

COMPLETION-FIRST ANALYSIS OF SINGLE-LEG BALANCE: ESTIMANDS, PARTICIPANT-LEVEL INFERENCE AND SIMULATION

Jill M. Lucas

University of Lynchburg, Virginia, USA

Received: 21/01/2026 Revised: 19/4/2026 Accepted: 28/4/2026 Published: 20/06/2026

Correspondence: lucas.j@lynchburg.edu

ABSTRACT

To estimate the probability advantage of one single-leg stance condition over another while retaining attempts that terminate before 30 s. A pairwise score assigned priority to stance completion, then compared duration between two terminated attempts or centre-of-pressure velocity between two completed attempts. Independent and matched effects, equal participant weighting, and bounds for technically missing velocity were derived. Eight six-state probability models were evaluated at 20, 60 and 120 independent participants per group, with 2,000 generated datasets per combination. Four additional settings contained three correlated attempts per participant. The 56,000 datasets were analysed using projection-based 95% Wald intervals; independent-attempt analyses also used a 499-resample percentile bootstrap. In an exact 16-record dataset, 37 wins, 25 losses and two ties gave a net advantage of 0.1875; pairing the same records by row gave zero. Across independent-attempt models, absolute simulated bias was at most 0.007351. Coverage was 0.9300–0.9530 for projection intervals and 0.9310–0.9570 for bootstrap intervals. A completion-probability difference of 0.20 was offset by a completed-sway contribution of -0.190125 , leaving a net advantage of 0.009875. With correlated attempts, participant-level coverage was 0.9315–0.9420; treating attempts as independent reduced it to 0.7790–0.8080 and narrowed mean intervals by 33.97–34.50%. Completion, termination time and completed-stance sway define distinct contributions to the balance comparison. Participant-level inference preserves within-person dependence; the evaluated intervals retain finite-sample undercoverage. The evidence consists of mathematical derivations, constructed records and Monte Carlo data, without human force-platform measurements.

Keywords: kinesiology; single-leg stance; noncompletion; centre of pressure; generalized pairwise comparisons; participant weighting.

INTRODUCTION

During a 30-s single-leg stance task, a person can terminate the stance before the acquisition window ends or complete the window with measurable centre-of-pressure (CoP) movement. Completion answers whether the posture was sustained; CoP velocity describes movement during sustained stance. Deleting a terminated attempt removes the first outcome from a comparison restricted to successful execution.

Attempt-selection rules determine which question a balance assessment answers. Zumbrunn et al. used acceptance and selection rules for single-leg stance trials in children (Zumbrunn et al., 2011). Zaghlul et al. obtained three successful recordings through repeated attempts on a Lafayette stability platform (Zaghlul et al., 2023). The Lafayette outcome concerns time in balance on a tilting platform, whereas a fixed force platform permits CoP analysis. In both cases, success-conditioned performance differs from performance across initiated attempts.

The acquisition window also affects the outcome. Richmond et al. identified duration-related differences in CoP-derived

postural-control quantities (Richmond et al., 2023). A velocity calculated over an interrupted segment is therefore not automatically equivalent to a full 30-s value. In single-leg stance after fatigue, Kozinc et al. identified sex-dependent and interval-dependent sway changes (Kozinc et al., 2021). Lower velocity is consequently an explicit preference within the present comparison, subordinate to completing the stance, rather than a universal sign of superior physiological control.

Prioritized generalized pairwise comparisons (GPC) express the difference between probabilities of favourable and unfavourable cross-group comparisons (Pocock et al., 2012; Verbeeck et al., 2019). Dependence between endpoints and comparison thresholds affect their operating characteristics (Deltuvaite-Thomas and Burzykowski, 2022); exact permutation and bootstrap distributions are available for GPC statistics (Anderson and Verbeeck, 2023). Here the score has three branches: completion versus termination, longer duration between two terminated attempts, and lower velocity between two completed attempts. Its independent-group and matched-person effects are evaluated separately.

The analysis addresses three numerical questions. First, how much can completed-stance sway offset a completion advantage? Second, how does within-person correlation alter confidence-interval coverage? Third, how many comparisons remain determined when a completed record loses its velocity value? The calculations use six-state probability models, an exact 16-record dataset, and 56,000 generated datasets with specified sampling units.

METHODS

TARGET TASK AND ATTEMPT RECORD

The comparison concerns valid initiated single-leg stance attempts with a 30-s completion threshold. Let $\tau = 30$ s, let F be the time of the first motor termination, and define

$$\begin{cases} X = (C, T, V), \\ T = \min(F, \tau), \\ C = \mathbb{I}(F \geq \tau). \end{cases} \quad (1)$$

If no termination occurs by τ , later F is unobserved, whereas $C = 1$ and $T = \tau$ are observed. When $C = 0$, full-window sway V is undefined. An interrupted trace is not assigned a full-window velocity. Completion witnessed independently of a corrupted force signal remains $C = 1$, with technically missing V .

A motor termination is the event that ends the prescribed stance: non-support-foot contact with the floor, displacement of the support foot, grasping support, or assessor assistance. Technical interruption is recorded separately. The computational records contain the terminal status and duration directly; no physical stance session or force-platform acquisition was conducted. Table 1 identifies the values entering the calculations.

TABLE 1. RECORD DEFINITIONS AND NUMERICAL INPUTS. EVERY VALUE REFERS TO THE CONSTRUCTED OR GENERATED DATA USED IN THIS STUDY.

Quantity	Values and analysis role
Completion threshold	$\tau = 30$ s. A completed record has $C = 1$, $T = 30$ s; a terminated record has $C = 0$, $T < 30$ s.
Generated durations	Terminated states have $T = 6, 16, 26$ s. All generated completed states have $T = 30$ s.
Generated velocities	Completed states have $V = 10, 14, 18$ mm/s. Full-window velocity is undefined for every terminated state.
Exact-enumeration records	Eight records in each condition. Completed velocities: A, 12–24 mm/s in increments of 2; B, 8, 10, 12 mm/s. Terminated durations: A, 14 s; B, 6, 10, 14, 18, 22 s.
Repeated attempts	Three generated attempts per participant. A shared latent state is copied with probability 0.8; otherwise an independent state is drawn from the same group distribution.
Technical omission	The 24-mm/s value in A's seventh completed record is withheld for the missing-velocity calculation, while its $C = 1$ and $T = 30$ s remain known.

For a completed force-platform recording, resultant CoP path velocity over the common window is defined by

$$V = \frac{1}{\tau} \sum_{k=2}^K \sqrt{(x_k - x_{k-1})^2 + (y_k - y_{k-1})^2}, \quad (K-1)d = \tau, \quad (2)$$

where K is the number of samples, d the sampling interval in seconds, and (x_k, y_k) the CoP position in millimetres. The velocity values used in this study are numerical state labels in mm/s; they were not calculated from measured CoP trajectories. Completion and full-window velocity retain their separate roles in the score.

COMPARISON KERNEL AND ESTIMAND

For $\delta \geq 0$, define $\text{sgn}_\delta(z) = 1$ when $z > \delta$, -1 when $z < -\delta$, and zero otherwise. For an A attempt and a B attempt,

$$h(X_A, X_B) = \begin{cases} C_A - C_B, & C_A \neq C_B, \\ \text{sgn}_{\delta_T}(T_A - T_B), & C_A = C_B = 0, \\ \text{sgn}_{\delta_V}(V_B - V_A), & C_A = C_B = 1. \end{cases} \quad (3)$$

Positive values favour A. Completion defeats any motor-terminated attempt; among two failed attempts, longer stance is preferred; among two completed attempts, lower velocity is preferred. A difference exactly at a tolerance is a tie. Tolerances describe the comparison rule, not estimated measurement error or an established clinically important difference.

For independent draws from the two target attempt distributions,

$$\begin{cases} \Delta = E\{h(X_A, X_B)\} = P(h = 1) - P(h = -1), \\ \theta = \frac{1 + \Delta}{2} = P(h = 1) + \frac{1}{2}P(h = 0). \end{cases} \quad (4)$$

Δ is a net pairwise advantage and θ a tie-adjusted superiority probability. Neither is a difference in seconds or millimetres per second. The tie-adjusted win odds are $\theta/(1 - \theta)$ when $0 < \theta < 1$, whereas the win ratio is $P(h = 1)/P(h = -1)$ when its denominator is positive (Brunner et al., 2021). They differ when ties occur.

Write $p_g = P(C_g = 1)$, $\gamma_T = E\{\text{sgn}_{\delta_T}(T_A - T_B) \mid C_A = C_B = 0\}$ and $\gamma_V = E\{\text{sgn}_{\delta_V}(V_B - V_A) \mid C_A = C_B = 1\}$. Independence between groups gives

$$\Delta = p_A - p_B + (1 - p_A)(1 - p_B)\gamma_T + p_A p_B \gamma_V. \quad (5)$$

A component with zero stratum probability contributes zero and does not require an otherwise undefined conditional expectation. The identity follows by partitioning the four completion-status combinations. It shows how opposing secondary outcomes can offset a completion advantage. It also yields the sufficient dominance condition $p_A(1 - p_B) > 1/2$: more than half of all pairs are then automatic A wins, so $\Delta > 0$ irrespective of the remaining comparisons.

At $\delta_T = \delta_V = 0$, the observed records have a total preorder: completed attempts precede failures, completed attempts are ordered by lower velocity, and failures by longer duration. Eq. (4) is consequently the ordinary tie-adjusted Mann-Whitney superiority estimand on that ordering. Positive tolerances retain antisymmetry, $h(x, y) = -h(y, x)$, but the resulting pairwise ties need not be transitive. The zero-tolerance score therefore has the rank-comparison interpretation of GPC.

TABLE 2. QUESTIONS ANSWERED BY ALTERNATIVE ANALYSES OF THE SAME TASK. CHOICE OF ESTIMAND PRECEDES CHOICE OF ESTIMATOR.

Analysis	Target comparison	Information not summarized
Completed-attempt sway	Sway conditional on successful execution	Frequency and duration of motor-terminated attempts.
Completion proportion	Probability of reaching τ	Duration among failures and sway among completers.
Capped stance duration	Distribution of $\min(F, \tau)$	Full-window sway differences among completers.
Completion-first GPC	Wins minus losses under Eq. (3)	Component differences in seconds and mm/s; these remain separate outcomes.

INDEPENDENT-ATTEMPT ESTIMATION AND CONFIDENCE INTERVALS

With one valid attempt from each independently sampled participant, let $h_{ij} = h(X_{Ai}, X_{Bj})$ and

$$\hat{\Delta} = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} h_{ij}, \quad \hat{\theta} = (1 + \hat{\Delta})/2. \quad (6)$$

Every term has expectation Δ , so the estimator is unbiased under this sampling model. Its $n_A n_B$ cells are dependent because each record appears in multiple cells. A binomial standard error using that number of cells is inappropriate.

Define row means $r_i = n_B^{-1} \sum_j h_{ij}$, column means $c_j = n_A^{-1} \sum_i h_{ij}$, and their sample variances

$$\begin{cases} s_A^2 = \frac{\sum_i (r_i - \hat{\Delta})^2}{n_A - 1}, \\ s_B^2 = \frac{\sum_j (c_j - \hat{\Delta})^2}{n_B - 1}, \\ \hat{\text{se}} = \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}. \end{cases} \quad (7)$$

This is a first-order projection standard error, not an exact finite-sample variance formula. For nondegenerate distributions and increasing independent sample sizes, it supports normal approximation. The evaluated 95% Wald interval was

$$\left[\max(-1, \hat{\Delta} - 1.96\hat{\text{se}}), \min(1, \hat{\Delta} + 1.96\hat{\text{se}}) \right]. \quad (8)$$

Endpoints for θ follow by the affine transformation in Eq. (4). Both group sizes must exceed one. A zero estimated standard error produces a point interval and is flagged as degenerate, rather than interpreted as proof of population certainty.

The second evaluated interval independently resampled n_A A participants and n_B B participants with replacement, recomputed the complete comparison mean, and used the 2.5th and 97.5th percentiles of 499 replicates. Quantiles used linear interpolation. The resampling unit is the participant; pair cells are never independently resampled. The evaluated intervals are normal-projection and percentile-bootstrap approximations. Exact GPC distribution methods (Anderson and Verbeek, 2023) lie outside this interval comparison. The condition $\Delta = 0$ is a weak null; label permutation additionally depends on exchangeability or a randomized sharp null.

REPEATED ATTEMPTS AND MATCHED CONDITIONS

For participant i in A with $m_i > 0$ valid initiated attempts and participant j in B with $q_j > 0$, define

$$H_{ij} = \frac{1}{m_i q_j} \sum_{a=1}^{m_i} \sum_{b=1}^{q_j} h(X_{Aia}, X_{Bjb}), \quad \hat{\Delta}_{\text{person}} = \frac{1}{N_A N_B} \sum_{i,j} H_{ij}. \quad (9)$$

This estimand selects a participant uniformly in each target group, then selects an available valid attempt uniformly within that participant. It avoids giving participants with more recorded attempts more weight. Eq. (7) applies to row and column means of H with N_A, N_B in place of n_A, n_B , because the whole participant record is the independent unit. A cluster bootstrap resamples these whole records and preserves all within-person attempts.

If valid attempt counts depend on technical availability, Eq. (9) targets the available-attempt distribution, not automatically the distribution of every planned attempt. Motor-terminated attempts remain valid records and are not excluded for lacking V . Participants with $m_i = 0$ cannot enter this estimator; their absence changes its eligible population. To target every scheduled opportunity, additional assumptions or unresolved-outcome bounds are required. A participant with zero valid attempts has no within-person attempt distribution in Eq. (9); the comparison is conditional on a positive valid-attempt count in each included group.

When A and B are conditions experienced by the same participant, an alternative target is

$$\hat{\Delta}_{\text{matched}} = \frac{1}{N} \sum_{i=1}^N H_{ii}. \quad (10)$$

This compares attempts within the same person and generally differs from the independent-participant target. If $\pi_{ab} = P(C_A = a, C_B = b)$ under the matched attempt-selection rule, then

$$\Delta_{\text{matched}} = \pi_{10} - \pi_{01} + \pi_{00}\gamma_{T,\text{matched}} + \pi_{11}\gamma_{V,\text{matched}}. \quad (11)$$

Joint status probabilities replace the products in Eq. (5). With independent matched participants, a first-order standard error is $\text{sd}(H_{11}, \dots, H_{NN})/\sqrt{N}$; bootstrap resampling keeps each participant's two conditions together. The matched standard error is an analytical result. The Monte Carlo calculations used independent groups; the row-paired 16-record dataset evaluates the matched point effect exactly.

TECHNICAL LOSS AND UNRESOLVED COMPARISONS

For a complete independent-attempt grid of $M = n_A n_B$ cells, let W, L, R, U count known wins, losses, ties and unresolved comparisons, with $M = W + L + R + U$. Completion can resolve a cell even when a completed trace is missing. Without assuming values for unresolved scores,

$$\frac{W - L - U}{M} \leq \hat{\Delta}_{\text{full}} \leq \frac{W - L + U}{M}, \quad (12)$$

$$\frac{W + R/2}{M} \leq \hat{\theta}_{\text{full}} \leq \frac{W + R/2 + U}{M}. \quad (13)$$

These are finite-sample outer bounds, not confidence intervals. They are conservative because shared missing records can constrain multiple cells simultaneously; they need not be sharp in general. For Eq. (9), apply the same argument to normalized cell weights $1/(N_A N_B m_{ij})$; W, L, R, U then denote their total weights and $M = 1$. Eqs. (12)–(13) delimit unresolved values within a fixed comparison grid; sampling variation is a separate uncertainty.

COMPUTATIONAL DESIGN

The Monte Carlo calculation estimated bias, RMSE, interval coverage and zero-exclusion frequency for each model and sample-size combination (Morris et al., 2019). Eight discrete models used the states

$$\mathcal{S} = \{(0, 6, \text{NA}), (0, 16, \text{NA}), (0, 26, \text{NA}), (1, 30, 10), (1, 30, 14), (1, 30, 18)\},$$

where duration is in seconds and velocity in mm/s. For completion probability p and conditional mass vectors f, v on the three failure and completion states, the six-state mass vector is $q = ((1-p)f, pv)$. Table 3 gives every input. Primary simulation comparisons used $\delta_T = \delta_V = 0$. Completion probabilities range from 0.05 to 0.95. The conditional state distributions are fixed numerical inputs, without estimation from a clinical sample.

If $K_{st} = h(\mathcal{S}_s, \mathcal{S}_t)$, the exact population truth is $q_A^T K q_B$. Equal-distribution null models have truth set to exactly zero using the antisymmetry of K , avoiding floating-point artifacts at interval boundaries. At $n = 20, 60, 120$ participants per group, each model generated 2,000 independent datasets; each participant supplied one attempt. Both intervals in Eqs. (7)–(8) and the percentile bootstrap were evaluated on the same dataset. Bias, root mean squared error (RMSE), inclusive-endpoint coverage, interval width and the frequency of excluding zero were calculated. Exclusion of zero is a two-sided rejection rate in exact-null models and a detection probability otherwise.

TABLE 3. FULLY SPECIFIED INDEPENDENT-ATTEMPT SIMULATION MODELS. VECTORS ARE ORDERED BY INCREASING FAILURE DURATION OR INCREASING COMPLETED VELOCITY. HERE $w = (0.25, 0.50, 0.25)$, $\ell = (0.60, 0.30, 0.10)$, $u = (0.10, 0.30, 0.60)$, $e = (0, 1, 0)$, $\ell_* = (0.80, 0.15, 0.05)$ AND $u_* = (0.05, 0.15, 0.80)$.

Model	p_A	p_B	f_A	f_B	v_A	v_B	Δ	Purpose
S1	.50	.50	w	w	w	w	0	Midrange null
S2	.95	.95	w	w	w	w	0	Completion ceiling
S3	.05	.05	w	w	w	w	0	Completion floor
S4	.65	.45	w	w	u	ℓ	.009875	Opposing sway
S5	.60	.60	w	w	ℓ	u	.234000	Sway advantage
S6	.35	.35	u	ℓ	w	w	.274625	Duration advantage
S7	.55	.50	ℓ	u	u_*	ℓ_*	-.3334375	Conflicting components
S8	.80	.80	e	e	e	e	0	Many exact ties

A further four settings used S1 and S4 with $N = 20$ or 60 participants per group and three attempts per participant. Each participant received a latent state $Z \sim q$. Each attempt copied Z with probability $\rho = 0.8$ and otherwise drew a fresh state independently from q . Attempts therefore retained marginal distribution q ; the correlation between identical nondegenerate state indicators for two attempts was $\rho^2 = 0.64$. Different participants were independent. For each of 2,000 datasets per setting, participant-weighted projection intervals were compared with projection intervals incorrectly treating the $3N$ attempts as independent. Equal attempt counts made the two point estimates identical, isolating the effect of the variance calculation.

The total was $8 \times 3 \times 2,000 + 2 \times 2 \times 2,000 = 56,000$ datasets. The maximum Monte Carlo standard error (MCSE) of a proportion with 2,000 replications is 0.0112; near coverage 0.95 it is approximately 0.0049. MCSE of bias was the standard deviation of the estimates divided by $\sqrt{2,000}$. Separate deterministic seeds were used for data generation and bootstrap sampling. The computational environment was Python 3.12.14, NumPy 2.3.5 and Matplotlib 3.10.8. The analysis was not preregistered.

RESULTS

EXACT ENUMERATION AND ESTIMAND DIFFERENCES

Table 4 contains 16 constructed records: seven completed and one terminated in A, and three completed and five terminated in B. At zero tolerances, the 64 independent cross-condition comparisons gave $W = 37$, $L = 25$ and $R = 2$, hence $\hat{\Delta} = 0.1875$ and $\hat{\theta} = 0.59375$. Completed-attempt mean velocity was 18 mm/s in A and 10 mm/s in B. The difference in conclusions reflects different targets: conditional sway favoured B, while the defined completion-first target favoured A.

TABLE 4. SIXTEEN CONSTRUCTED RECORDS FOR THE 64 CROSS-CONDITION COMPARISONS. A DASH DENOTES FULL-WINDOW VELOCITY UNDEFINED BY MOTOR TERMINATION. MATCHING A AND B BY ROW DEFINES THE SEPARATE EIGHT-PAIR CALCULATION.

Row	Condition A			Condition B		
	C	T (s)	V (mm/s)	C	T (s)	V (mm/s)
1	1	30	12	1	30	8
2	1	30	14	1	30	10
3	1	30	16	1	30	12
4	1	30	18	0	6	–
5	1	30	20	0	10	–
6	1	30	22	0	14	–
7	1	30	24	0	18	–
8	0	14	–	0	22	–

The completion contribution was 0.5, the completed-completed contribution $-20/64 = -0.3125$, and the failed-failed contribution zero. Here $(7/8)(5/8) = 35/64 > 1/2$, so A's positive net advantage was already guaranteed by the dominance condition. Completion alone resolves 35 A wins and three A losses; the remaining 26 pairs compare duration or velocity. The win ratio was $37/25 = 1.4808$, and the tie-adjusted win odds $38/26 = 1.4615$.

Across the defined nine-combination grid $\delta_T \in \{0, 2, 5\}$ s and $\delta_V \in \{0, 2, 4\}$ mm/s, net advantages were 0.1875, 0.21875 and 0.265625 for $\delta_V = 0, 2, 4$, respectively, at every tested δ_T . At $(\delta_T, \delta_V) = (5, 4)$, the counts were 36 wins, 19 losses and nine ties. Increasing a tolerance can remove either favourable or unfavourable differences and need not reduce $|\Delta|$.

Matching A and B by table row gave four wins and four losses, so $\hat{\Delta}_{\text{matched}} = 0$ and $\hat{\theta}_{\text{matched}} = 0.5$. The matched status probabilities were $\pi_{10} = 4/8$, $\pi_{01} = 0$, $\pi_{11} = 3/8$ and $\pi_{00} = 1/8$; both conditional matched comparisons had $\gamma = -1$. Eq. (11) therefore gave $4/8 - 3/8 - 1/8 = 0$. This difference is created by the specified joint pairing, not by changing marginal records.

If only A row 7's velocity was technically lost while completion remained known, three cells became unresolved: $W = 37$, $L = 22$, $R = 2$, $U = 3$. Eqs. (12)–(13) gave $\hat{\Delta}_{\text{full}} \in [0.1875, 0.28125]$ and $\hat{\theta}_{\text{full}} \in [0.59375, 0.640625]$. Both endpoints are attainable in this example by positioning the missing velocity above or below every completed B velocity. No sampling confidence interval was attached to these fixed records.

INDEPENDENT-ATTEMPT SIMULATION

Across the 24 settings, the largest absolute simulated bias was 0.007351 in S6 at $n = 20$, equivalent to 1.80 MCSEs of bias. The unbiasedness conclusion follows from Eq. (6); the simulation supplies a numerical check and assesses interval behaviour. Projection coverage ranged from 0.9300 to 0.9530 and bootstrap coverage from 0.9310 to 0.9570. Undercoverage was most evident in several small-sample alternatives, and the percentile bootstrap did not uniformly correct it.

Table 5 contains the $n = 60$ results. Under exact-null models S1, S2, S3 and S8, the zero-exclusion frequencies across all sample sizes were 0.0485–0.0625 for projection intervals and 0.0430–0.0650 for bootstrap intervals. At $n = 60$, the largest null rejection frequency was 0.0625 for projection intervals and 0.0650 for bootstrap intervals, both in S3.

In S4, conditional sway contributed $-.190125$, almost cancelling the completion advantage of .20, so $\Delta = .009875$ despite higher completion in A. In S7, the adverse duration and sway distributions outweighed the .05 completion advantage, giving $\Delta = -.3334375$. These examples show that a higher completion proportion alone does not generally determine the aggregate sign. In S5, projection detection probability increased from 0.2580 at $n = 20$ to 0.8765 at $n = 120$; in S6 it increased from 0.3305 to 0.9505. The detection probabilities apply to the stated conditional mass vectors and completion probabilities.

TABLE 5. INDEPENDENT-ATTEMPT PERFORMANCE AT 60 PARTICIPANTS PER GROUP, BASED ON 2,000 DATASETS PER MODEL. C IS COVERAGE AND E THE FREQUENCY OF EXCLUDING ZERO; SUBSCRIPTS P AND B DENOTE PROJECTION AND PERCENTILE-BOOTSTRAP INTERVALS. BIAS AND RMSE REFER TO THE COMMON POINT ESTIMATOR. NUMERICAL ENTRIES ARE ROUNDED.

Model	Δ	Bias	RMSE	C_P	C_B	E_P	E_B
S1	0.0000	0.0017	0.1022	0.9510	0.9475	0.0490	0.0525
S2	0.0000	0.0001	0.0987	0.9475	0.9460	0.0525	0.0540
S3	0.0000	0.0030	0.1013	0.9375	0.9350	0.0625	0.0650
S4	0.0099	-0.0024	0.1105	0.9405	0.9365	0.0595	0.0620
S5	0.2340	0.0019	0.1037	0.9425	0.9440	0.6185	0.6105
S6	0.2746	0.0030	0.1037	0.9460	0.9440	0.7500	0.7485
S7	-0.3334	0.0000	0.1010	0.9450	0.9440	0.8840	0.8790
S8	0.0000	0.0020	0.0723	0.9515	0.9570	0.0485	0.0430

REPEATED-ATTEMPT SIMULATION

Figure 1 shows the consequence of using the wrong independent unit. Participant-based coverage ranged from 0.9315 to 0.9420, whereas pooled-attempt coverage ranged from 0.7790 to 0.8080. In null model S1, the false-positive frequency was 0.0615 at $N = 20$ and 0.0580 at $N = 60$ with participant inference, compared with 0.1920 and 0.2050 with pooled-attempt inference. In S4, mean interval widths were 0.6499 and 0.3734 for participant inference at $N = 20$ and 60, compared with 0.4257 and 0.2464 for pooled-attempt inference. Zero-standard-error or zero-width intervals did not occur in the evaluated settings.

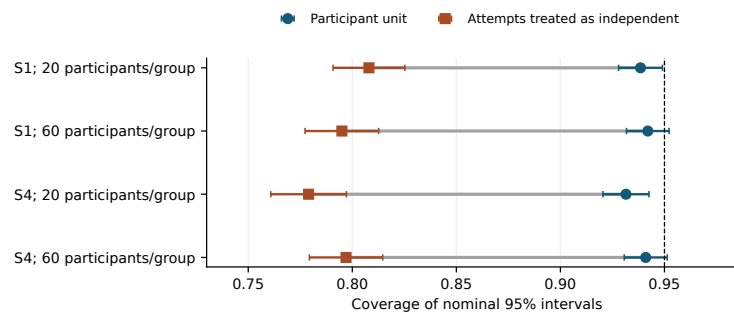


FIGURE 1. COVERAGE UNDER THREE CORRELATED ATTEMPTS PER PARTICIPANT. CIRCLES USE PARTICIPANT-LEVEL ROW AND COLUMN PROJECTIONS; SQUARES TREAT THE SAME ATTEMPTS AS INDEPENDENT. HORIZONTAL BARS REPRESENT APPROXIMATELY 95% MONTE CARLO UNCERTAINTY, $\pm 1.96 \sqrt{\hat{c}(1-\hat{c})/2,000}$, FOR ESTIMATED COVERAGE $\hat{c}$; THEY ARE NOT CONFIDENCE INTERVALS FOR A BALANCE EFFECT. THE DASHED LINE MARKS NOMINAL COVERAGE 0.95. EACH ROW REPRESENTS 2,000 INDEPENDENT DATASETS WITH LATENT-STATE COPYING PROBABILITY 0.8.

DISCUSSION

COMPLETION AND SWAY CAN FAVOUR DIFFERENT CONDITIONS

The 16-record calculation separates the two outcomes directly. A completed seven of eight attempts, compared with three of eight in B, but its completed-attempt mean velocity was 18 mm/s rather than 10 mm/s. The completion contribution of 0.5 exceeded the adverse completed-sway contribution of -0.3125 , producing $\hat{\Delta} = 0.1875$. Because 35 of 64 pairs were automatic A wins, the sign was fixed by completion before the secondary comparisons were evaluated.

The six-state models remove that dominance. In S4, $p_A = 0.65$ and $p_B = 0.45$ gave a completion contribution of 0.20, while the completed-sway contribution was -0.190125 . The resulting $\Delta = 0.009875$ was only 4.94% of the completion-probability difference. In S7, a 0.05 completion advantage was outweighed by duration and sway, leaving $\Delta = -0.3334375$. The sign of the completion difference therefore does not generally determine the sign of the pairwise effect. Completed-only velocity and the completion-first score answer different questions; their opposing conclusions are compatible with the same records.

THE 30-S THRESHOLD IS A SUBSTANTIVE PRIORITY

Eq. (3) gives a completed 30-s attempt priority over a terminated 29.9-s attempt. The duration tolerance acts only when both attempts terminate, so it leaves this cross-status discontinuity intact. Completion first is appropriate to a question

that assigns priority to sustaining the entire stance window. A question centred on control during a shorter successful segment has a different observation window and outcome.

The tolerance calculations also show that tie assignment need not reduce the absolute effect. In the 16-record dataset, increasing the velocity tolerance from zero to 4 mm/s changed $\hat{\Delta}$ from 0.1875 to 0.265625. At $\delta_T = 5$ s and $\delta_V = 4$ mm/s, 36 wins and 19 losses remained after nine ties. Removing comparisons from one direction more often than the other alters the net advantage asymmetrically; the tolerance values here are numerical choices, not validated clinical thresholds.

PARTICIPANT DEPENDENCE DETERMINES INTERVAL CALIBRATION

With three attempts and latent-state copying probability 0.8, identical state indicators had within-person correlation 0.64. Equal attempt counts made participant-weighted and pooled-attempt point estimates identical, while the pooled variance calculation ignored this dependence. Across the four settings, mean intervals narrowed by 33.97–34.50% and coverage fell by 0.1305–0.1525. In S1 at 60 participants per group, the null was rejected in 410 of 2,000 pooled-attempt analyses, compared with 116 participant-level analyses. The effect estimate alone did not reveal the error in the sampling unit.

Matching creates a second distinction. Row-pairing the exact dataset produced four wins and four losses, whereas independent cross-condition enumeration produced 37 wins and 25 losses. Eq. (11) uses joint completion probabilities because the two conditions belong to the same person. Independent and matched effects consequently cannot be interchanged on the basis of identical marginal completion counts.

INFERENTIAL AND MEASUREMENT LIMITS

The strongest independent-attempt undercoverage occurred in S6 at 20 participants per group: projection coverage was 0.9300 and bootstrap coverage 0.9310. The projection interval covered the true effect in 1,860 of 2,000 datasets; the bootstrap interval covered it in 1,862. The 40 and 38 additional failures relative to nominal 95% coverage identify finite-sample approximation error. At 60 participants per group, S3 coverage was 0.9375 and 0.9350, showing that increasing sample size did not eliminate every observed shortfall. The 499-resample bootstrap quantiles also contain resampling error. Exact and studentized intervals were outside the calculation, as were Monte Carlo evaluations of positive tolerances, unequal attempt counts and matched inference.

Technical omission had a narrower consequence in the exact dataset. Withholding A's 24-mm/s completed value left 61 of 64 scores known because completion still resolved its five comparisons against terminated B attempts. The three unresolved scores bounded $\hat{\Delta}_{\text{full}}$ between 0.1875 and 0.28125. This fixed-grid bound addresses absent velocity values; it does not include sampling uncertainty.

The evidence covers specified state distributions and algebraic identities. CoP trajectories, event-coding agreement, test-retest reliability and responsiveness were not measured. Intraclass-correlation methods address component reliability in repeated physical measurements (Koo and Li, 2016); this Monte Carlo calculation addresses confidence-interval behaviour for an aggregate comparison. Balance-test evidence for fall discrimination concerns observed clinical populations (Kozinc et al., 2020). The present net advantage has no estimated association with falls, injury or independently assessed physical function.

CONCLUSION

Completion-first analysis retains the 30-s completion outcome alongside terminated-stance duration and completed-stance velocity. The exact dataset gave an independent net advantage of 0.1875 and a matched advantage of zero. In the correlated-attempt calculations, participant-level intervals covered the true effect in 93.15–94.20% of datasets, compared with 77.90–80.80% under independent-attempt pooling. These effects and interval frequencies refer to the specified state distributions.

DECLARATIONS

CONFLICT OF INTEREST

The author declares that there are no conflicts of interest related to this study.

DATA AVAILABILITY

The data supporting the findings of this study are available from the author.

FUNDING

This research received no external funding.

REFERENCES

1. Anderson WN, Verbeeck J. Exact permutation and bootstrap distribution of generalized pairwise comparisons statistics. *Mathematics*. 2023;11(6):1502.
2. Brunner E, Vandemeulebroecke M, Mütze T. Win odds: an adaptation of the win ratio to include ties. *Stat Med*. 2021;40(14):3367–3384.
3. Deltuvaite-Thomas V, Burzykowski T. Operational characteristics of generalized pairwise comparisons for hierarchically ordered endpoints. *Pharm Stat*. 2022;21(1):122–132.
4. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155–163.
5. Kozinc Ž, Löfler S, Hofer C, Carraro U, Šarabon N. Diagnostic balance tests for assessing risk of falls and distinguishing older adult fallers and non-fallers: a systematic review with meta-analysis. *Diagnostics (Basel)*. 2020;10(9):667.
6. Kozinc Ž, Trajković N, Smajla D, Šarabon N. The effect of fatigue on single-leg postural sway and its transient characteristics in healthy young adults. *Front Physiol*. 2021;12:720905.
7. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med*. 2019;38(11):2074–2102.
8. Pocock SJ, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur Heart J*. 2012;33(2):176–182.
9. Richmond SB, Otto G, Dames KD. Characterization of trial duration in traditional and emerging postural control measures. *J Biomech*. 2023;147:111438.
10. Verbeeck J, Spitzer E, de Vries T, van Es GA, Anderson WN, Van Mieghem NM, et al. Generalized pairwise comparison methods to analyze (non)prioritized composite endpoints. *Stat Med*. 2019;38(30):5641–5656.
11. Zaghlul N, Goh SL, Razman R, Danaee M, Chan CK. Test-retest reliability of the single leg stance on a Lafayette stability platform. *PLoS One*. 2023;18(1):e0280361.
12. Zumbunn T, MacWilliams BA, Johnson BA. Evaluation of a single leg stance balance test in children. *Gait Posture*. 2011;34(2):174–177.