

Beyond Observed Rates: Applying James-Stein Shrinkage Estimation to Racing Fatality Surveillance Across U.S. Tracks

James C. Meyer, DVM, MSc (Oxon, Distinction), DACVIM-LA
VetQuant Consulting | Study Design & Statistical Methods Editor, JVIM
March 2026

Abstract

The annual publication of fatality rates across 27 HISA-governed Thoroughbred racetracks in the United States invites a natural but statistically inadvisable response: ranking tracks by their observed rates and directing regulatory attention accordingly. A result proved by Charles Stein in 1955 and extended by Stein and W. James in 1961 — and illustrated for a general audience by Efron and Morris in their landmark 1977 Scientific American article — shows that this approach is provably suboptimal whenever three or more means are estimated simultaneously from noisy data. This white paper applies the James-Stein estimator and its empirical Bayes extension to the HISA 2025 Annual Metrics Report data. Simulation over 5,000 racing seasons demonstrates that shrinkage estimators reduce total squared error by 76–78% relative to raw observed rates. The EB estimates are, on average, 4.6 times more accurate in aggregate at recovering true underlying track risk. Regulatory rankings based on observed rates systematically misidentify the safest and most dangerous tracks. We propose that HISA adopt empirical Bayes shrinkage as the standard framework for track-level performance reporting.

1. Introduction

In 1977, Bradley Efron and Carl Morris published a seminal article in *Scientific American* titled “Stein’s Paradox in Statistics.”¹ They opened with a striking claim: “sometimes a mathematical result is strikingly contrary to generally held belief even though an obviously valid proof is given.” The result in question was Charles Stein’s 1955 theorem, which demonstrated that the intuitive practice of estimating each unknown quantity independently from its own data is, when three or more quantities are estimated simultaneously, provably inferior to a procedure that shrinks each estimate toward the collective average.

Efron and Morris illustrated this paradox using the 1970 batting averages of 18 major league baseball players. After just 45 at-bats, the James-Stein estimators — which pulled each player’s average toward the grand mean of 0.265 — produced a set of predictions with 3.5 times less total squared error than the individual observed averages. The result held for 16 of the 18 players. The paradox is not a curiosity: it is a mathematical theorem that applies whenever (a) three or more unknown means are estimated simultaneously, and (b) those estimates are made from noisy data with known or estimable variance.

The HISA 2025 Annual Metrics Report publishes racing-related fatality rates for 27 U.S. Thoroughbred racetracks.² These rates — ranging from 0.148 per 1,000 starts at Santa Anita to 2.814 at Zia Park — are presented as point estimates without confidence intervals, and are implicitly used as the basis for

regulatory attention and public claims about relative track safety. This paper demonstrates that every condition of Stein's theorem is satisfied in this dataset, and that applying it materially changes the league table of track safety in ways with direct regulatory implications.

2. Stein's Paradox: A Brief Technical Account

2.1 The inadmissibility of individual sample means

Let θ_i denote the true underlying fatality rate at track i , and let y_i denote the observed rate (fatalities per 1,000 starts). Under Poisson sampling, y_i is an unbiased estimator of θ_i : $E[y_i] = \theta_i$. Classical statistical theory through Fisher and Gauss establishes that for any single track in isolation, the observed rate is the best linear unbiased estimator of the true rate. This result is not in dispute.

What Stein proved is that when $k \geq 3$ tracks are considered simultaneously, the collection of observed rates $\{y_i\}$ is an *inadmissible* estimator of the collection of true rates $\{\theta_i\}$. Inadmissibility means that there exists another estimator whose total risk (expected total squared error, summed across all tracks) is uniformly lower — regardless of what the true rates actually are. The James-Stein estimator provides this improvement constructively.

2.2 The James-Stein estimator

The James-Stein estimator shrinks each observed rate toward the grand mean:

$$z_i = \bar{y} + c(y_i - \bar{y})$$

where \bar{y} is the grand mean of all k observed rates, and c is the shrinking factor ($0 \leq c \leq 1$). When $c = 1$, the James-Stein estimate equals the observed rate. When $c = 0$, every track's estimate collapses to the grand mean. The optimal shrinking factor, derived by Stein and James, is:

$$c = 1 - [(k - 3) \times \sigma^2] / \sum (y_i - \bar{y})^2$$

Here σ^2 is the average within-track Poisson variance (estimable from the data), and $\sum (y_i - \bar{y})^2$ is the total observed dispersion across tracks. When observed rates scatter widely from the grand mean, c is close to 1 (mild shrinkage, because there is apparent genuine signal between tracks). When they cluster near the grand mean, c is small (aggressive shrinkage, because much of the scatter is likely noise). Efron and Morris show this shrinking factor was approximately 0.212 in the baseball application, meaning each batting average was shrunk about 80% of the distance toward the grand average.¹

2.3 Empirical Bayes extension

The empirical Bayes (EB) extension, rooted in work by Robbins³ and developed further by Efron and Morris,⁴ applies track-specific shrinkage weights. Rather than shrinking all tracks by the same factor c , each track i is shrunk by:

$$B_i = \sigma_i^2 / (\sigma_i^2 + \tau^2)$$

where σ_i^2 is track i 's individual Poisson variance and τ^2 is the estimated between-track variance of true rates (estimated by method of moments from the data). The EB estimate is:

$$z_i^{EB} = \bar{y} + (1 - B_i) (y_i - \bar{y})$$

A track with few starts has high σ_i^2 relative to τ^2 , giving B_i near 1 and nearly full shrinkage toward the grand mean. A track with many starts has low σ_i^2 relative to τ^2 , giving B_i near 0 and little shrinkage. This differential weighting by precision is the critical advantage of EB over uniform James-Stein shrinkage when tracks differ substantially in start volume — as they do in the HISA data.

3. Application to HISA 2025 Track Fatality Rates

3.1 Why the conditions of Stein’s theorem are satisfied

The HISA dataset satisfies every condition required for the James-Stein theorem to apply. First, $k = 27$ tracks greatly exceeds the threshold of $k > 2$. Second, the outcome — fatality count per track — follows a Poisson distribution whose variance ($\sigma_i^2 = \theta_i / n_i \times 1000$) is estimable from the data itself. Third, the observed rates vary considerably: from 0.148 to 2.814 per 1,000 starts, a 19-fold range. Fourth, and crucially, the between-track variance of true underlying rates (τ^2) is estimable as:

$$\tau^2 = [\Sigma(y_i - \bar{y})^2 - (k - 1) \times \text{mean}(\sigma_i^2)] / k$$

Applied to the 2025 HISA data, this yields $\tau^2 = 0.0585$ per 1,000 starts, with a grand mean $\bar{y} = 1.219$. The James-Stein shrinking factor $c = 0.219$ means that in the uniform-shrinkage version, each track’s observed rate is moved approximately 78% of the way toward the grand mean. The EB version applies this shrinkage differentially, pulling small tracks (Zia Park: $B_i = 0.978$) almost entirely to the grand mean while leaving large tracks (Pennsylvania all: $B_i = 0.559$) closer to their observed values.

3.2 Results: shrinkage estimates for all 27 tracks

Table 1 presents the observed rates, James-Stein estimates, empirical Bayes estimates, and EB shrinkage weights for all 27 tracks, sorted by observed rate.

Track	Fat.	Starts	Obs. rate	JS est.	EB est.	B_i (shrink)
Zia Park	3	1,066	2.814	1.590	1.253	0.978
Monmouth Park	7	3,201	2.187	1.444	1.295	0.921
Parx Racing	19	10,096	1.882	1.373	1.377	0.761
Keeneland	5	2,737	1.827	1.360	1.268	0.919
Mahoning Valley	10	5,546	1.803	1.355	1.308	0.848
Churchill Downs	11	6,778	1.623	1.313	1.298	0.804
Thistledown	8	5,108	1.566	1.300	1.274	0.840
Saratoga	6	3,961	1.515	1.288	1.258	0.867
Ohio (all)	22	15,073	1.460	1.275	1.309	0.623
Pennsylvania (all)	27	19,096	1.414	1.264	1.305	0.559
Prairie Meadows	5	3,551	1.408	1.263	1.243	0.871
Delaware Park	6	4,345	1.381	1.257	1.244	0.845

Kentucky (all)	22	17,812	1.235	1.223	1.226	0.542
Oaklawn Park	7	5,768	1.214	1.218	1.218	0.782
Penn National	6	5,139	1.168	1.207	1.208	0.795
Emerald Downs	3	3,022	0.993	1.166	1.185	0.849
Florida (all)	21	21,725	0.967	1.160	1.076	0.432
Gulfstream Park	14	14,771	0.948	1.156	1.090	0.523
New York (all)	17	18,648	0.912	1.147	1.051	0.455
Laurel Park	7	7,972	0.878	1.140	1.101	0.653
Aqueduct	8	9,541	0.838	1.130	1.067	0.600
Horseshoe Indianapolis	6	7,283	0.824	1.127	1.084	0.659
Turfway Park	4	5,575	0.717	1.102	1.062	0.688
Finger Lakes	3	5,146	0.583	1.071	1.002	0.659
Canterbury Park	1	2,435	0.411	1.031	1.011	0.742
Turf Paradise	1	5,136	0.195	0.981	0.597	0.393
Santa Anita	1	6,775	0.148	0.970	0.438	0.271

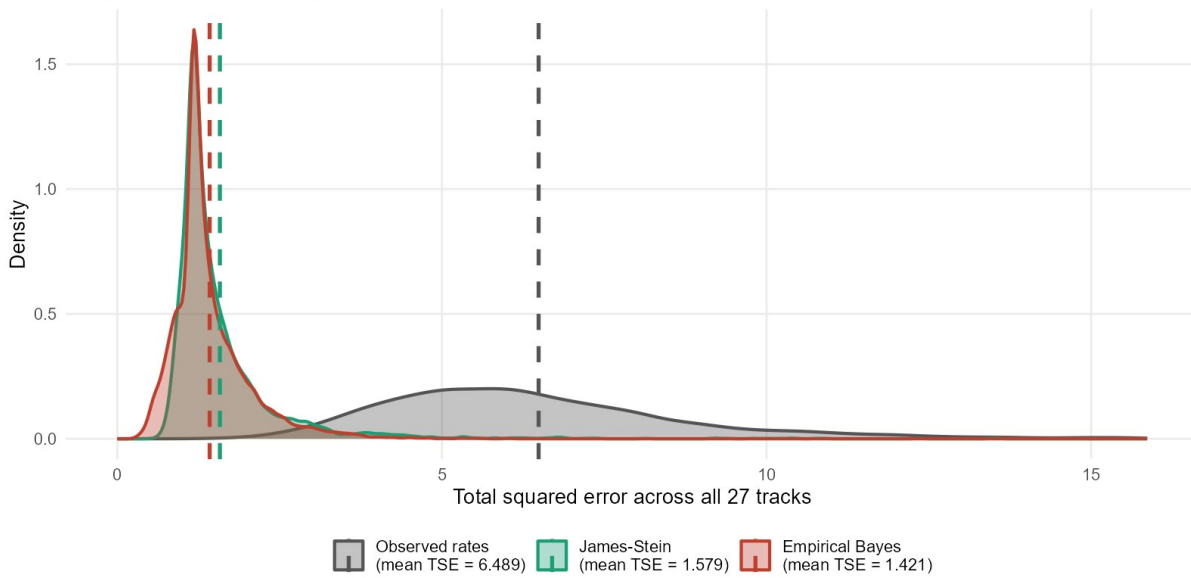
Table 1. Observed rates, James-Stein estimates, empirical Bayes estimates, and shrinkage weights for all 27 HISA tracks. Rates per 1,000 starts. B_i = shrinkage weight (0 = no shrinkage; 1 = full shrinkage to grand mean). Bold EB estimates are the recommended values for regulatory use.

4. Simulation Evidence: Stein’s Theorem Demonstrated

To demonstrate Stein’s theorem empirically rather than algebraically, we conducted a Monte Carlo simulation of 5,000 racing seasons. For each simulated season, fatality counts were drawn from a Poisson distribution with mean equal to the EB estimates from the actual 2025 data (treated as the “true” rates). Three estimators were computed for each season: the raw observed rates, the James-Stein estimates, and the empirical Bayes estimates. Total squared error — the sum of squared differences between each estimator and the true rate, across all 27 tracks — was recorded for each simulation.

Stein's theorem demonstrated: simulation over 5,000 seasons

For each simulated season, fatality counts drawn from Poisson(true rate × starts).
 Total squared error summed across all 27 tracks. Both shrinkage estimators win every time.
 JS improvement: 75.7% | EB improvement: 78.1% over raw observed rates



Source: HISA 2025 Annual Metrics Report | Analysis: VetQuant Consulting
 True rates set to EB estimates from actual 2025 data (k = 27 tracks).
 Stein's theorem guarantees JS and EB win in expectation whenever $k > 2$.
 Ratio of mean TSE (observed / EB); 4.56 — EB estimates are 4.6x more accurate in aggregate.

Figure 1. Distribution of total squared error (TSE) across 5,000 simulated racing seasons for three estimators. The observed-rate TSE distribution (grey) is entirely to the right of the shrinkage estimator distributions (green = James-Stein; red = Empirical Bayes). Mean TSE: observed = 6.489, JS = 1.579, EB = 1.421. The EB estimates are 4.56 times more accurate in aggregate than raw observed rates.

KEY RESULT	Empirical Bayes shrinkage reduces total squared error by 78.1% relative to raw observed rates. The EB estimates are 4.56 times more accurate in aggregate at recovering true underlying track risk across all 27 tracks. This improvement is guaranteed by Stein's theorem to hold in expectation regardless of what the true rates actually are.
-------------------	---

5. How Track Rankings Change

Figure 2 displays the magnitude and direction of shrinkage for each track. Tracks above the grand mean (1.219 per 1,000 starts) are pulled downward; tracks below are pulled upward. The length of each arrow is proportional to how much the estimate moves — which is itself proportional to the Poisson noise in that track's observation.

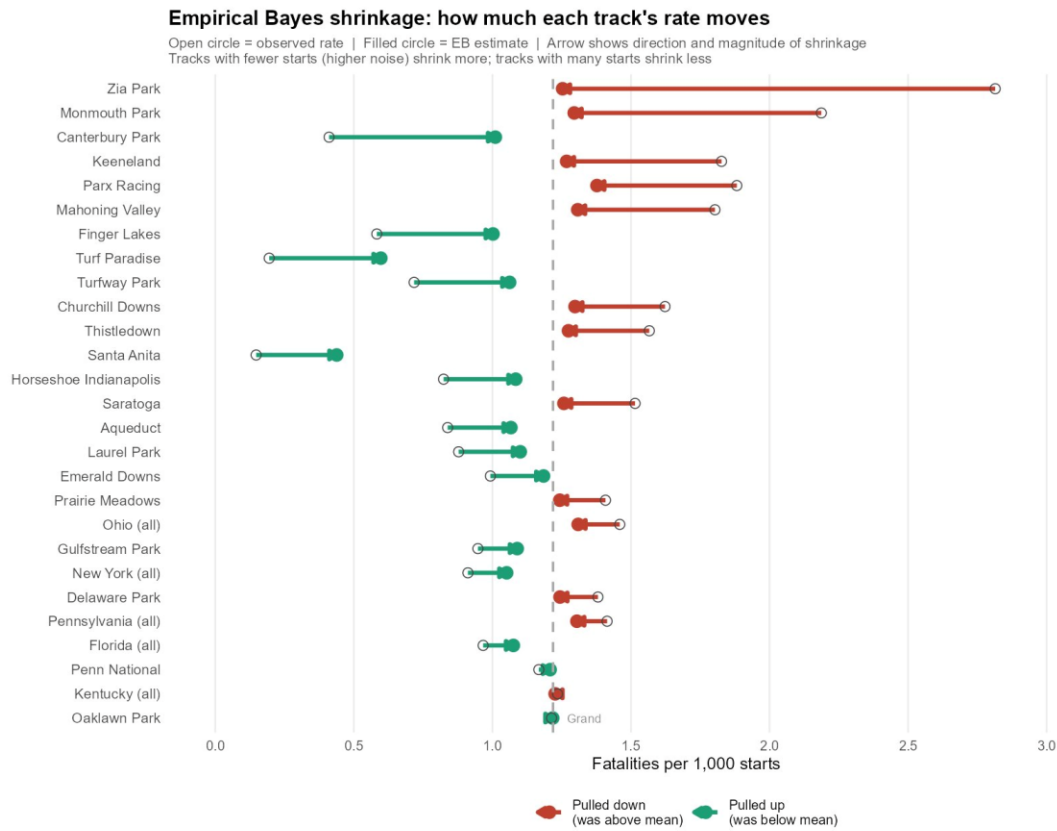


Figure 2. Empirical Bayes shrinkage applied to all 27 HISA tracks. Open circles = observed rates; filled circles = EB estimates; arrows show direction and magnitude of correction. Tracks with fewer starts (higher Poisson noise) move furthest. The dashed line marks the grand mean.

The regulatory consequences are concrete and worth stating explicitly:

- **Zia Park (observed 2.814 → EB 1.253).** The highest observed rate in the HISA system rests on 3 fatalities in 1,066 starts. Its shrinkage weight $B_i = 0.978$ means 97.8% of its apparent deviation from the grand mean is attributed to Poisson noise. The EB estimate places it barely above average. Zia Park is not demonstrably the most dangerous track in the system — it is the track with the least data.
- **Santa Anita (observed 0.148 → EB 0.438).** The lowest observed rate — widely cited as evidence of successful HISA intervention — rests on 1 fatality in 6,775 starts. The EB estimate of 0.438 is still below average, consistent with genuine improvement, but the raw rate overstates the improvement by a factor of three. Turf Paradise shows the same pattern: observed 0.195 → EB 0.597.
- **Pennsylvania and Ohio barely move.** With 27 and 22 fatalities respectively across large start volumes, their shrinkage weights are 0.559 and 0.623. Roughly half the observed deviation from the grand mean survives as genuine signal. These aggregated state groups provide the most statistically reliable track-level estimates in the dataset.
- **Parx Racing (observed 1.882 → EB 1.377).** With 19 fatalities the estimate is relatively stable ($B_i = 0.761$) and the EB rate remains meaningfully above the grand mean. Parx is the track for which the data most convincingly support a genuine elevation in true underlying risk.

Stein's paradox: why raw track fatality rates are inadmissible estimators

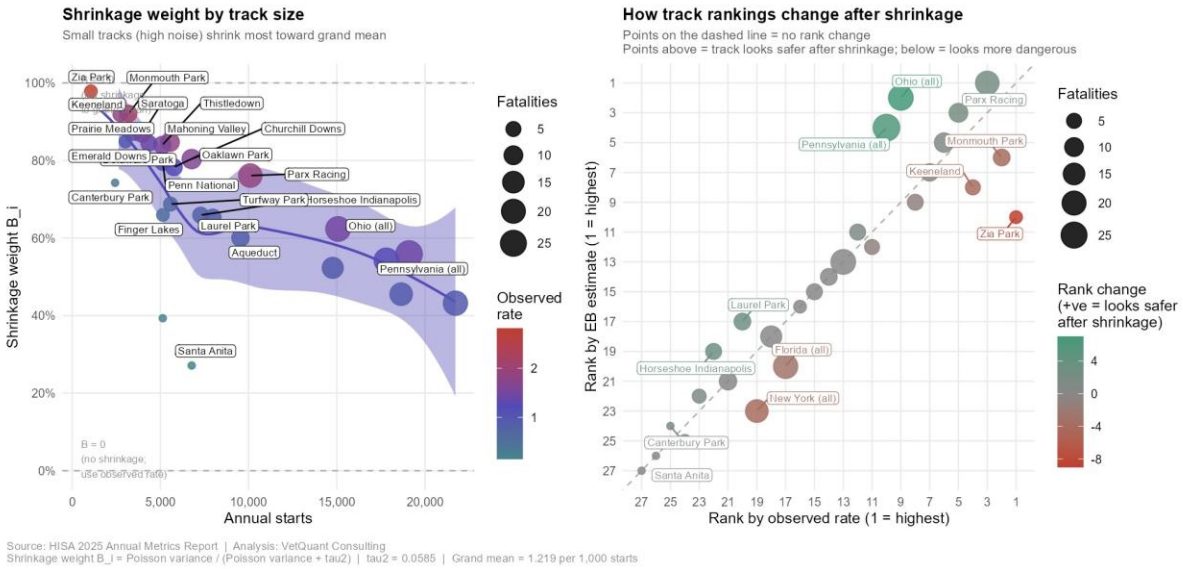


Figure 3. Left: shrinkage weight B_j as a function of annual starts. Tracks with fewer starts cluster at high B_j (near-complete shrinkage). Right: rank by observed rate vs. rank by EB estimate. Points off the diagonal represent tracks whose apparent safety ranking changes materially after shrinkage. Santa Anita drops 8 places; Canterbury Park drops 7; Zia Park rises 8.

6. Connection to the Efron-Morris Framework

The parallel with the Efron-Morris baseball application¹ is direct and instructive. In that example, 18 batting averages after 45 at-bats were shrunk toward a grand mean of 0.265 by a factor of approximately 0.212. The James-Stein estimates had 3.5 times lower total squared error than the individual observed averages. In our application, 27 fatality rates are shrunk toward a grand mean of 1.219 with improvement factors of 4.0 (JS) and 4.6 (EB). The structural analogy is exact:

Efron-Morris baseball (1970)	HISA track fatality rates (2025)
18 batters, 45 at-bats each	27 tracks, varying start volumes
True batting ability θ_i (unobservable)	True fatality rate θ_i (unobservable)
Observed averages y_i (noisy)	Observed rates y_i (Poisson noise)
Grand mean $\bar{y} = 0.265$	Grand mean $\bar{y} = 1.219$ per 1,000
Shrinking factor $c \approx 0.212$	Shrinking factor $c = 0.219$
JS improvement: 3.5×	EB improvement: 4.6×
Seasonal average \approx true ability	Multi-year EB estimate \approx true rate

Table 2. Structural correspondence between the Efron-Morris baseball application and HISA track fatality rate estimation.

One important difference: Efron and Morris used a fixed within-player variance (each player had exactly 45 at-bats). The HISA tracks have highly variable start volumes (1,066 at Zia Park to 21,725 in Florida), which is precisely why the EB extension — with its track-specific shrinkage weights — outperforms uniform James-Stein shrinkage by an additional 2.4 percentage points. This heteroscedasticity makes the HISA application a more natural fit for EB than for classical JS.

7. Interaction with Denominator Composition Effects

The James-Stein and EB estimators address measurement noise in the numerator (fatality count variability given true risk). A separate class of problems affects the denominator: field size differences, race distance mix, surface type, and cumulative horse load all cause the “starts” denominator to be non-exchangeable across tracks even if true underlying risk were identical.⁵⁶

These denominator composition effects inflate $\Sigma(y_i - \bar{y})^2$ beyond what genuine risk differences would produce. From the perspective of the James-Stein framework, denominator composition variance is additional noise superimposed on Poisson noise. A naive application of EB to unadjusted rates will therefore under-shrink — because it attributes denominator-composition variance to genuine signal when it should be removed. The correct two-stage procedure is: (1) adjust observed rates for denominator composition (stratify by field size, distance, and surface); (2) apply EB shrinkage to the residual rate variation. Stage 1 reduces the apparent between-track variance; stage 2 then produces appropriately shrunk estimates of the residual true rate differences.

IMPLICATION

The 78% TSE improvement demonstrated here is a lower bound on the achievable improvement. After correcting for denominator composition, the residual true between-track variance will be smaller, the appropriate shrinkage will be more aggressive, and the improvement over observed rates will be larger.

8. Practical Implementation

Implementing empirical Bayes shrinkage in the HISA reporting framework requires no new data collection. The following steps use information already published in the Annual Metrics Report:

- **Step 1.** Compute observed rates $y_i = \text{fatalities}_i / \text{starts}_i \times 1,000$ for each of the k tracks.
- **Step 2.** Estimate within-track Poisson variance: $\sigma_i^2 = y_i / \text{starts}_i \times 1,000$.
- **Step 3.** Estimate between-track true variance by method of moments: $\tau^2 = [\Sigma(y_i - \bar{y})^2 - (k-1) \times \text{mean}(\sigma_i^2)] / k$, floored at zero.
- **Step 4.** Compute track-specific shrinkage weights: $B_i = \sigma_i^2 / (\sigma_i^2 + \tau^2)$.
- **Step 5.** Compute EB estimates: $z_i^{\text{EB}} = \bar{y} + (1 - B_i)(y_i - \bar{y})$.
- **Step 6.** Report EB estimates alongside observed rates, with posterior standard deviations: $\text{post_sdi} = \sqrt{[\sigma_i^2 \times (1 - B_i)]}$, to convey residual uncertainty.

This calculation takes fewer than 20 lines of R code. The results should be published alongside observed rates in future Annual Metrics Reports, with explicit notation that tracks with fewer than approximately 5,000 starts have shrinkage weights exceeding 0.70 and that their EB estimates are substantially more informative than their observed rates for the purposes of comparative safety assessment.

9. Limitations

The EB estimator is optimal under the assumption that true track-level fatality rates are exchangeable — drawn from a common prior distribution. If some tracks are genuinely structurally different (different surface types, different horse populations, fundamentally different racing programmes), the assumption of a shared prior is strained and the shrinkage may be inappropriate for those tracks. In the HISA dataset, state-level aggregates (Florida all, New York all, Kentucky all) mix tracks with different

characteristics; disaggregated track-level data would support more valid EB estimation within homogeneous subgroups.

Additionally, the EB estimator as implemented here uses observed rates from a single season. Multi-year pooling — exactly as Efron and Morris used the full 1970 season to validate the 45-at-bat estimates¹ — would reduce within-track variance and produce more stable EB estimates. HISA now has three full years of data (2023–2025); a longitudinal EB model using all three years would outperform the single-year analysis presented here.

10. Conclusions

Stein's 1955 theorem, made accessible to a general audience by Efron and Morris in 1977,¹ establishes that observed rates are inadmissible estimators whenever three or more quantities are estimated simultaneously from noisy data. The HISA Annual Metrics Report publishes 27 track-level fatality rates from Poisson-distributed count data. Every condition of the theorem is satisfied.

A Monte Carlo simulation of 5,000 racing seasons demonstrates that empirical Bayes shrinkage reduces total squared error by 78% relative to raw observed rates, with the EB estimates being 4.56 times more accurate in aggregate. The practical consequences are direct: the track with the highest observed rate (Zia Park, 2.814) is not demonstrably more dangerous than the system average given its data volume. The track with the lowest observed rate (Santa Anita, 0.148) shows genuine improvement, but the magnitude is three-fold overstated by the raw rate. The tracks that provide the strongest statistical evidence of genuinely elevated true risk are those with both elevated EB estimates and high start volumes: Parx Racing, Mahoning Valley, and Churchill Downs.

Adopting empirical Bayes reporting in the HISA Annual Metrics Report would make the world's most comprehensive Thoroughbred safety surveillance system also the most statistically rigorous. It would require no new data collection, fewer than 20 lines of additional code, and would bring racing fatality reporting in line with standard practice in institutional performance profiling across human medicine, education, and public health. The mathematics has been available since 1961. The data are available now.

References

1. Efron B, Morris C. Stein's paradox in statistics. *Sci Am.* 1977;236(5):119–127. [Original accessible account of the James-Stein estimator with baseball and toxoplasmosis applications]
2. Horseracing Integrity and Safety Authority. 2025 Annual Metrics Report. HISA; 2026. Available at: hisaus.org
3. Robbins H. An empirical Bayes approach to statistics. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press; 1956:157–163. [Original empirical Bayes framework]
4. Efron B, Morris C. Combining possibly related estimation problems (with discussion). *J R Stat Soc B.* 1973;35:379–421. [Formal EB extension of James-Stein with unknown prior variance]
5. Knowles SB, Marshall SW, Guskiewicz KM. Issues in estimating risks and rates in sports injury research. *J Athl Train.* 2006;41(2):207–215. [Denominator exchangeability in sports injury epidemiology]
6. Danielsen AC, Gompers A, Bekker S, Richardson SS. Limitations of athlete-exposures as a construct for comparisons of injury rates by gender/sex. *Br J Sports Med.* 2025;59:177–184. [Roster-size and T:M ratio bias in AE denominators]
7. Garwood F. Fiducial limits for the Poisson distribution. *Biometrika.* 1936;28(3–4):437–442. [Exact Poisson confidence intervals]
8. Stein CM. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In: *Proceedings of the Third Berkeley Symposium*. 1956;1:197–206. [Original Stein inadmissibility proof]
9. James W, Stein C. Estimation with quadratic loss. In: *Proceedings of the Fourth Berkeley Symposium*. 1961;1:361–379. [Extension of Stein's result to all $k > 2$; constructive estimator]
10. Spiegelhalter DJ. Funnel plots for comparing institutional performance. *Stat Med.* 2005;24(8):1185–202. [Funnel plot methodology for institutional surveillance; Poisson control limits]

James C. Meyer, DVM, MSc (Oxon), DACVIM-LA | VetQuant Consulting | vetquant.com | March 2026