XC-SPI Formula & Statistics Guide

Version 2.7 plain-language guide to the equations, terms, and interpretation rules

Colin C. Caso

Oklahoma State University

August 28, 2026

Abstract

This guide explains the equations used by XC-SPI for readers without formal statistical training. It distinguishes four things that are easy to blur together: the rule-based Week 0 Preseason Score, the Postmodern Era Historical Success Model, the future fitted athlete-performance model, and the arithmetic used to translate a defensible probability into betting odds. The August 2026 model-design audit tightened several equations without changing the frozen Week 0 rankings. In particular, Preseason Score is no longer described as expected NCAA points, the depth term is explicitly a fragility measure, athlete uncertainty grows with elapsed time, context uncertainty is carried into race updates, difficult conditions may widen outcome variance, team results can be correlated through a team-day shock, and the former Preseason Score-to-lognormal probability bridge is retired.

1. Basic Notation

A subscript identifies an object: i athlete, j team, r race, t time, m simulation. A hat means an estimate. A tilde means a simulated draw. $N(a, b)$ denotes a normal distribution with mean a and variance b. $I(\cdot)$ is 1 when a statement is true and 0 otherwise. $\sum$ means add.

Table 1: Core vocabulary.

Term	Plain-language meaning
Index	A constructed score used to order or summarize cases. An index need not have a natural zero or ratio interpretation.
Probability	Chance under a specified model, between 0 and 1.
Expected value	A probability-weighted long-run mean. This term is reserved for quantities that are actually probabilistic expectations.
Variance	Squared uncertainty scale around a mean.
Standard deviation	Square root of variance.
Simulation	One generated possible version of an event.
Calibration	Whether repeated stated probabilities agree with observed frequencies.

2. NCAA Scoring Comes First

XC-SPI keeps raw overall finish P_i and NCAA team-scoring place Q_i separate. Individual qualifiers and runners from incomplete teams can affect the raw finish order without receiving team-scoring places. Five runners score; runners six and seven may displace.[1]

Plain language. The simulator must reproduce NCAA scoring before any model ranking can claim to represent championship team performance.

Method status. This is rule implementation, not a statistical invention.

3. Week 0: Evidence Confidence, Not Probability

X_i is an athlete's comparable prior scoring value. c_i is a declared Confidence weight. The launch ledger uses 0.99 for verified cases, 0.85 for probable cases, 0.70 for materially unresolved cases, and 0 for verified unavailable cases.

Do not read this as: $c_i = .85$ does not mean an 85% chance that the athlete runs at NCAAs.

3.1. Adjusted Runner Score

$$C_i = c_i X_i + (1 - c_i)R_0. \tag{1}$$

Plain language. Pull the historical value toward a conservative replacement baseline when current evidence is weaker. If $X_i = 20$, $c_i = .85$, and $R_0 = 200$, then $C_i = .85(20) + .15(200) = 47$.

Do not read this as: $C_i = 47$ is not a predicted finish and is not a mathematical expectation because c_i is not a calibrated probability.

Method status. The exact formula is an XC-SPI index construction. The general logic resembles shrinkage/partial pooling, but the launch constants are declared and sensitivity-tested rather than statistically estimated.[5]

3.2. A useful edge case

If $X_i > R_0$, uncertainty can improve the adjusted index value because the formula pulls the historical value toward replacement. For example, $X_i = 250$ and $c_i = .85$ gives $C_i = 242.5$. This is not an error: it reveals that C_i is a shrinkage index combining current-evidence confidence with a replacement baseline, not a literal participation expectation.

3.3. Positive-part operator and Depth Risk

$$(x)_+ = \max(x, 0). \tag{2}$$

Order candidate contributions $C_{j(1)} \leq \cdots \leq C_{j(7)}$ and define

$$F_j = \lambda[(C_{j(6)} - C_{j(5)})_+ + (C_{j(7)} - C_{j(5)})_+]. \tag{3}$$

Plain language. F_j gets larger when the sixth and seventh options are far behind the fifth scorer.

Do not read this as: This is not actual NCAA displacement scoring. Real displacement depends on the full field and is handled in the championship simulation.

Method status. This is an XC-SPI roster-fragility heuristic; launch $\lambda = .05$ is sensitivity-tested.

3.4. Preseason Score

$$PS_j = \sum_{k=1}^{5} C_{j(k)} + F_j. \tag{4}$$

Plain language. Add the five strongest adjusted contributions, then add a small fragility penalty. Lower is better.

Do not read this as: Preseason Score is not predicted NCAA points, not title probability, and not a ratio scale. A Preseason Score of 300 is not "twice as bad" as 150.

4. Preseason Athlete Distribution

$$\alpha_{i,0} \sim N(\theta_{0,s} + \theta_{1,s} XC_{i,2025} + \theta_{2,s} TRK_{i,2026}, V_{i,0}). \tag{5}$$

Plain language. Before a current XC race, estimate athlete ability from prior XC plus standardized track evidence while carrying uncertainty. Men's and women's parameters are fit separately or explicitly stratified.

Method status. Hierarchical/regression priors are standard; the predictors are XC-SPI choices to be estimated and backtested.[5]

5. Ability Changes with Time

$$\alpha_{i,t}^{-} = \alpha_{i,t-1} + \eta_{i,t}, \tag{6}$$
$$\eta_{i,t} \sim N(0, \sigma_\eta^2 \Delta t), \tag{7}$$
$$V_{i,t}^{-} = V_{i,t-1} + \sigma_\eta^2 \Delta t. \tag{8}$$

Plain language. The longer it has been since we saw the athlete, the more uncertain current fitness becomes. Five weeks without evidence should not be treated like five days.

Method status. This is a state-space/random-walk idea with elapsed-time scaling.[7]

6. Observed Race Time

$$\log T_{ir} = \mu_s + \alpha_{i,t}^{-} + \delta_s(d) + \gamma_{c,s} + h_s(w) + g_s(q) + u_r + v_{jr} + \epsilon_{ir}. \tag{9}$$

Plain language. Observed logged time is athlete ability plus distance, course, weather/ground context, a common race effect, a shared team-day effect, and individual noise.

Method status. Additive transformed-response modeling is standard; this decomposition is the XC-SPI specification.

7. Why Weather Does Not Automatically Reorder a Future Race

If every athlete in the same championship receives the same additive course/weather term, subtracting Runner B's equation from Runner A's cancels those common terms. Therefore an additive "mud penalty" changes predicted absolute time but not by itself who beats whom.

XC-SPI v2.7 handles this by allowing difficult conditions to alter residual spread:

$$\log \sigma_{\epsilon,r} = \xi_0 + \xi^\top z_r. \tag{10}$$

Plain language. Mud, heat, wind, or technical conditions can make the race more variable and therefore increase upset probability, even when there is not enough evidence to say which specific athlete is uniquely good in mud.

Do not read this as: Athlete-specific condition effects should not be added just because they sound plausible; repeated historical evidence must support them.

8. Team-by-Race Shared Shock

$$v_{jr} \sim N(0, \sigma^2_{team}). \tag{11}$$

Plain language. Runners from the same team can move together on a particularly good or bad team day. This captures modest correlation that independent athlete errors would miss.

Method status. A shared random effect is standard multilevel modeling; its use for team-day execution is an XC-SPI application.[5]

9. Distance and Course

$$\delta_s(d) = \sum_b \kappa_{b,s} B_b(\log d), \qquad \sum_r \omega_r \delta_s(d_r) = 0. \tag{12}$$

Plain language. Fit a smooth distance relationship rather than assuming one universal pace conversion.[6]

$$\gamma_{k,s} \sim N(x_k^\top \beta_{c,s}, \tau^2_{c,s}). \tag{13}$$

Plain language. A sparsely observed course is pulled toward what measurable course features predict; repeated evidence can move it away from that baseline.

10. How a Race Updates an Athlete

First remove estimated context to create z_{ir}. Then define observation variance

$$R_{ir} = \sigma^2_\epsilon + V_{course} + V_{weather} + V_{ground} + V_{measurement}. \tag{14}$$

The gain is

$$K_{ir} = \frac{V^-_{i,t}}{V^-_{i,t} + R_{ir}}. \tag{15}$$

Then

$$\hat{\alpha}_{i,t} = \hat{\alpha}^-_{i,t} + K_{ir}(z_{ir} - \hat{\alpha}^-_{i,t}), \tag{16}$$

$$V_{i,t} = (1 - K_{ir})V^-_{i,t}. \tag{17}$$

Plain language. A race with uncertain course/weather measurement receives less effective weight because its observation variance is larger.

Method status. These are Kalman-style explanatory equations; production may fit the equivalent joint hierarchical model.[7]

11. Why Bradley–Terry Is Not in the Production Model

A separate paired-comparison likelihood can model who beats whom, but the same-race clock already contains that ordering. Adding both without a clear validation gain risks double-counting. XC-SPI v2.7 therefore removes the Bradley–Terry layer from production. It can return only if preregistered rolling validation shows unique predictive information beyond normalized times.

12. Availability and Selection

$$p_i^{avail} = P(A_i = 1 \mid \mathcal{E}_t), \tag{18}$$
$$p_i^{select} = P(L_i = 1 \mid A_i = 1, \mathcal{E}_t). \tag{19}$$

Plain language. Being healthy/eligible enough to race and being chosen for the championship seven are different questions.

Do not read this as: Week 0 Confidence weights are not substitutes for these probabilities. Until the data support them, XC-SPI uses transparent roster gates and rule-based selection.

13. Championship Simulation and NCAA Score

For simulation m,

$$S_j^{(m)} = \sum_{k=1}^{5} Q_{jk}^{(m)}. \tag{20}$$

Plain language. Generate athlete performances, apply the actual NCAA scoring filter, and sum the five scorers. Important outputs include

$$\bar{S}_j = M^{-1} \sum_m S_j^{(m)}, \tag{21}$$
$$\widetilde{S}_j = \text{median}_m \, S_j^{(m)}, \tag{22}$$
$$\widehat{P}_j(title) = M^{-1} \sum_m I(teamplace_j^{(m)} = 1). \tag{23}$$

Do not read this as: Lowest expected score and highest title probability are related but not mathematically identical. A volatile team can have a better mean score but lose head-to-head more often than a stable team.

Method status. The primary power ranking uses expected simulated score; title, podium, median score, and expected place are separate outputs.

14. Monte Carlo Precision

$$\text{MCSE}(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{M}}. \tag{24}$$

Plain language. MCSE is only computer-simulation noise. More draws reduce MCSE but do not fix a bad model.

Method status. XC-SPI uses adaptive stopping: common probabilities need small absolute MCSE, while longshots also need acceptable relative precision/minimum event counts.[2]

15. Probability Calibration

For binary event forecasts, the Brier score is

$$BS = \frac{1}{N} \sum_{n=1}^{N} (p_n - y_n)^2. \tag{25}$$

Plain language. A model that repeatedly says 20% should see those events happen about 20% of the time across comparable cases.

Method status. Brier score and proper scoring/calibration literature provide the evaluation framework.[3;8]

16. Why One Championship Is Not Enough

One held-out men's champion and one held-out women's champion cannot establish calibration across probability ranges. Race-level components can be developed on recent seasons, but national-title calibration needs a longer historical championship archive, ideally roughly a decade or more where clean reconstruction is feasible.

17. Postmodern Era Historical Success

The public historical layer uses a simple name: **Historical Success Model**. It asks what happened one season later to Postmodern Era teams with a similar NCAA scoring profile. XC-SPI defines the Postmodern Era (Super-Shoe Era onward) as 2020 forward.[9]

Each profile uses four plain-language inputs: Team Score, Front Runner, Fifth Runner, and Depth Gap. The model has five possible next-season outcome categories. In formal notation,

$$P(Y = k \mid \mathbf{z}) = \frac{\exp(\alpha_k + \boldsymbol{\beta}_k^T \mathbf{z})}{\sum_{\ell=1}^{5} \exp(\alpha_\ell + \boldsymbol{\beta}_\ell^T \mathbf{z})}. \tag{26}$$

Read it aloud: the model compares the team's four standardized profile measures with patterns in the Postmodern Era and assigns probabilities to champion, 2nd–4th, 5th–10th, 11th–20th, and other.

The directly reported quantities are Historical Title Rate, Historical Podium Rate, Historical Top-10 Rate, and Historical Top-20 Rate. Historical Success Rank is ordered by podium rate because title is a much rarer event.

17.1. Historical Title Odds

Raw Historical Title Rates are estimated independently for each profile, so they do not form a mutually exclusive current-field market by themselves. For the public Top-25-conditional historical title board,

$$p_j^H = \frac{r_j}{\sum_{\ell=1}^{25} r_\ell}. \tag{27}$$

Read it aloud: take each team's historical title tendency and divide it by the total across the displayed Top 25. The shares now sum to 100%. This creates Historical Title Odds for the displayed field; it does not turn the historical benchmark into the full current-athlete XC-SPI probability model.

Podium and top-10 rates are different: several teams can satisfy those events in the same championship, so those direct historical rates are not normalized across teams.

18. The Retired Preseason Odds Bridge

Previous releases temporarily drew team performance from a lognormal distribution with mean equal to Preseason Score. That construction has been retired because it makes probabilities depend on Preseason Score ratios, which gives the index a natural zero/ratio interpretation it does not possess. The old scenario is kept internally only for provenance.

19. Fair Odds Once a Defensible Probability Exists

$$d = 1/p, \tag{28}$$

$$O_{US} = \begin{cases} +100(1-p)/p, & p < .5, \\ -100p/(1-p), & p \geq .5, \end{cases} \tag{29}$$

$$EV = pd - 1. \tag{30}$$

Market-implied probabilities and vig removal are ordinary arithmetic. Kelly is

$$f^* = \frac{bp - (1-p)}{b}, \qquad b = d - 1. \tag{31}$$

Do not read this as: These formulas translate a probability; they do not validate it. XC-SPI does not publish staking recommendations.[4]

20. Method Provenance Summary

Component	Status	Source/interpretation
NCAA scoring	Governing rule	NCAA rules.[1]
Preseason Score / Depth Risk	XC-SPI construction	Declared constants, sensitivity-tested; not expected points.
State-space athlete evolution	Established framework adapted	Durbin/Koopman.[7]
Hierarchical course/team effects	Established framework adapted	Gelman/Hill.[5]
Penalized distance/weather functions	Established framework adapted	Wood.[6]
Monte Carlo	Established framework	Metropolis/Ulam.[2]
Brier/proper scoring	Established framework	Brier; Gneiting/Raftery.[3;8]
Kelly	Established criterion, diagnostic only	Kelly.[4]
Historical Success model	XC-SPI historical benchmark	Postmodern Era profile-persistence model; separate from current athlete simulation.

References

[1] National Collegiate Athletic Association. Cross Country and Track and Field Playing Rules. Accessed August 28, 2026. https://www.ncaa.org/championships/playing-rules/cross-country-track-and-field-playing-rules/
[2] Metropolis N, Ulam S. The Monte Carlo method. *J Am Stat Assoc.* 1949;44(247):335–341.
[3] Brier GW. Verification of forecasts expressed in terms of probability. *Mon Weather Rev.* 1950;78(1):1–3.
[4] Kelly JL Jr. A new interpretation of information rate. *Bell Syst Tech J.* 1956;35(4):917–926.
[5] Gelman A, Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press; 2007.
[6] Wood SN. *Generalized Additive Models: An Introduction With R*. 2nd ed. Chapman & Hall/CRC; 2017.
[7] Durbin J, Koopman SJ. *Time Series Analysis by State Space Methods*. 2nd ed. Oxford University Press; 2012.
[8] Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. *J Am Stat Assoc.* 2007;102(477):359–378.
[9] Caso CC. *XC-SPI Postmodern Era Historical Success Model: Super-Shoe Era onward, 2020–2025 team-profile persistence, 2026 success rates, and historical title odds*. Oklahoma State University; 2026.