



Presented at the 2022 Conference on Test Security
October 26-28 Princeton, NJ

Visualizing the Prevalence of Collusion/Content Exposure in the Testing Population

Brett P. Foley, Ph.D.
Senior Psychometrician

VALIDITY
Fair, Reliable, Secure

Visualizing the Prevalence of Collusion/Content Exposure



- Score/Response Similarity Analyses increasing in popularity
- Uses pairwise expected/observed matches
- Effective for identifying collusion & content exposure
- Challenging to convey extent of problem in population

Session Topics

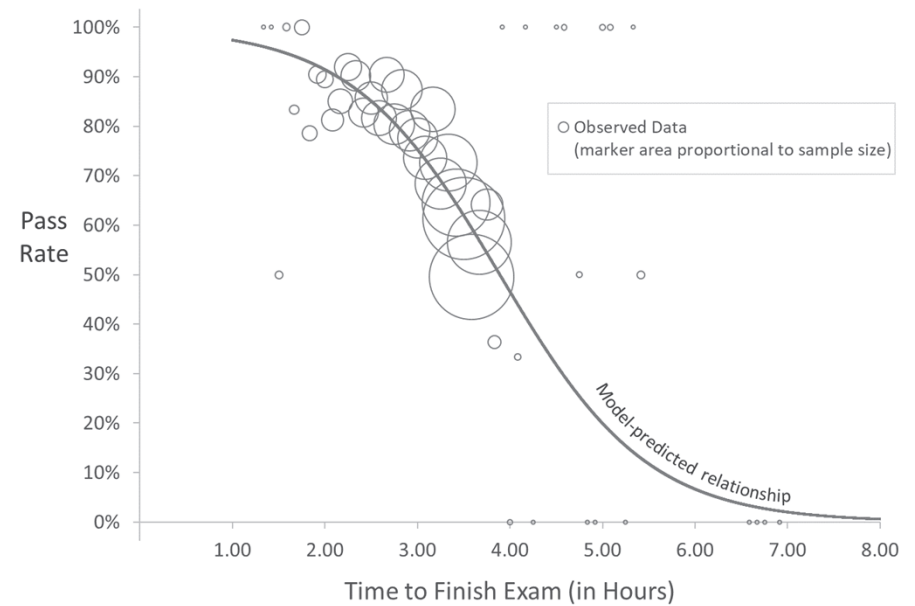
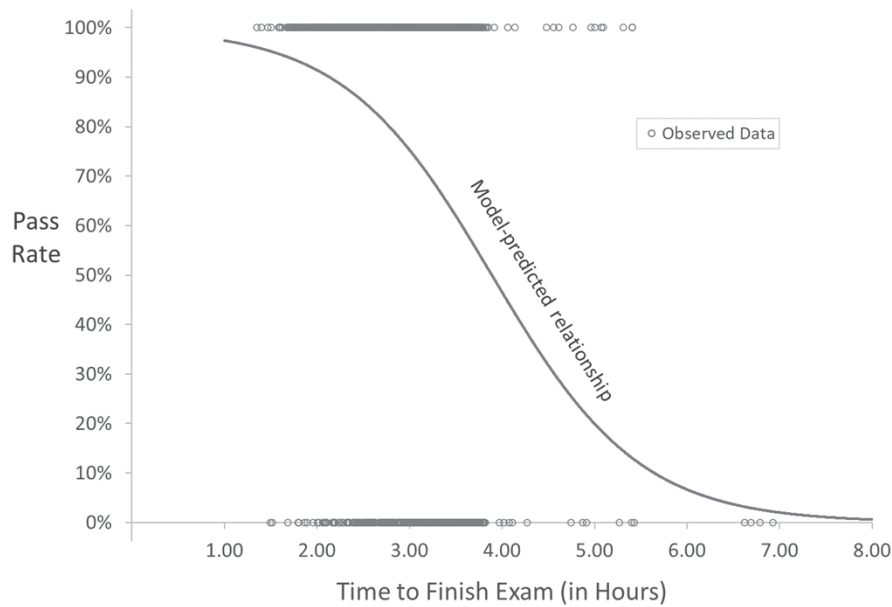
- Best practices for exam-security visual displays
- Examples of visual displays for score/answer similarity
- Scaling strategies to highlight outliers or emphasize common patterns

Best Practices for Exam-Security Visual Displays

VALIDITY
Fair, Reliable, Secure

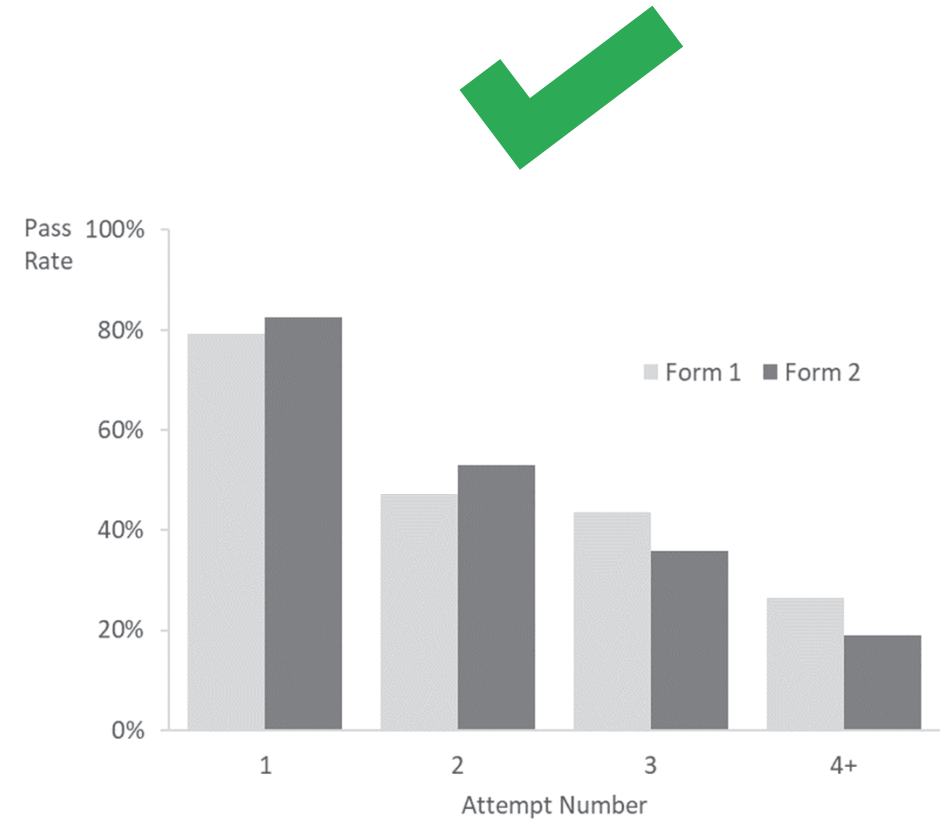
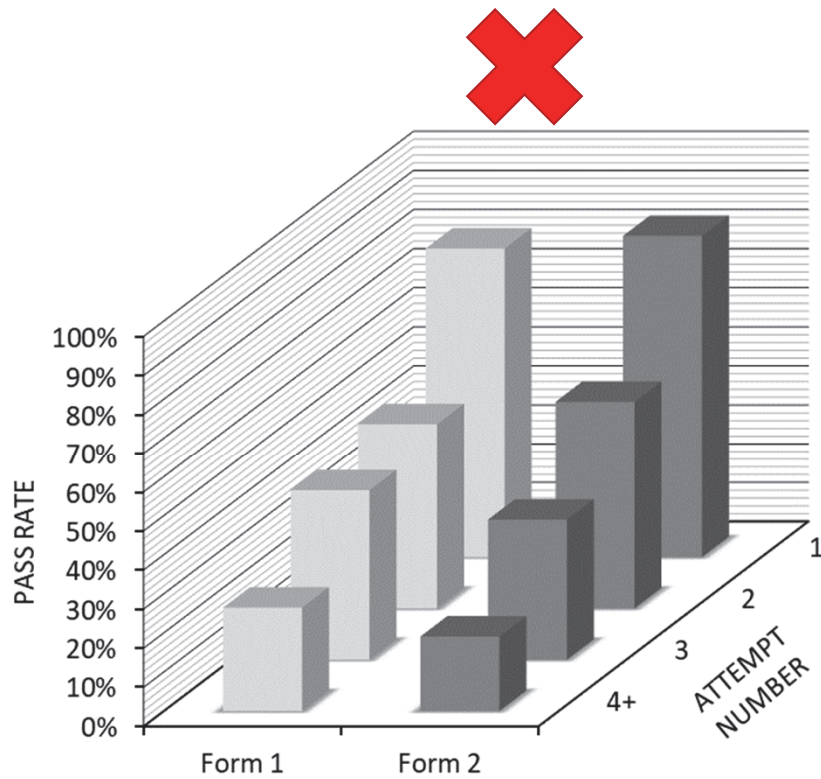
Best Practices

Choose displays that are appropriate for the data



Best Practices

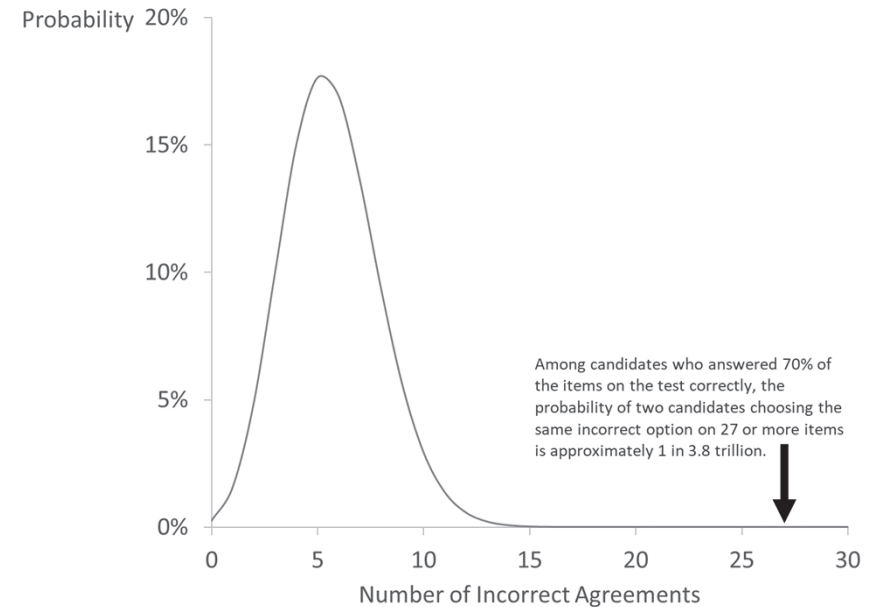
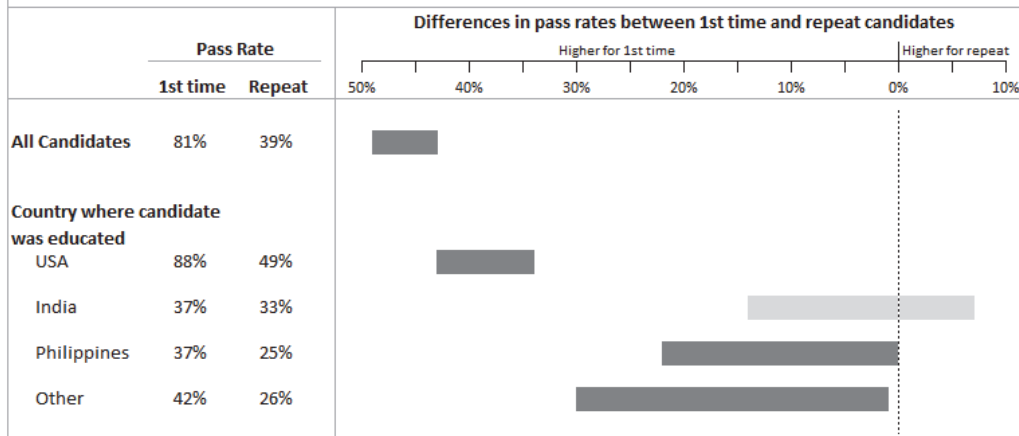
Aim for simplicity



Best Practices

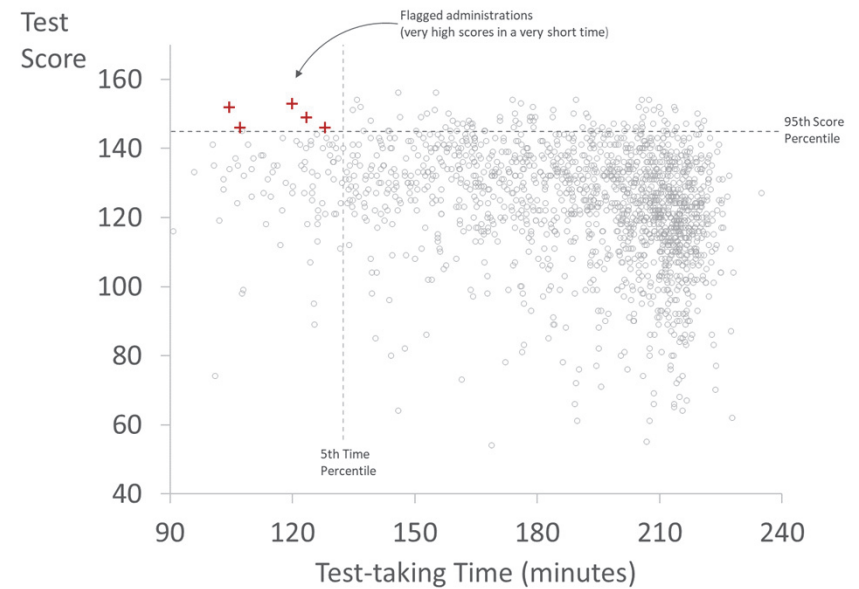
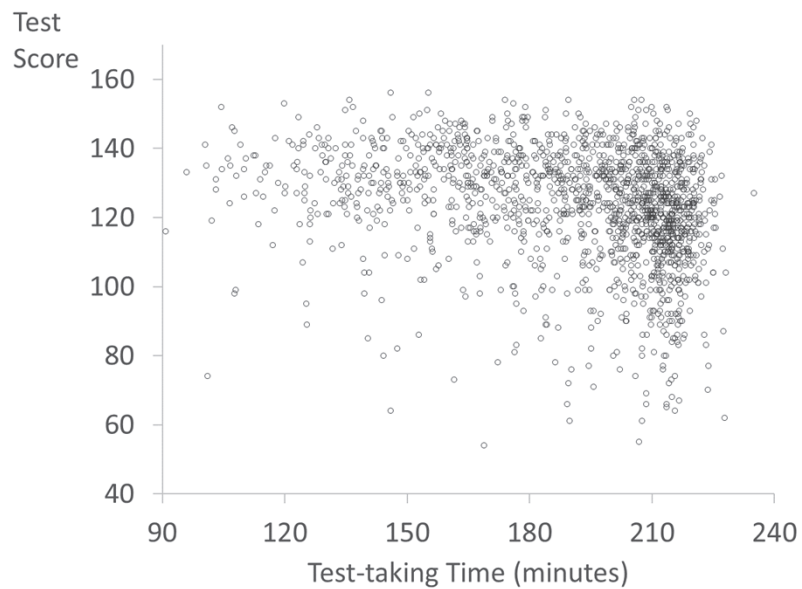
Integrate text, numbers, and figures

This figure shows comparisons of pass rates for first time and repeat test takers for several different countries. The graphic on the right illustrates ranges that, with 95% certainty, capture the true difference in the pass rates between the two types of candidates within each country. If the range is completely on the "Higher for 1st time" or "Higher for repeat" side of the dashed line, the difference between the two pass rates is significant.



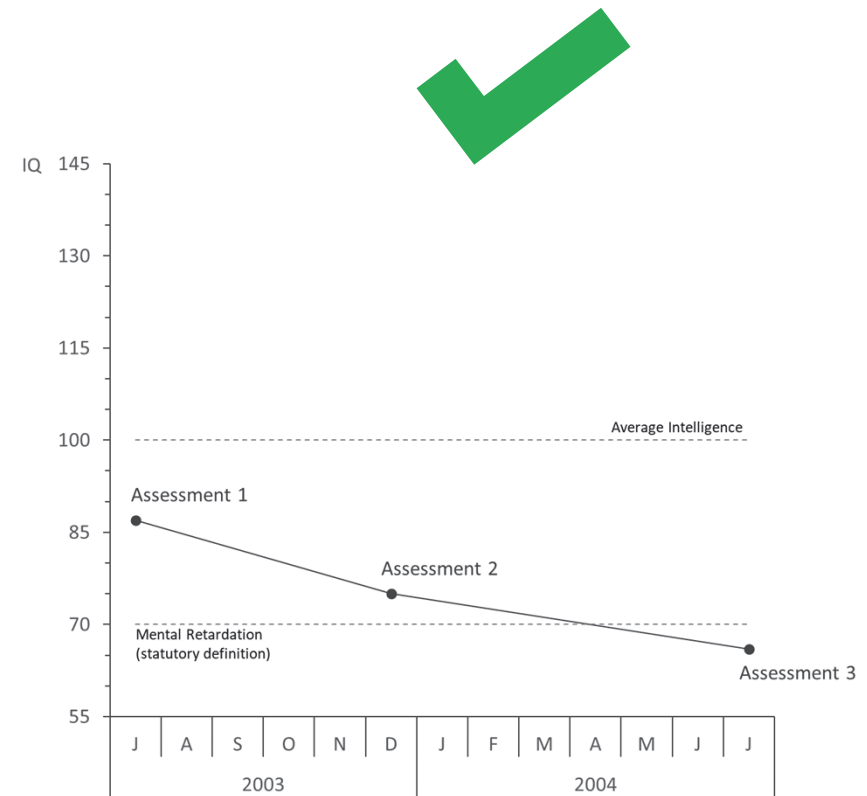
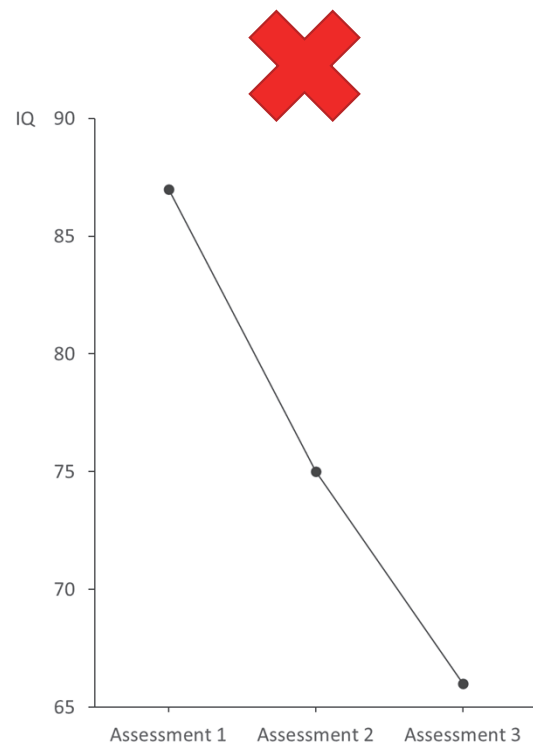
Best Practices

Highlight what is important



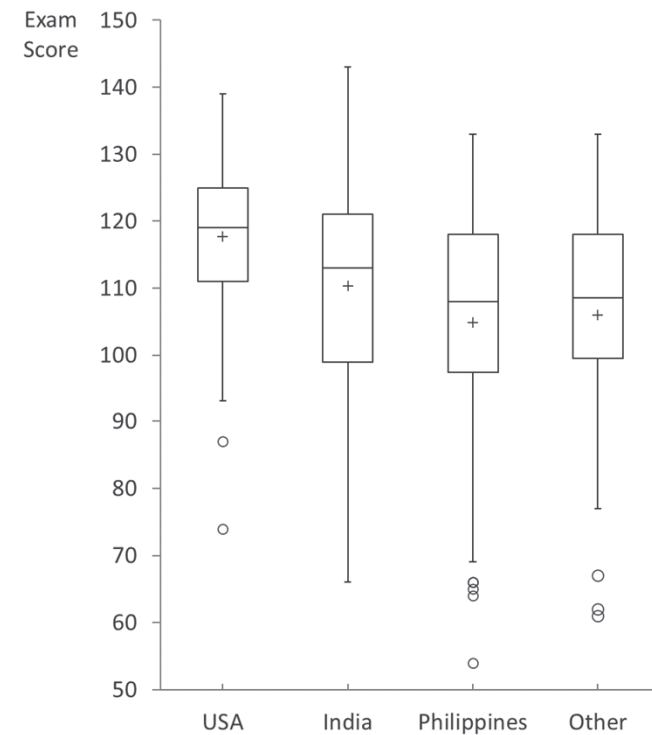
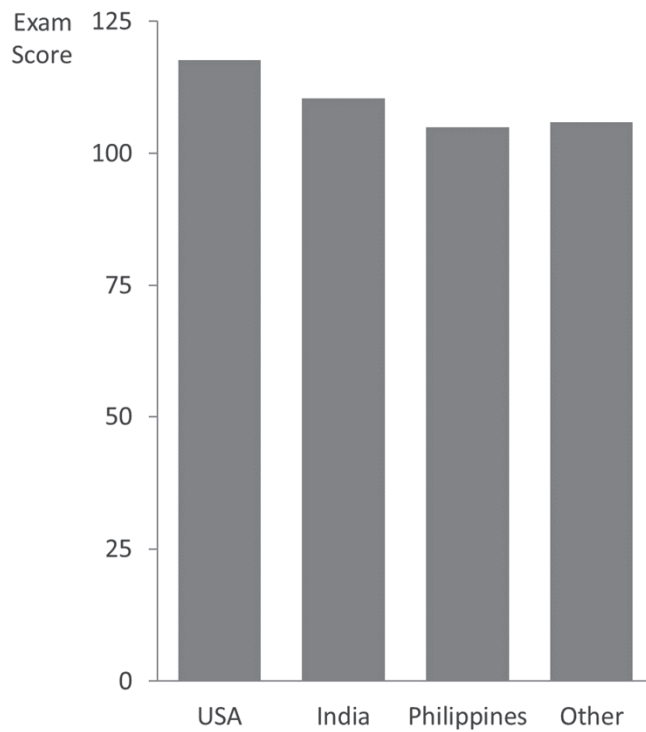
Best Practices

Do not intentionally mislead



Best Practices

Pair the design with the audience



Visual Displays for Score/Answer Similarity

VALIDITY
Fair, Reliable, Secure

Score/Answer Similarity

- Data Structure
 - Candidate Pairs
 - Expected vs. Observed
- Data Challenges
 - 1,000 Candidates -> 499,550 Candidate Pairs
- More Effective to Cluster Around Scores
 - 100 items -> 4,950 possible expected/observed combinations
 - Many/most will never occur.

Data Restructuring

Candidate 1	Candidate 2	Expected Matches	Observed Matches
A	B	66.24	54
A	C	67.73	70
A	D	66.65	63
B	C	56.14	56
B	D	58.85	51
C	D	56.36	51

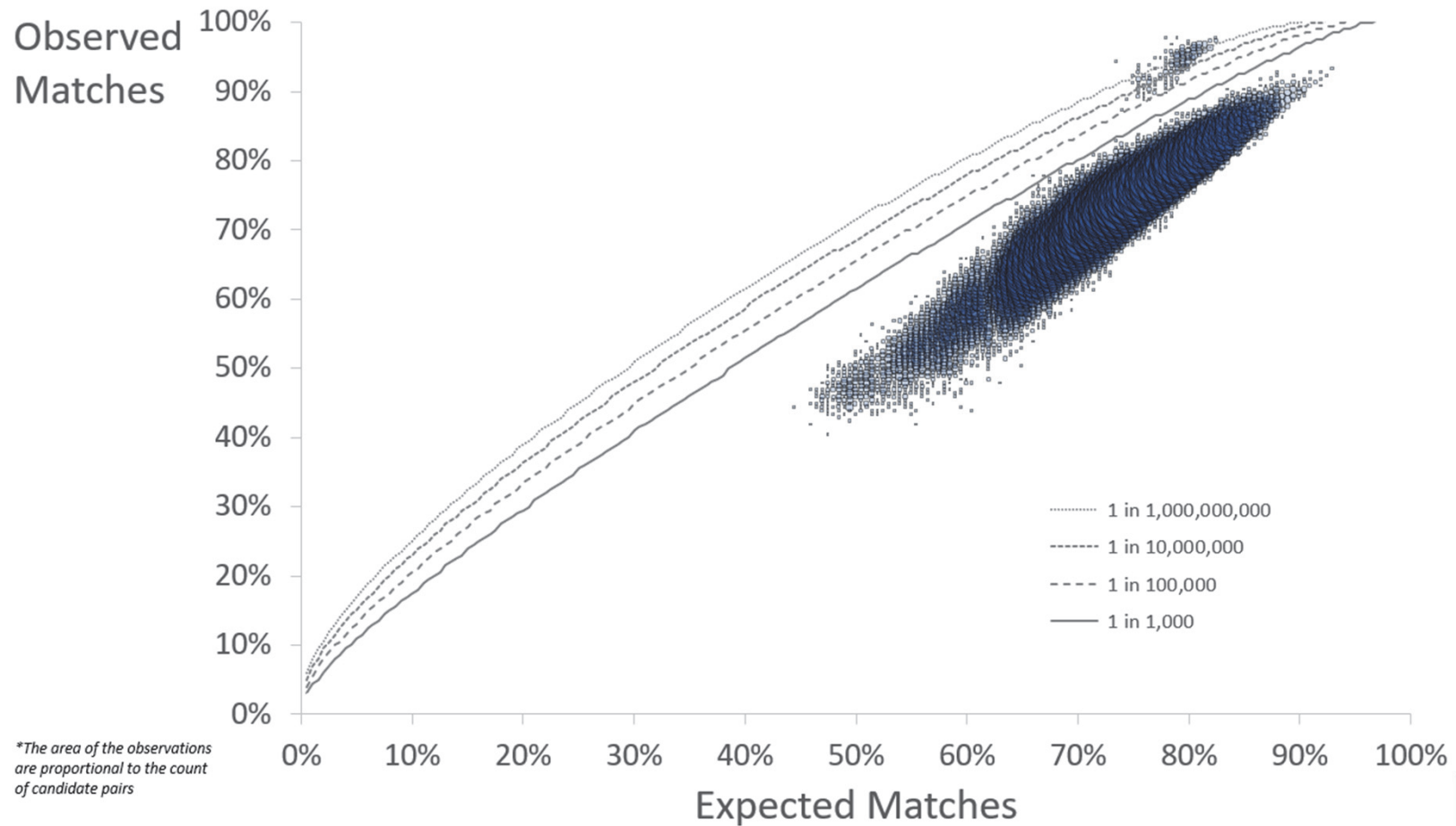
Round to whole numbers

Count candidate pairs for each expected/observed combination

Expected Matches	Observed Matches	Candidate Pairs
50	50	20
50	51	45
50	52	10
50	53	5
50	54	2
50	55	0

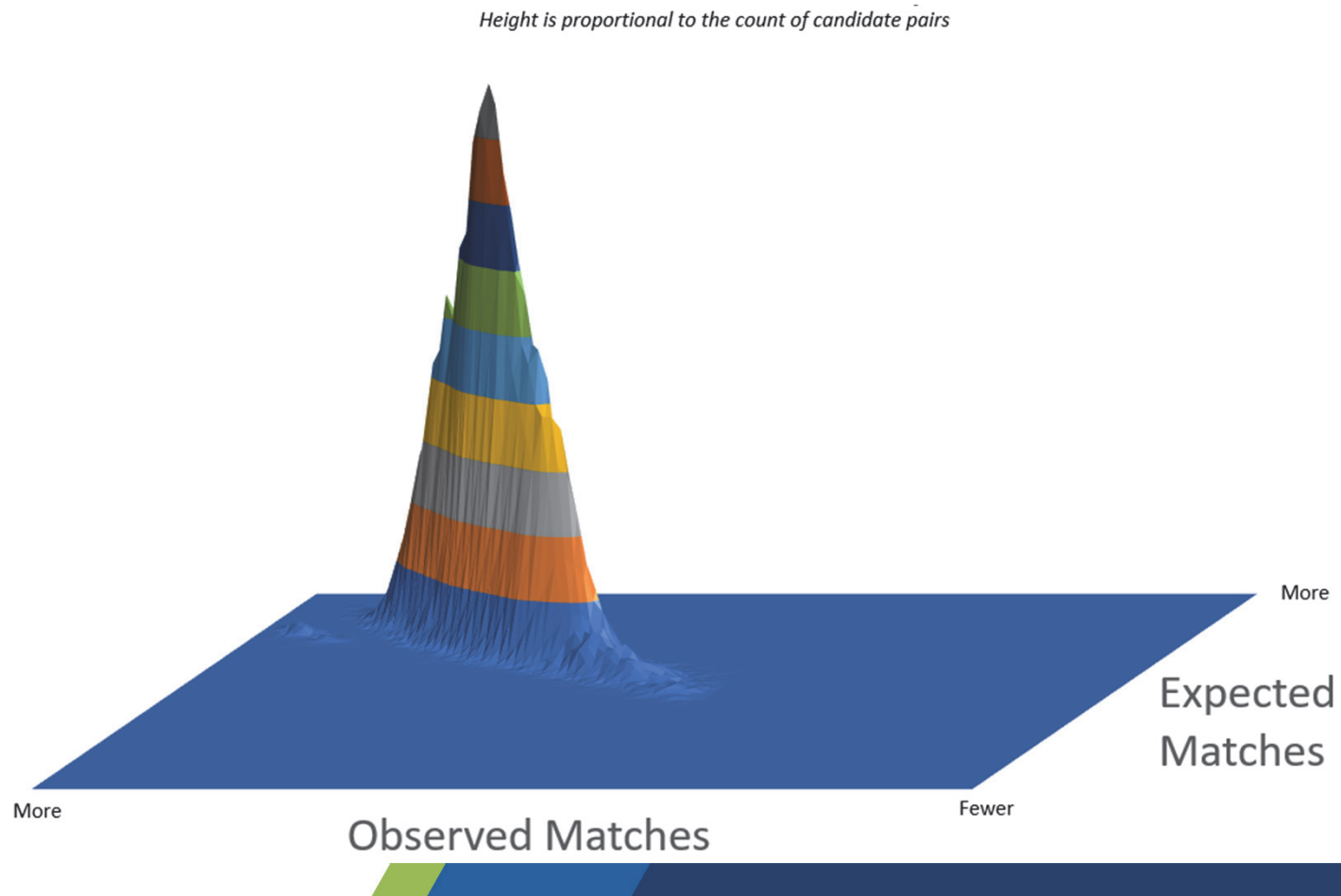
Flavor 1

Bubble Plot with Semi-Transparent Shading



Flavor 2

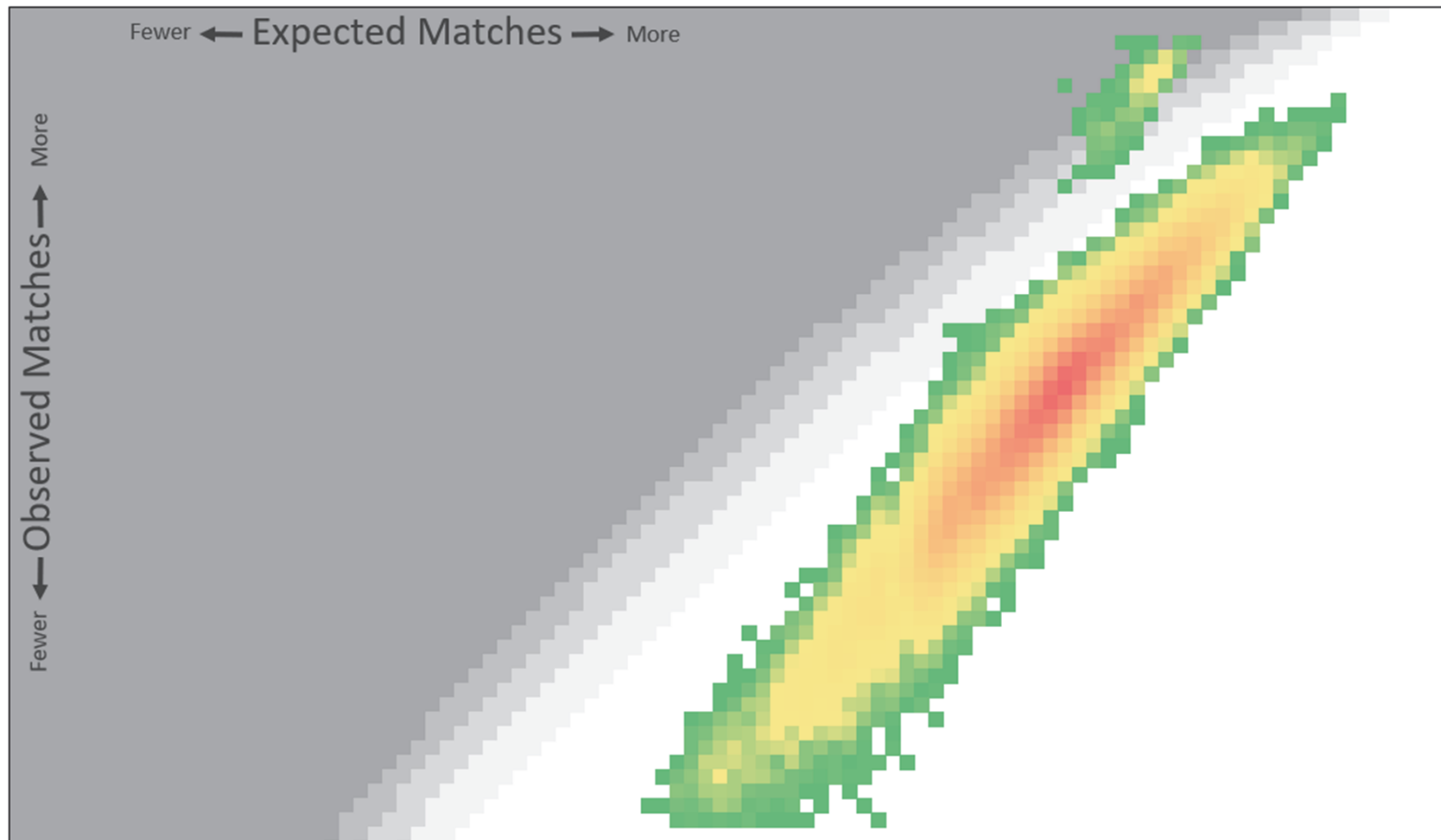
Rotated Surface Plot



Flavor 3

Heat Map (a.k.a., Doppler Radar)

Red = more pairs of candidates Green = fewer pairs of candidates



Scaling to Direct Focus

VALIDITY
Fair, Reliable, Secure

Count Scaling Example

Expected Matches	Observed Matches	Candidate Pairs	Sqrt of Candidate Pairs	Candidate Pairs ²
50	50	20	4	400
50	51	45	7	2025
50	52	10	3	100
50	53	5	2	25
50	54	2	1	4
50	55	0	0	0



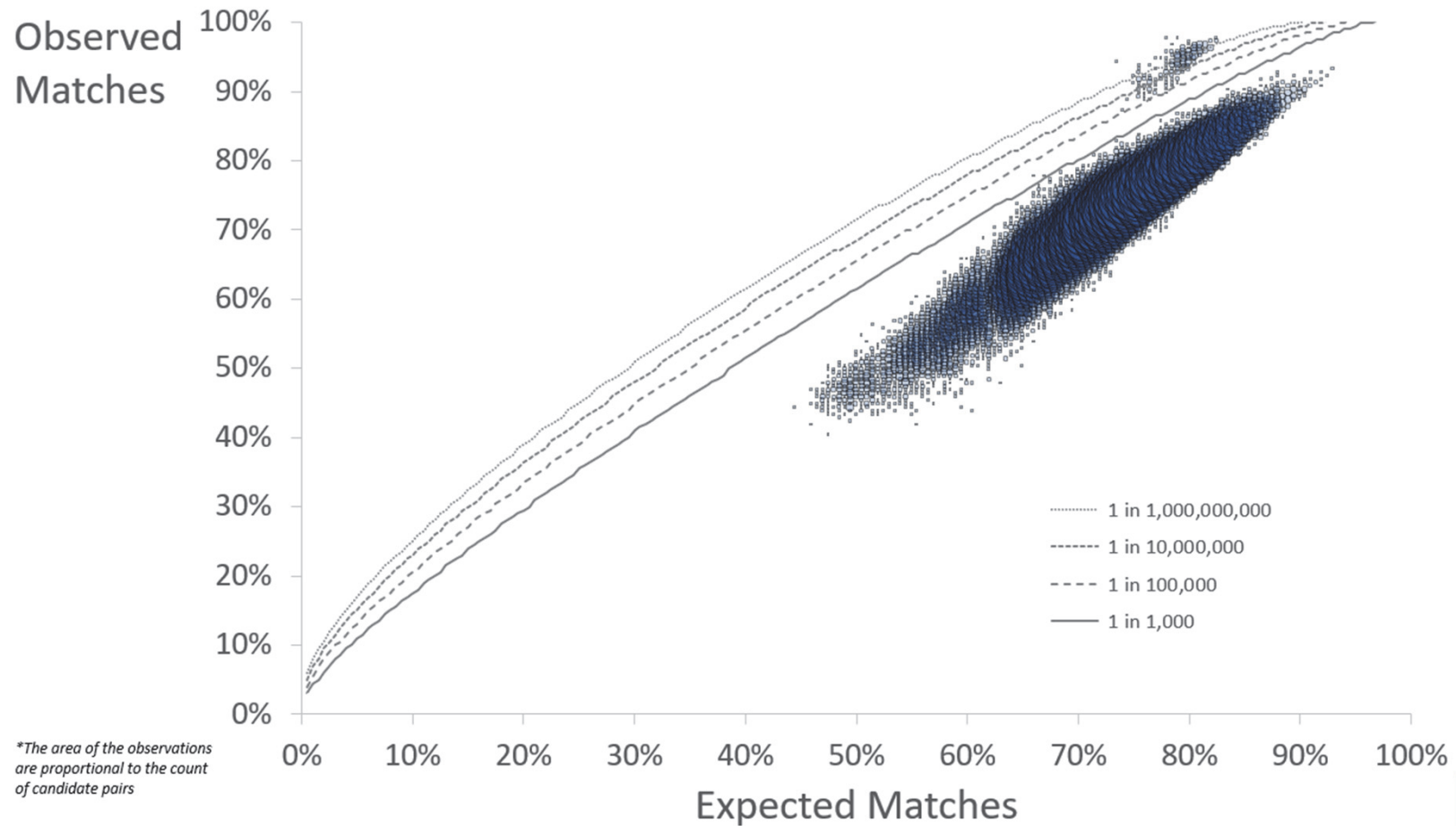
Diminishes group differences



Exaggerates group differences

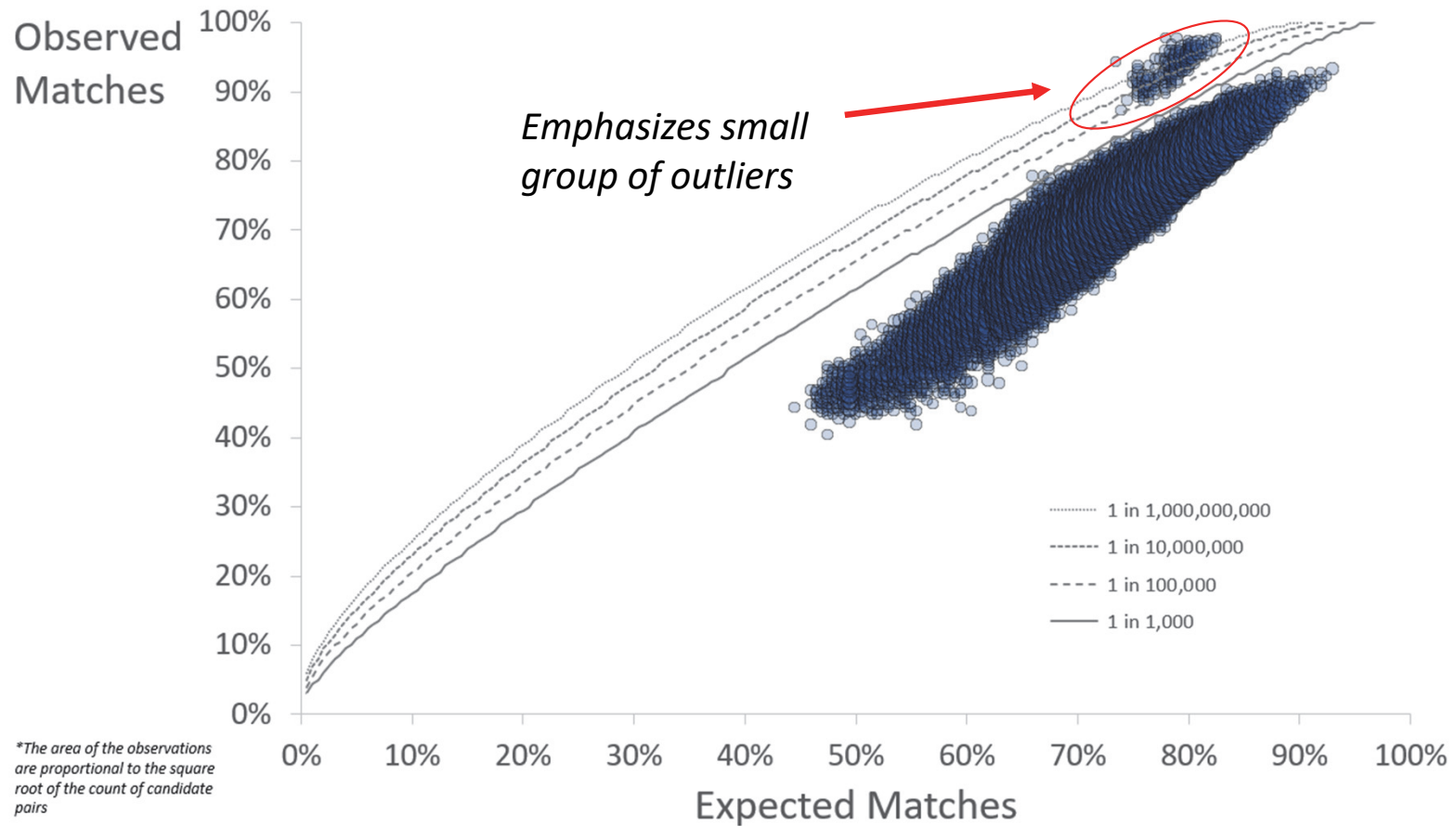
Flavor 1

Bubbles proportional to counts



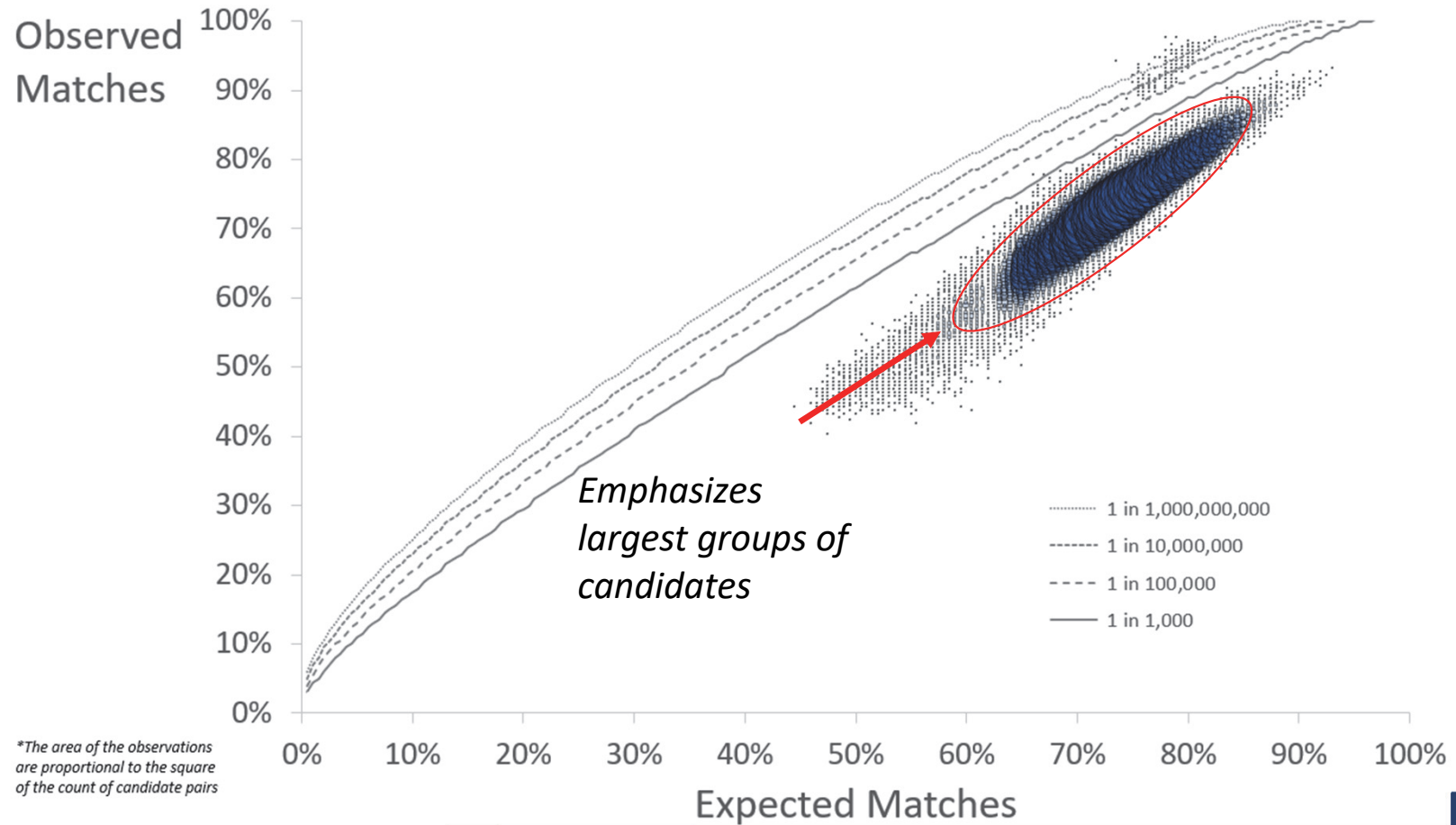
Flavor 2

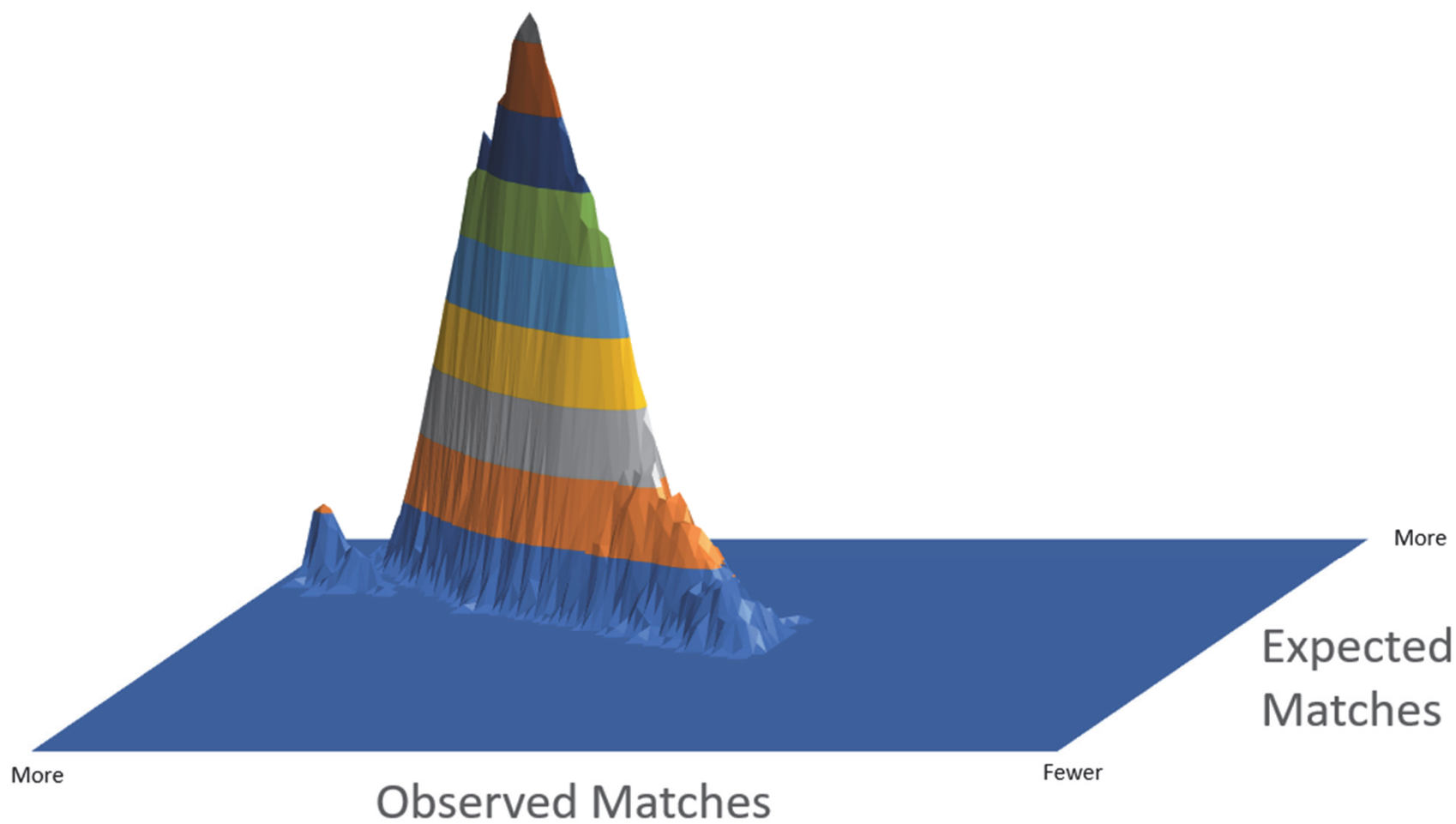
Bubbles proportional to square root of counts



Flavor 3

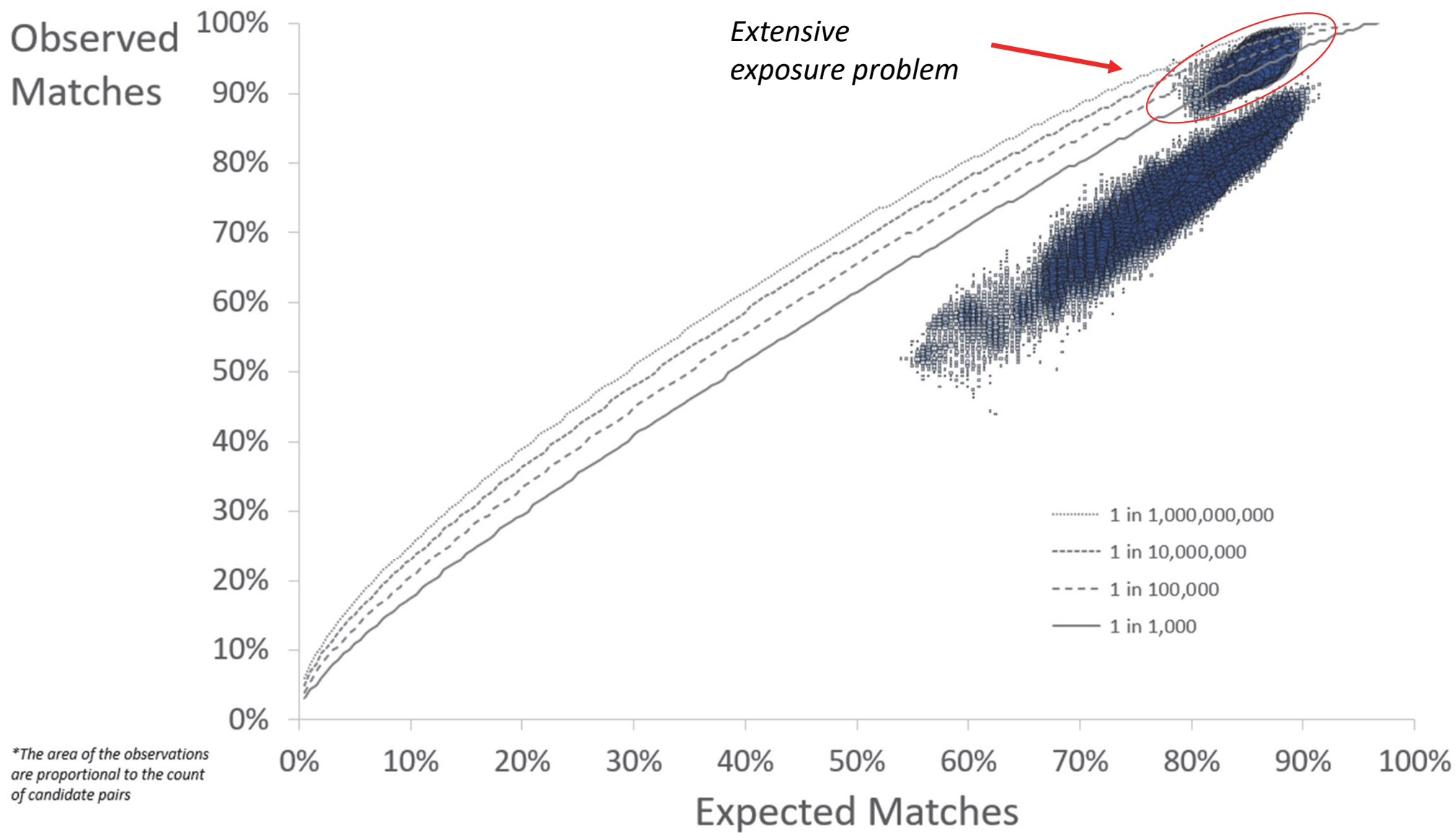
Bubbles proportional to square of counts





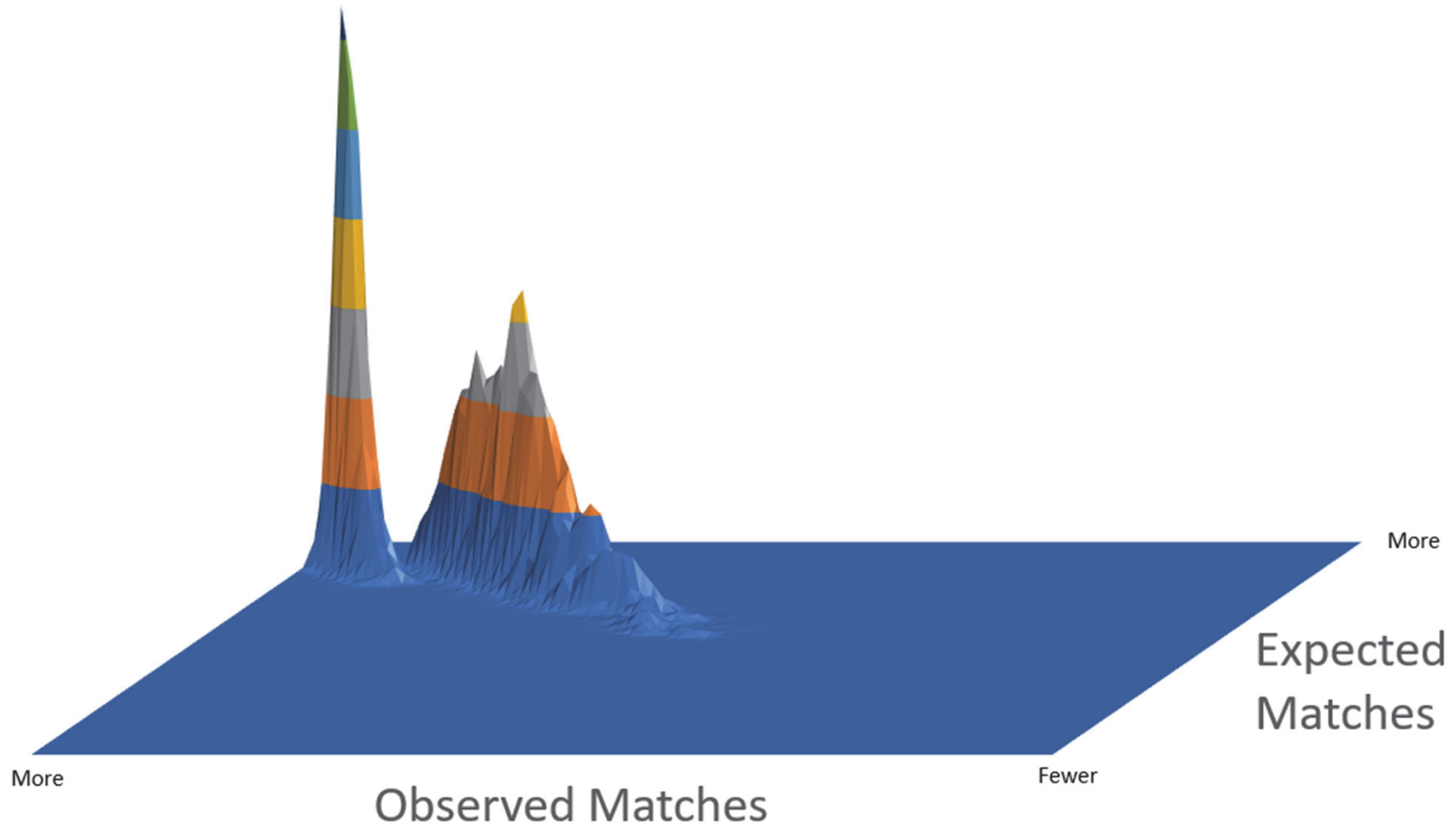
More Examples

VALIDITY
Fair, Reliable, Secure



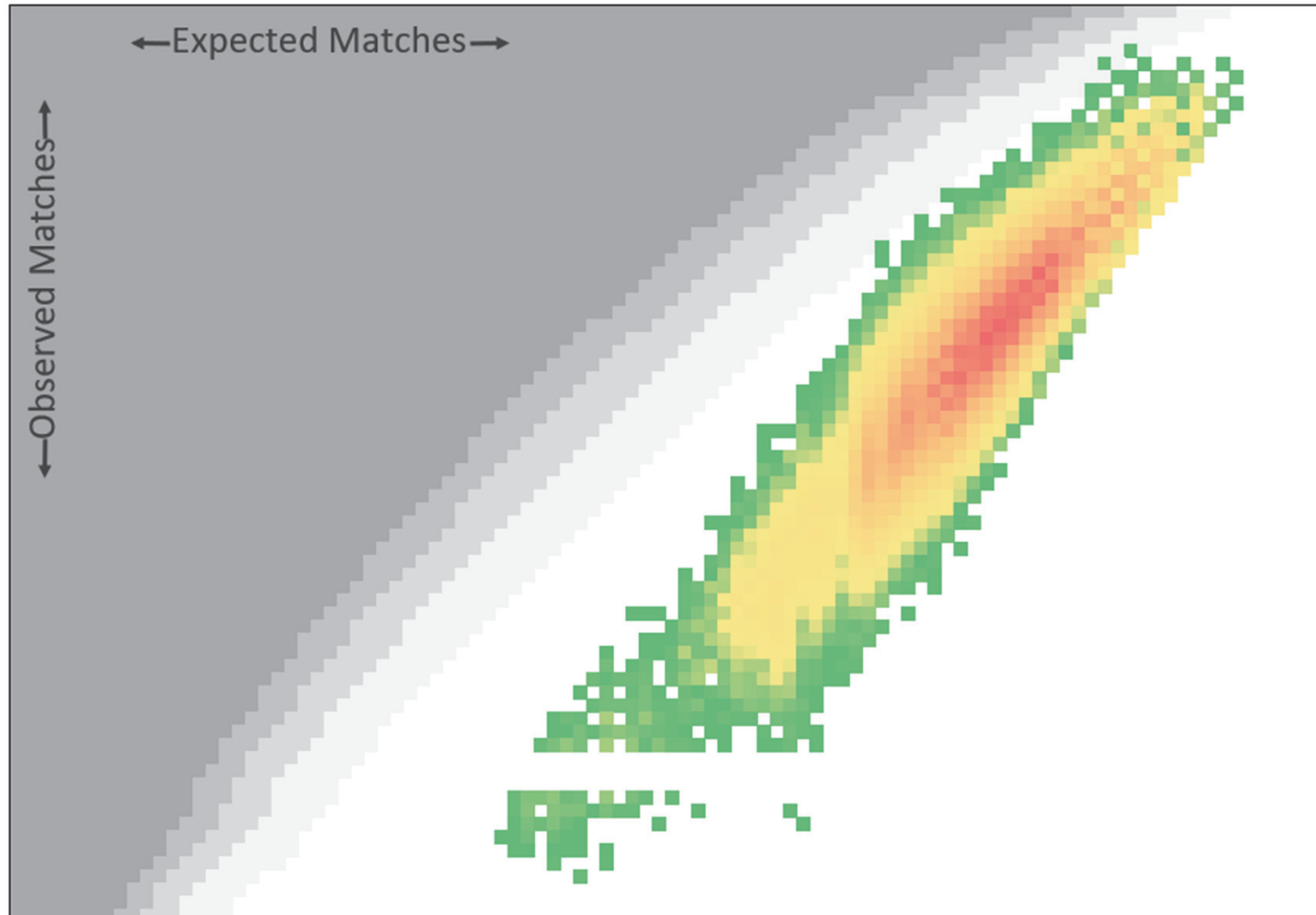
Rotated Surface Plot of Observed vs Expected Matches

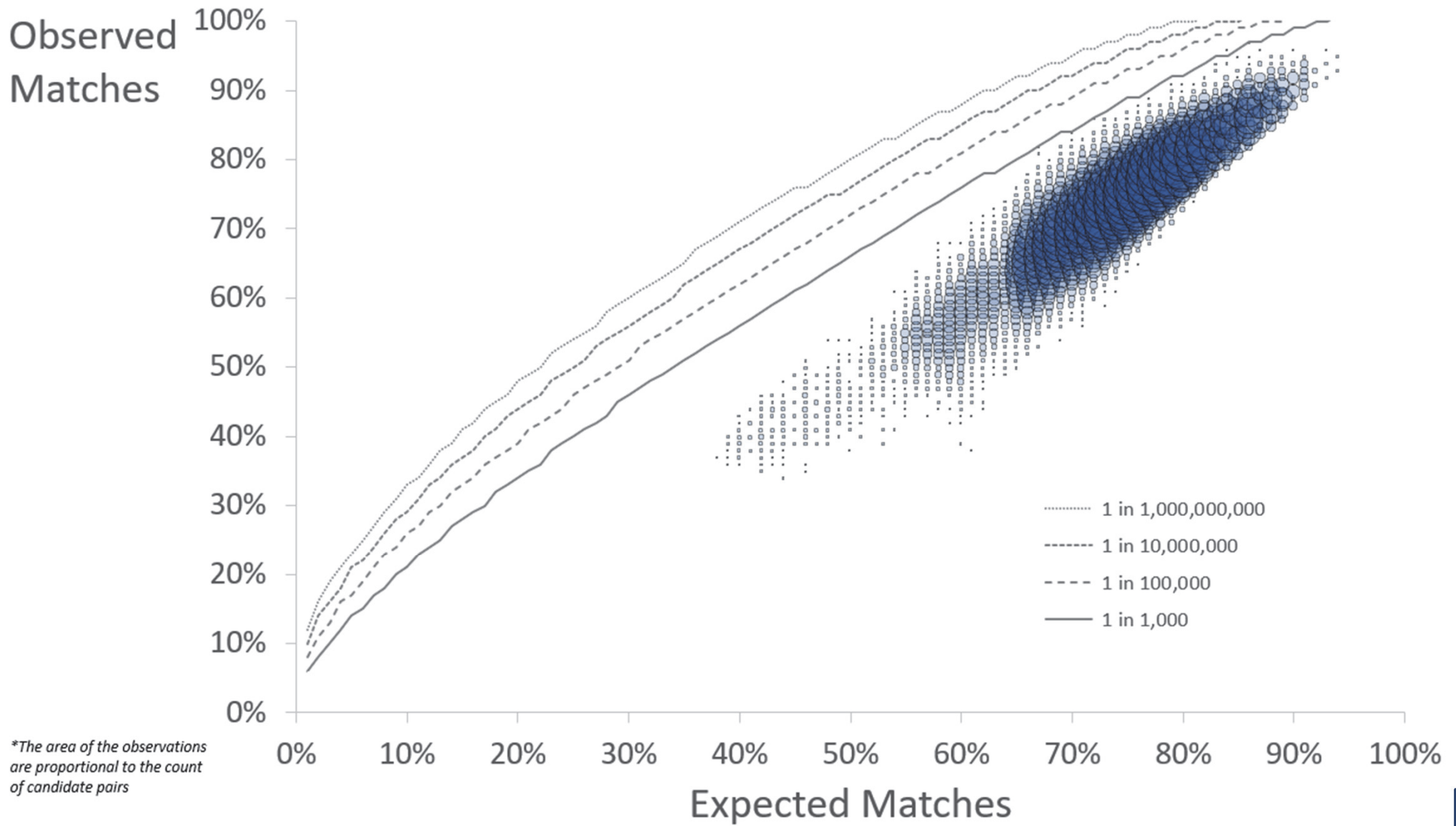
Height is proportional to the count of candidate pairs



Heat Map of Observed vs Expected Matches

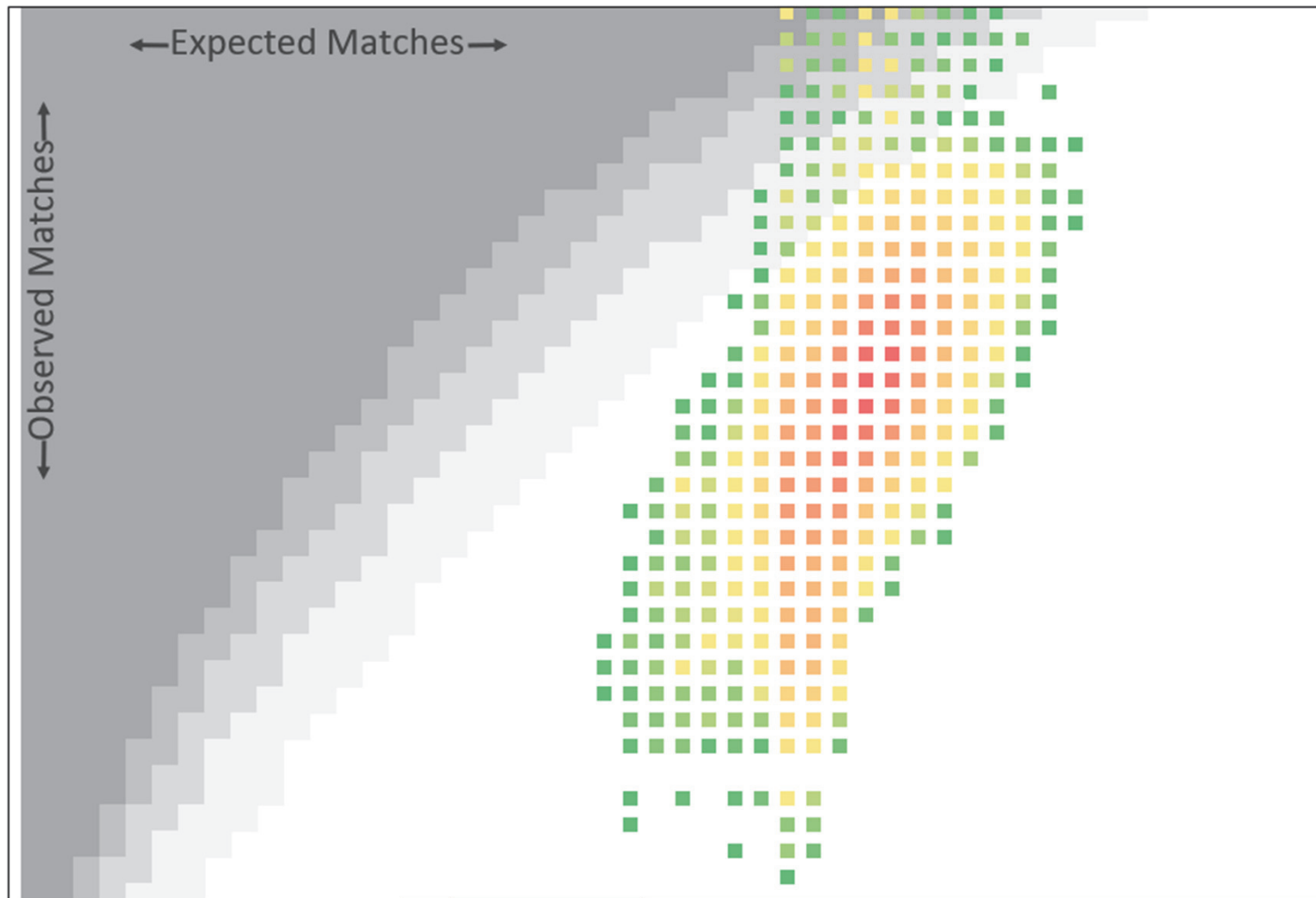
Red = more pairs of candidates Green = fewer pairs of candidates

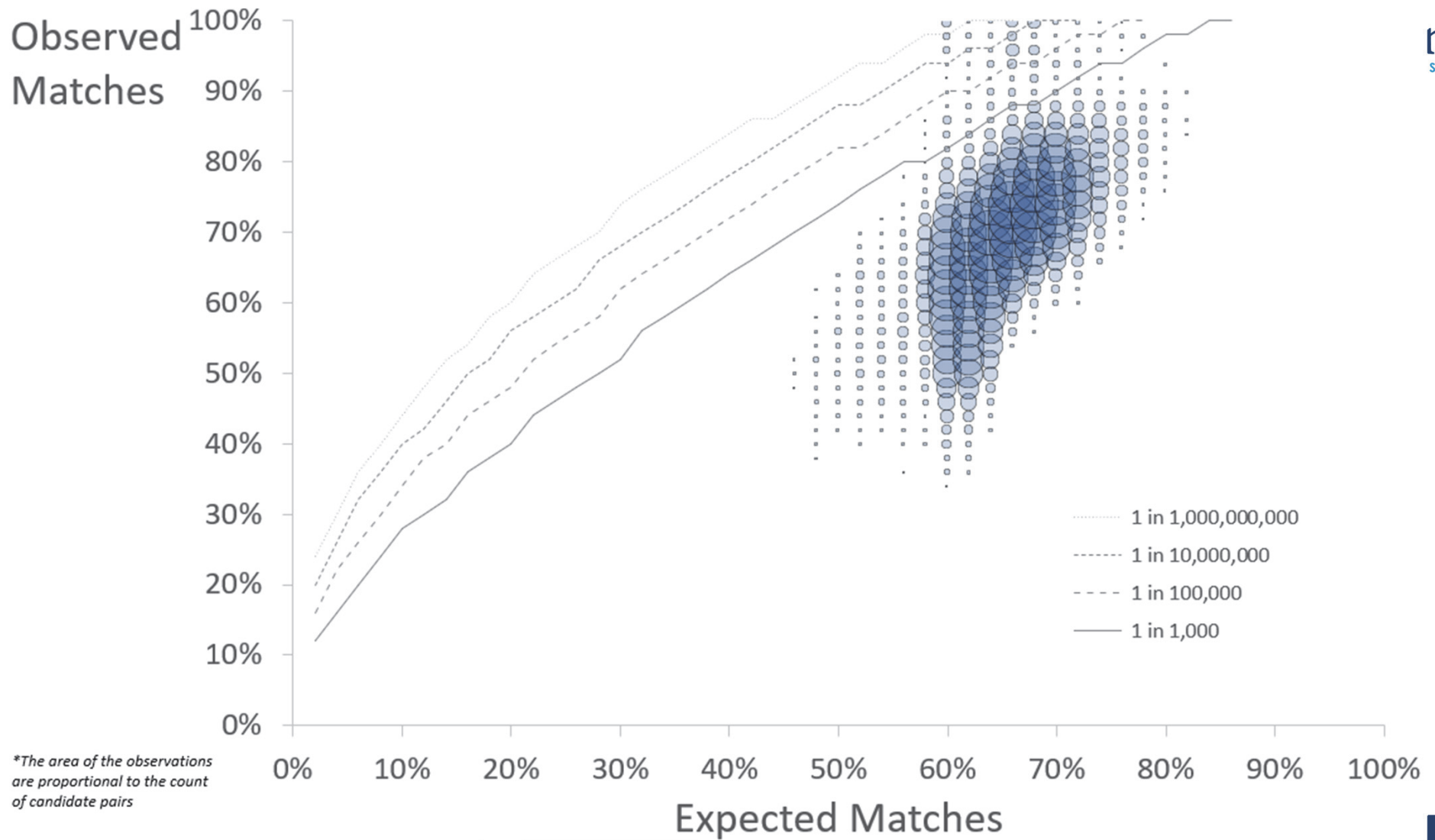




Heat Map of Observed vs Expected Matches

Red = more pairs of candidates Green = fewer pairs of candidates





Visualizing the Prevalence of Collusion/Content Exposure

- Score/Response similarity analyses are effective
- Focus should be on both:
 - Identifying aberrant candidates
 - Overall health
- Graphs shown here help understand health/extent of problems
 - Different versions work better, depending on exam/candidate counts
 - Scaling counts can help direct attention where needed

Questions?

brett.foley@alpinetesting.com

Foley, B. P. (2016). Visual displays of test fraud data. In G. J. Cizek & J. A. Wollack (Editors). *Handbook of quantitative methods for detecting cheating on tests* (pp. 323-345). New York, NY: Routledge.