data
  – *E.g.*, STAR detector experts during RHIC running or offline calibration experts
  – *All must recuse themselves from blind physics analysis*

• Selection of high quality runs for analyses must proceed prior to mixing of events
**Vital Stats**

- 2017 BNL NPP Program Advisory Committee recommended **blind analyses** of **CME studies** of Run-18 isobar data
- Published analysis blinding manuscript:
  
  *Methods for a blind analysis of isobar data collected by the STAR collaboration, J. Adam et al. (STAR Collaboration), Nuclear Science and Techniques 32, 48 (2021).*

- **Methods developed and accepted by collaboration in January 2018, well before 2018 data-taking**
  
  - **Step-0, Initial steps**
    - Calibrations and quality assurance (QA) of data acquisition “runs” by calibration experts
    - “Mock data challenge”: *Sanity-check of feasibility and implementation*
  
  - **Step-1, “The Reference”**
    - Provide output files composed of collision data from a *mix* of the two isobar species
    - As much as possible, order of collision “events” *respects time-dependent changes in detector conditions*
    - Analysis code and time-dependent QA tuned and frozen
  
  - **Step-2, “The run by run QA sample”**
    - Provide files that blind the isobar species but do not “mix” data from different data acquisition runs
    - Only allow “run-by-run” corrections and code alteration directly resulting from these corrections
  
  - **Step-3, Full un-blinding**
Data-taking for Isobar Collisions

**RHIC Running**

- Switch isobar species each time beam is inserted into RHIC
- Stable luminosity (matched between species) with long (~20 hour) beam circulation time
- Adjust and level luminosity to optimize data collection rate while minimizing backgrounds and systematics
- Restrict species-related information to those necessary for successful data-taking
- Calibration experts (recused from CME analyses) evaluate data quality “in real time”
Step-0: Initial Steps

“The Tune-up”

- Calibrations and quality run selection by un-blind experts
- Develop software infrastructure to implement the blinding procedure
  - Event mixing procedure and run-numbers encrypted
  - Additional information obfuscated in data
    • Event ID, run ID, event timestamp, collision species, hit/coinccidence/background rates from certain detectors
- “Mock data challenge”
  - Sanity-check of feasibility and implementation
  - Utilize blinding procedures on 2018 27 GeV Au+Au data
  - Analysts tune code on “mock data”
    • Check that data blinding infrastructure works as intended
    • Verify the appropriate information is blinded as intended
    • Ensure appropriate information is accessible to analysts
    • Check that analysis codes run properly on “blind” data structures
    • Confirm “blind” and “unblind” results are the same
      - sanity check of procedures
Step-1: Isobar Blind and Mixed

“The Reference”

- Provide output files composed of events from a mix of the two isobar species
  - Mixing procedure encrypted and known only by two computing experts (recused)
- As much as possible, order of events respects time-dependent change in run conditions
- Analysis code and time-dependent QA tuned
- Critical analysis needs enabled by this step:
  - Extraction of time-dependent spectra for quality assessment
  - Detection of time-dependent anomalies
  - Measurement of peak widths relevant to momentum resolution

Following completion of Step-1, analysis codes are frozen and committed to the repository
Before moving to Step-2, codes are documented and reviewed by the isobar paper review committee
Step-2: Isobar Blind

“The run by run QA sample”

• Provide data files that obscure the species but do not mix events across different runs
  – Limit the number of events to prevent deciphering species by simple counting
• Only run-by-run corrections and code alteration directly resulting from these corrections are allowed at this stage
• Additional bad runs identified based on physics quantities and discarded
  – Analysts perform run-by-run QA using a predefined and frozen algorithm
• This step enables analysts to perform QA using quantities relevant to their specific analysis

Following completion of Step-2...

• Analysis codes are reviewed, frozen, and committed to the repository
• Fully un-blind data are released and analyzed with the frozen codes
• Only changes to correct “mistakes” are allowed after unblinding
  – Errors in arithmetic
  – Unintended departures from documented and approved procedures, cuts, corrections, and systematic uncertainty estimates
• **STAR has developed a procedure for the CME isobar blind analyses**
  - Step-0: Calibrations, run-QA, and mock data challenge
  - Step-1: Isobar blind and mixed (analysis codes tuning)
  - Step-2: Isobar blind and un-mixed (run-by-run QA and correction)
  - Step-3: Full un-blinding (physics analysis)

• Development and implementation has been a substantial, collective undertaking
  - Innovative RHIC running
  - New software and computing infrastructure
  - Cooperation across analysis groups, physics working groups, committees, etc.

*Thank you to all who supported the effort!*