

SINGLE-LEG STANCE DURATION: INTERRUPTION BOUNDS AND EXACT CONFIDENCE COVERAGE

Arjun Sharma, Aditya Singh and Nikhil Gupta*

MYAS-GNDU Department of Sports Sciences & Medicine, Guru Nanak Dev University, Amritsar, Punjab, India

Received: 12/7/2024 Revised: 12/9/2024 Accepted: 03/12/2024 Published: 20/12/2024

Correspondence: gnikhil78@yahoo.com

ABSTRACT

An observation-only stop during single-leg stance leaves a bounded amount of unobserved task time. At a fixed horizon, motor failure and withdrawal terminate the duration score, while verified horizon completion determines it exactly. This study derives sharp closed bounds for mean task duration and paired condition differences from termination codes and verified elapsed times. A bounded-mean confidence interval controls coverage of the entire population identified set for a fixed number of independent participants and prespecified endpoint ranges. Exact enumeration covers 90 binomial and 12 multinomial configurations. Twelve constructed pairs at a 30-second horizon give a mean difference of $[-1.33, 2.00]$ seconds; deleting the two interrupted pairs gives 2.20 seconds for the remaining ten. With 20 independent pairs, a 5-second interruption time, and interruption probability 0.02, the normal-approximation interval has whole-set coverage of 33.24%. The bounded interval attains at least 99.59% coverage across all 102 configurations, above its 95% guarantee. Interruption associated with potential performance shifts the complete-case limit by ± 3.43 seconds despite a zero underlying mean difference. These results quantify the effects of event classification, interruption timing, and endpoint range on duration uncertainty.

Keywords: single-leg stance; postural control assessment; restricted mean duration; procedural interruption; partial identification; finite-sample confidence interval.

INTRODUCTION

At a 30-second single-leg stance horizon, motor failure at 9 seconds fixes the task score at 9 seconds. An observation-only stop at the same instant leaves a score between 9 and 30 seconds, provided stance remains valid through the stop. Verified completion fixes the score at 30 seconds. The elapsed time therefore acquires its meaning from the event that ends the record. Combining these three events in one stopping code obscures the amount of task performance established by observation.

Ceiling effects occur in standing-balance tests for people with hip osteoarthritis and in timed static balance assessment in children (Choi et al., 2014; Condon and Cremin, 2014). A fixed ceiling restricts discrimination between performances beyond that ceiling. Within-horizon duration remains fully known for a person who completes the task. Shen and Mak used Kaplan-Meier analysis for single-leg stance, censoring the absence of balance loss at 30 seconds (Shen and Mak, 2014). For mean duration restricted to that horizon, the same record supplies a complete score. The restricted-mean identity connects task seconds to the area under the survival curve (Royston and Parmar, 2011, 2013). Defining how each termination enters the outcome precedes estimation, as the ICH E9(R1) guidance specifies for events affecting the interpretation of a measurement (International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, 2019).

Statistical approaches differ in the assumptions they place on unobserved duration. Zhang and colleagues developed restricted-mean estimation for interval-censored event times using smoothing (Zhang et al., 2020). Missing-outcome sensitivity analyses vary assumptions about unobserved means or proportions (Cole et al., 2023; White et al., 2011);

uncertainty intervals under non-ignorable missingness retain the distinction between absent outcome information and sampling error (Genbäck et al., 2015). Partial-identification inference also separates coverage of one admissible parameter from coverage of the entire set of admissible population values (Stoye, 2021).

Here, the observation record determines a lower and an upper task duration. Their participant-level differences determine sharp mean-contrast bounds. A fixed-sample concentration inequality then adds sampling protection to those bounds. Exact count enumeration quantifies the confidence coverage obtained when interruptions are rare or depend on potential stance duration. The derivation uses restricted-mean and bounded-variable results to address three specific decisions in stance testing: which event fixes the score, how many seconds remain unresolved, and how much uncertainty a population comparison must retain.

METHODS

STANCE ONSET, TERMINATION, AND OBSERVATION STOPS

Let $\tau > 0$ denote the horizon chosen before testing. Define the support limb, visual condition, footwear, surface, arm position, permitted supporting-foot displacement, familiarization, recovery period, and scored attempt schedule. Timing starts once the non-supporting foot has cleared the surface and the prescribed posture has been established; the later event sets onset. Both compared conditions use the same onset and posture criteria.

Motor failure occurs when the non-supporting foot contacts the surface, the supporting foot leaves its permitted position, unplanned external support is used, or balance-related rescue is required. Withdrawal occurs at a participant request, discomfort-related cessation, or a non-balance safety termination. Both events fix the duration score at termination. Separate codes preserve their respective contributions to observed time loss.

A procedural observation stop requires independent confirmation of valid stance through its recorded time. A silent display failure qualifies when a synchronized recording verifies stance up to that instant. A sound, physical perturbation, instruction, or examiner action may alter balance itself. Classification as an observation-only stop then requires evidence that removing the event preserves the task and its trajectory through the observed interval.

RESTRICTED TASK DURATION AND OBSERVED DURATION

For a fixed stance condition and attempt schedule, let $T^{(0)}$ be the time to motor failure or withdrawal under a specified measurement procedure that prevents observation-only interruption. With no task-ending event, set $T^{(0)} = \infty$. Define

$$\begin{cases} S^{(0)} = \min(T^{(0)}, \tau), \\ \mu = E(S^{(0)}), \\ X = \min(T^{(0)}, C, \tau), \\ \nu = E(X), \end{cases} \quad (1)$$

where C is the procedural observation-stop time. The representation of X requires three conditions:

1. The uninterrupted procedure is well defined and retains the same task, withdrawal rules, and attempt schedule.
2. The actual stance trajectory agrees with the uninterrupted trajectory up to the first observed termination. Preventing the recording stop has no anticipatory effect on performance before that time.
3. Task criteria and last-confirmed valid stance are observed correctly, or their uncertainty is retained in the admissible interval.

These conditions permit dependence between C and $T^{(0)}$. The observation stop may occur more often during short or long tasks. Its classification still requires a well-defined interruption-free measurement procedure and agreement of the observed and potential trajectories up to the stop.

The mean μ includes motor failure and withdrawal under the defined task; ν also reflects interruption of observation. A motor-failure-only outcome requires separate treatment of withdrawal as a competing event. Causal interpretation of a paired condition difference additionally requires justified condition assignment, order, recovery, and carryover assumptions.

TERMINATION CODES AND TIMING EVIDENCE

Table 1 defines exact-time codes. Time uncertainty is handled subsequently. If a task-ending event is definitely simultaneous with a procedural stop, the task event identifies the score at that time. If their ordering cannot be resolved at the recording resolution, retain both possible interpretations. For confirmed simultaneous motor failure and withdrawal, the score is exact but the cause is U unless the first cause is independently established. At τ , a confirmed task event takes precedence over H for cause coding; either produces the same restricted score.

TABLE 1. EXACT-TIME ADJUDICATION AND THE RESTRICTED TASK INFORMATION RETAINED. ALL INTERVALS ARE INTERPRETED IN THEIR CLOSURE.

Code	Observation	Evidence required	Score interval
F	Motor termination at x	Prespecified balance criterion or balance-related rescue	$[x, x]$
W	Withdrawal at x	Participant request or non-balance safety termination	$[x, x]$
P	Observation-only stop at $x < \tau$	Valid stance through x ; measurement-consistency assumptions justified	$[x, \tau]$
H	Horizon completion	Valid stance to τ without an earlier task event	$[\tau, \tau]$
U	Unresolved first cause	Retain the set of admissible causes and their supporting times	Hull of admissible intervals
M	Wholly unrecorded prescribed attempt	No supported lower duration; attempt remains in the fixed target schedule	$[0, \tau]$

Each record contains participant and examiner identifiers, condition, attempt number, horizon, protocol identifier, timing instrument, onset and stop-time limits, last-confirmed valid time, termination code or admissible code set, rescue or withdrawal criterion, and adjudication status. The fixed attempt schedule retains the horizon and attempt count when a record is absent. All elapsed times enter the calculation in seconds.

Retain examiner and synchronized-recording classifications separately. Two reviewers independently apply the termination criteria. Document the evidence resolving each disagreement; retain U when the evidence does not determine the first cause. Examiner agreement requires evaluation in a separate observation study.

SCORE INTERVALS FROM VERIFIED ELAPSED TIME

The pointwise identity $\min(T^{(0)}, \tau) = \int_0^\tau \mathbf{1}(T^{(0)} > t) dt$ yields

$$\mu = \int_0^\tau \Pr(T^{(0)} > t) dt. \quad (2)$$

With only a common administrative limit and no earlier loss of observation, every restricted score is known. The sample mean of the scores equals the area under the empirical Kaplan–Meier curve to τ : before the horizon, there are no censoring removals from its risk set. A completed horizon observation therefore identifies $S^{(0)} = \tau$, although it leaves unrestricted $T^{(0)}$ unresolved. No continuation beyond τ is needed for Equation 2.

Write the justified score interval for record i as $[\ell_i, u_i] \subseteq [0, \tau]$. With exact time and an adjudicated F, W, P, or H code,

$$\ell_i = X_i, \quad u_i = X_i + (\tau - X_i)\mathbf{1}(D_i = P). \quad (3)$$

Table 2 gives the limits for uncertain times and causes. With several admissible causes, use the smallest justified lower endpoint and the largest justified upper endpoint. The resulting hull retains the extreme attainable mean values. Disconnected feasible sets may exclude individual durations inside that hull.

TABLE 2. CONVERSION OF UNCERTAIN OBSERVATION INTO CONSERVATIVE SCORE LIMITS. CHRONOLOGY AND THE MEASUREMENT ASSUMPTIONS MUST FIRST BE CHECKED.

Available evidence	Justified limits
F or W within elapsed interval $[a, b]$	$[\min(a, \tau), \min(b, \tau)]$, with $0 \leq a \leq b$
P with valid stance last confirmed at elapsed time a	$[\min(a, \tau), \tau]$; a later recording stop does not strengthen a
Cause could be F, W, or P	Hull of cause-specific intervals; $[x, \tau]$ when all share a verified elapsed stop x
Unknown cause; no positive supported duration	$[0, \tau]$, provided the uninterrupted target remains well defined
Absolute start in $[s_-, s_+]$, event in $[e_-, e_+]$	Elapsed limits $[\max(0, e_- - s_+), \max(0, e_+ - s_-)]$, clipped to $[0, \tau]$
Last valid absolute instant v , uncertain start	Procedural lower limit $\min\{\tau, \max(0, v - s_+)\}$
Missing prescribed attempt	$[0, \tau]$; retain its position in the fixed attempt schedule

Check incompatible chronology and an absent onset against the observation record. Without evidence supporting a narrower interval, retain $[0, \tau]$. Interpretation for μ also requires preservation of the defined task; widening the interval cannot establish that condition. Failure to initiate can receive a zero score only when the protocol defines it that way before testing. A wholly unrecorded scheduled attempt receives $[0, \tau]$.

MEAN AND PAIRED-DIFFERENCE BOUNDS

For n records representing the target units,

$$\begin{cases} \hat{L} = \frac{1}{n} \sum_i \ell_i \leq \bar{S}^{(0)} \leq \hat{U} = \frac{1}{n} \sum_i u_i, \\ \hat{U} - \hat{L} = \frac{1}{n} \sum_i (u_i - \ell_i). \end{cases} \quad (4)$$

Eq. 4 bounds the fixed-record mean. When the score intervals have no additional joint restrictions, endpoint assignments attain each extreme in the closure. Convex combinations attain intermediate means for interval feasible sets. Verified stance through a P stop can require $T_i^{(0)} > X_i$; the lower endpoint is then an infimum approached from above.

At population level, $L = E(\ell) \leq \mu \leq U = E(u)$. For exact adjudicated records,

$$U - L = E\{(\tau - X)\mathbf{1}(D = P)\} = \begin{cases} p_P\{\tau - E(X | D = P)\}, & p_P > 0, \\ 0, & p_P = 0. \end{cases} \quad (5)$$

The unconditional expression remains defined at zero interruption probability. With broader uncertainty codes, the general width is $E(u - \ell)$ and includes their unresolved durations.

For paired participant-level scores, define

$$\begin{cases} d_i^- = \ell_{iA} - u_{iB}, \\ d_i^+ = u_{iA} - \ell_{iB}, \\ \bar{d}^- \leq \bar{S}_A^{(0)} - \bar{S}_B^{(0)} \leq \bar{d}^+. \end{cases} \quad (6)$$

The population contrast satisfies $\Delta = E(S_A^{(0)} - S_B^{(0)}) \in \mathcal{I} = [L_\Delta, U_\Delta]$, where $L_\Delta = E(d^-)$ and $U_\Delta = E(d^+)$. Under unrestricted joint continuations, these limits are jointly sharp. Pairing preserves within-person dependence for sampling calculations; it does not itself restrict unresolved continuations.

If each participant has a fixed R attempts, average endpoints before forming the contrast for an average-performance outcome. For maximum-of- R performance, take $\max_r \ell_{ir}$ and $\max_r u_{ir}$. A missing prescribed attempt retains its interval. Dropping it or changing R changes the target. Participants, not attempts, are the independent sampling units throughout the inference below.

Write an interrupted score as $S_i^{(0)} = \ell_i + r_i(u_i - \ell_i)$, $0 \leq r_i \leq 1$. The fraction r_i specifies how much of the unresolved duration is assigned to continued stance. Setting equal fractions across records imposes an additional assumption. The full paired bounds in Equation 6 retain all admissible fractions.

CONFIDENCE COVERAGE OF THE POPULATION BOUNDS

Suppose the paired participants are independent and identically distributed from the target population. Let $d_i^- \in [a_-, b_-]$ and $d_i^+ \in [a_+, b_+]$ almost surely, with ranges $R_- = b_- - a_-$ and $R_+ = b_+ - a_+$ specified independently of the observed sample. For any fixed $n \geq 1$ and $0 < \alpha < 1$, put

$$\begin{cases} h_n = \sqrt{\frac{\log(2/\alpha)}{2n}}, \\ \mathcal{C}_B = [\max\{a_-, \bar{d}^- - R_- h_n\}, \min\{b_+, \bar{d}^+ + R_+ h_n\}]. \end{cases} \quad (7)$$

Then

$$\Pr\{\mathcal{I} \subseteq \mathcal{C}_B\} \geq 1 - \alpha. \quad (8)$$

A one-sided bounded-mean tail inequality assigns error probability $\alpha/2$ to each endpoint; the union bound gives Equation 8. Within-participant dependence between d^- and d^+ is permitted. Concentration inequalities also support time-uniform inference (Howard et al., 2021); the interval here requires a sample size fixed before observation.

Without further design information, use $a_- = a_+ = -\tau$ and $b_- = b_+ = \tau$, so both ranges equal 2τ . Narrower ranges are admissible only if guaranteed by the protocol or an explicit probability law before observation. Sample minima and maxima do not supply such guarantees. A known constant endpoint has range zero; an endpoint that merely happens to be constant in a small sample does not.

For comparison, the normal-approximation outer interval is

$$\mathcal{C}_N = [\max\{a_-, \bar{d}^- - z_{1-\alpha/2}s_-/\sqrt{n}\}, \min\{b_+, \bar{d}^+ + z_{1-\alpha/2}s_+/\sqrt{n}\}], \quad (9)$$

where s_- and s_+ are sample standard deviations and $n > 1$. For a fixed sampling law with positive endpoint variances, the marginal central limit theorems and consistent variances give asymptotic full-set coverage of at least $1 - \alpha$. A truly deterministic endpoint needs no tail approximation. The procedure has no distribution-free finite-sample guarantee, and observing zero variance does not establish determinism. No universal sample-size threshold makes the approximation reliable across arbitrarily rare interruptions.

Coverage is defined by containment of the entire $\mathcal{I}$. Containment of one particular $\Delta \in \mathcal{I}$ imposes a weaker requirement. A positive population contrast supported across all admissible continuations requires the lower confidence limit to exceed zero. A clinical claim requires a separately justified threshold in seconds.

TIME LOSS ASSIGNED TO THE FIRST TERMINATION

When the observed stop duration X is known, include U as an unresolved first-cause category and define

$$A_k = E\{(\tau - X)\mathbf{1}(D = k)\}, \quad k \in \{F, W, P, U\}, \quad \tau - \nu = \sum_k A_k. \quad (10)$$

H contributes zero. With $F_k(t) = \Pr(X \leq t, D = k)$, integration gives $A_k = \int_0^\tau F_k(t) dt$, the corresponding restricted mean time-lost identity for competing risks (Conner and Trinquart, 2021). Each A_k assigns the observed shortfall to its first termination. It does not measure the causal effect of preventing that event. U retains unresolved causes. Missing durations require intervals for ν and the loss totals.

If correctly classified complete durations have absolute timing error at most δ , clipping to $[0, \tau]$ does not enlarge the error. A paired complete-score mean can consequently change by at most 2δ . This result does not protect against wrong codes or an incorrect time origin.

EXACT NUMERICAL EVALUATION

Set $\tau = 30$ seconds and $\alpha = 0.05$. Enumerate 90 binomial configurations from 15 observation distributions and six sample sizes, and 12 multinomial configurations from three distributions and four sample sizes. Every possible category-count outcome contributes its exact probability. Logarithmic factorials evaluate the count probabilities; their numerical sums differ from one by less than 2×10^{-10} .

Horizon-dominated law. B always completes 30 seconds. A also has $S_A^{(0)} = 30$ but is procedurally interrupted at c with probability p , otherwise completing the horizon. Use $n \in \{20, 50, 100, 200, 500, 1000\}$, $p \in \{0.005, 0.02, 0.10, 0.30, 0.50\}$, and $c \in \{5, 20, 29\}$, giving 90 configurations. With $g = 30 - c$, $d^- = -g\mathbf{1}(D_A = P)$ and $d^+ = 0$. Thus $\mathcal{I} = [-gp, 0]$, although the particular latent contrast generating the observations is zero. The law specifies $R_- = g$, $R_+ = 0$ in advance.

If $K \sim \text{Binomial}(n, p)$ and $q = K/n$, the normal lower endpoint before clipping is $-gq - z_{0.975}g\sqrt{q(1-q)/(n-1)}$; the bounded lower endpoint is $-g(q + h_n)$. Both upper endpoints equal zero under the law-specific support. Each is clipped to $[-g, 0]$. A third calculation uses only the general paired range $[-30, 30]$ for Equation 7. Its extra uncertainty reflects the information discarded by using that broader support.

Performance-dependent interruption law. Half the A tasks would end at 6 seconds and half would complete 30 seconds, so $\mu_A = 18$. B is exactly 18 seconds. A interruption, when assigned, occurs at 2 seconds. Conditional interruption probabilities for short and long tasks are $(p_s, p_l) \in \{(0.30, 0.30), (0.10, 0.50), (0.50, 0.10)\}$. The observable categories F, H, P have probabilities

$$(\pi_F, \pi_H, \pi_P) = \{0.5(1 - p_s), 0.5(1 - p_l), 0.5(p_s + p_l)\}. \quad (11)$$

Their lower differences are $(-12, 12, -16)$ and upper differences $(-12, 12, 12)$ seconds. The ranges specified by this law are $R_- = 28$ and $R_+ = 24$. Enumerate all multinomial counts at $n \in \{20, 50, 100, 200\}$, giving 12 configurations. Each law has $\Delta = 0$ and interruption probability 0.30, but different selection of complete records.

The generating task durations determine the probability evaluation. Estimation uses the observable endpoint supports and measurement-consistency assumptions, leaving post-interruption duration unrestricted. The generating contrast is therefore one member of the identified set; a zero generating contrast can coexist with a nonzero identification width.

For each configuration, calculate whole-set coverage and expected confidence-interval width. For category count vector $\mathbf{k}$, its probability is $n! \prod_j \pi_j^{k_j} / \prod_j k_j!$. Coverage sums those probabilities for which the lower confidence limit is at most L_Δ and the upper is at least U_Δ . Expected width uses the same probabilities. Both methods are clipped to the same law-specific overall range when compared using law-specific support; the general-range bounded interval is evaluated separately.

RESULTS

TWELVE PAIRED DURATIONS AT A 30-SECOND HORIZON

The twelve constructed pairs in Table 3 contain one A withdrawal at 22 seconds and A observation stops at 9 and 11 seconds. B has eleven motor terminations and one horizon completion, with no observation stop. The calculation treats these twelve pairs as a fixed set of records.

TABLE 3. CONSTRUCTED PAIRED DURATIONS AND ADMISSIBLE A TASK SCORES AT A 30-SECOND HORIZON. B SCORES ARE EXACT RESTRICTED SCORES.

Record	X_A (s)	X_B (s)	D_A	D_B	A score limits (s)
1	8	10	F	F	[8, 8]
2	12	13	F	F	[12, 12]
3	16	16	F	F	[16, 16]
4	19	18	F	F	[19, 19]
5	22	20	W	F	[22, 22]
6	25	21	F	F	[25, 25]
7	30	24	H	F	[30, 30]
8	30	25	H	F	[30, 30]
9	30	26	H	F	[30, 30]
10	30	27	H	F	[30, 30]
11	9	28	P	F	[9, 30]
12	11	30	P	H	[11, 30]

A has an observed total of 242 seconds and an admissible upper total of 282 seconds. B totals 258 seconds. Thus, A's task mean lies in [20.17, 23.50] seconds and B's mean equals 21.50 seconds. Subtraction gives the full-record paired bounds [−1.33, 2.00] seconds. After removal of pairs 11 and 12, the remaining A and B totals are 222 and 200 seconds. Their ten-record mean difference is 2.20 seconds, exceeding the full-record upper bound because deletion changes the records entering the mean.

Let r_{11} and r_{12} be the separate continuation fractions for the two P records. Then

$$\bar{\Delta}(r_{11}, r_{12}) = \frac{-16 + 21r_{11} + 19r_{12}}{12} \text{ s.} \quad (12)$$

The zero-difference line in Figure 1 is $21r_{11} + 19r_{12} = 16$. Along the equal-fraction diagonal, the difference becomes zero at $r = 0.40$ and equals 0.33 seconds at $r = 0.50$. The latter assignment gives A scores of 19.50 and 20.50 seconds for the two interrupted records. Positive differences require $21r_{11} + 19r_{12} > 16$. The observations leave both fractions undetermined.

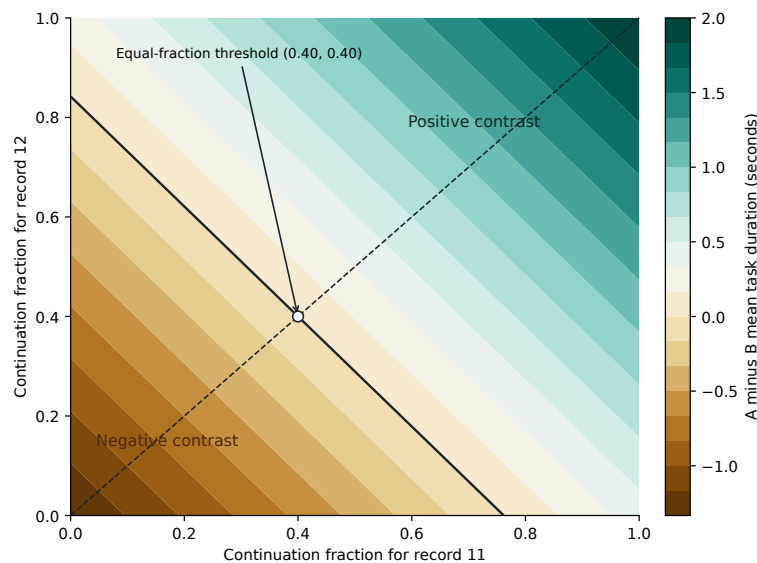


FIGURE 1. PAIRED MEAN DIFFERENCE $(-16 + 21r_{11} + 19r_{12})/12$ SECONDS FOR THE 9- AND 11-SECOND OBSERVATION STOPS. COLOUR GIVES THE DIFFERENCE IN SECONDS. THE SOLID LINE IS $21r_{11} + 19r_{12} = 16$; THE DASHED LINE SETS $r_{11} = r_{12}$. THEIR INTERSECTION AT (0.40, 0.40) IS THE EQUAL-FRACTION ZERO THRESHOLD. EACH FRACTION ASSIGNS A DURATION WITHIN THE CORRESPONDING SCORE INTERVAL.

For A, observed losses assigned to motor termination, withdrawal, and procedure total 70, 8, and 40 seconds, respectively. Their means are 5.83, 0.67, and 3.33 seconds, summing to $30 - 242/12 = 9.83$ seconds. B's motor loss is 102 seconds, or 8.50 seconds per record. A's smaller assigned motor shortfall does not establish greater motor ability because other terminations pre-empt observation of later motor events.

RARE INTERRUPTION AND FINITE-SAMPLE COVERAGE

In the horizon-dominated law with $n = 20$, $p = 0.02$, and $c = 5$, the identified set is $[-0.50, 0]$ seconds. No interruption is observed with probability $0.98^{20} = 0.6676$. On those samples, both endpoint estimates and their estimated standard errors are zero, so the normal interval is $[0, 0]$. It covers the particular generating contrast zero but fails to contain the nondegenerate population identified set. Exact enumeration gives full-set coverage of 33.24%.

Across the 90 binomial configurations, normal-approximation coverage ranges from 9.54% to 97.93% (selected values in Table 4). Coverage of the bounded interval with the stipulated endpoint ranges is 99.64–100.00%. The interval using the general paired range has coverage rounding to 100.00% at every evaluated configuration. All bounded intervals exceed the nominal 95% level within these laws.

TABLE 4. EXACT WHOLE-SET COVERAGE AND EXPECTED INTERVAL WIDTH IN SELECTED HORIZON-DOMINATED LAWS, WITH INTERRUPTION AT 5 SECONDS. BOUNDED INTERVALS USE THE LAW-SPECIFIED RANGES $R_- = 25$, $R_+ = 0$. BOTH INTERVALS ARE CLIPPED TO $[-25, 0]$.

n	p	Coverage (%)		Expected width (s)	
		Normal	Bounded	Normal	Bounded
20	0.005	9.54	100.00	0.36	7.72
20	0.020	33.24	100.00	1.37	8.09
20	0.100	87.84	100.00	5.50	10.09
50	0.020	63.58	100.00	1.25	5.30
200	0.020	91.06	100.00	0.97	2.90
1000	0.020	96.24	100.00	0.72	1.57

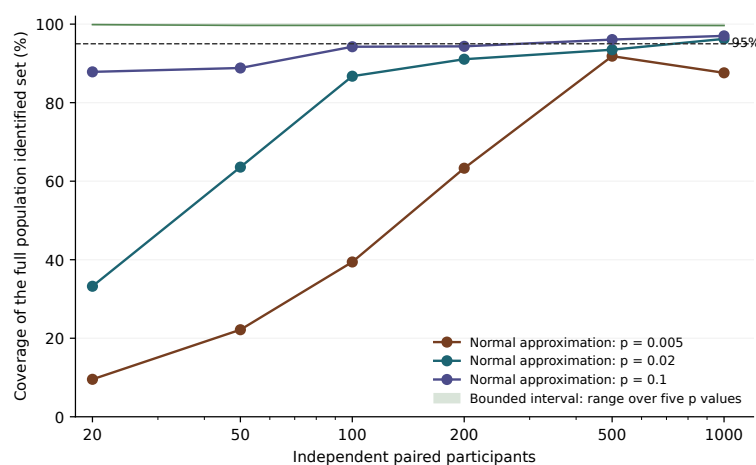


FIGURE 2. WHOLE-SET COVERAGE FROM BINOMIAL COUNT ENUMERATION WITH A 5-SECOND INTERRUPTION AT A 30-SECOND HORIZON. THE THREE CURVES GIVE NORMAL-APPROXIMATION COVERAGE FOR $p = 0.005$, 0.02 , AND 0.10 . THE GREEN BAND SPANS BOUNDED-INTERVAL COVERAGE ACROSS $p = 0.005$, 0.02 , 0.10 , 0.30 , AND 0.50 , USING $R_- = 25$ AND $R_+ = 0$. THE DASHED LINE MARKS 95%; SAMPLE SIZE IS SHOWN ON A LOGARITHMIC AXIS.

At $n = 20$, $p = 0.02$, and $c = 5$, expected interval width is 8.09 seconds with the stipulated endpoint ranges and 36.94 seconds with $[-30, 30]$ for both endpoints. The normal interval has width 1.37 seconds and coverage 33.24%. At $n = 1000$, the respective widths are 1.57, 5.65, and 0.72 seconds. The two bounded widths differ because the general range allows endpoint values excluded by the stipulated observation law.

PERFORMANCE DEPENDENCE AND COMPLETE-CASE LIMITS

The three performance-dependent laws all have an underlying contrast of zero and population identification width 8.40 seconds. Their identified sets differ because interruption selectively removes short or long tasks (Table 5). Deletion converges to a complete-case contrast of -3.43 or 3.43 seconds when interruption preferentially affects long or short tasks, respectively. Increasing sample size does not remove that selection shift.

TABLE 5. PERFORMANCE-DEPENDENT INTERRUPTION AT 2 SECONDS. ALL LAWS HAVE $\Delta = 0$, $\pi_P = 0.30$, AND B DURATION 18 SECONDS. COVERAGE IS EXACT AT $n = 100$; THE BOUNDED INTERVAL USES THE LAW-SPECIFIED RANGES.

p_s	p_t	Identified set (s)	Complete-case	Full-set coverage (%)	
			limit (s)	Normal	Bounded
0.30	0.30	$[-4.80, 3.60]$	0.00	95.29	99.59
0.10	0.50	$[-7.20, 1.20]$	-3.43	95.30	99.60
0.50	0.10	$[-2.40, 6.00]$	3.43	95.81	99.68

Across the twelve multinomial configurations, normal coverage is 95.25–96.20% and bounded coverage is 99.59–99.76%. The normal interval therefore exceeds 95% in these configurations. Its low coverage in the binomial calculations is associated with rare interruption and nearly complete horizon performance. Equation 8 provides the same finite-sample guarantee for both distribution families.

DISCUSSION

A duration endpoint restricted to 30 seconds and eventual time to balance loss answer different questions. A completed 30-second record fixes the first endpoint and right-censors the second. Raising the horizon changes the test duration and may alter fatigue, discomfort, and withdrawal. The observation horizon must therefore follow the intended task outcome.

Interruption timing determines identification width directly. At a 30-second horizon, a verified 29-second observation stop leaves one second unresolved; a 5-second stop leaves 25 seconds. Equation 5 combines interruption probability with the mean remaining duration. Counting interrupted attempts alone omits this timing contribution. Resolving the first cause and obtaining a later verified stance time can narrow the interval; adding participants does not recover the missing duration within an existing record.

Termination codes separate motor failure, withdrawal, observation stop, and horizon completion. Their usefulness depends on the evidence supporting each code. A synchronized-video study can measure examiner differences in elapsed seconds, code agreement, and the resulting changes in mean endpoints. For complete repeated-session scores, intraclass correlation analysis must identify its statistical model, measurement unit, and agreement definition (Koo and Li, 2016). Agreement measured only among complete records leaves interrupted-record classification untested.

The rare-interruption calculation isolates a failure of estimated-variance inference. With $n = 20$ and $p = 0.02$, the probability of observing no interruption is 66.76%. A zero observed variance on those samples makes the normal interval a point, even though the population lower endpoint is -0.50 seconds. The bounded interval retains a sampling allowance determined by n , α , and the known endpoint range. Its protection is conservative: the minimum coverage across all 102 configurations is 99.59%, and its width exceeds that of the normal interval in the tabulated binomial cases.

The support range must be justified before the sample is observed. In the binomial law, $d^+ = 0$ is known from the stipulated observation distribution, so its range is zero. Identical observed upper differences in an ordinary stance sample do not establish the same population restriction. Without design-based restrictions, both difference endpoints use $[-\tau, \tau]$. The confidence statement also requires fixed n and independent participant units; repeated inspection or clustering needs a corresponding inferential treatment.

Observation completeness and sampling precision require separate design calculations. For a permitted unresolved mean duration w_0 , the population condition is $E(u - \ell) \leq w_0$. For a common support range R and a maximum sampling allowance η at each confidence endpoint, Equation 7 gives

$$n \geq \frac{R^2 \log(2/\alpha)}{2\eta^2}. \quad (13)$$

At $\tau = 30$ seconds, the general paired range has width $R = 60$ seconds. An allowance of $\eta = 3$ seconds at 95% confidence requires $n \geq 738$. This value follows from the conservative concentration inequality. It is a precision calculation under that inequality and provides no empirical estimate of a clinical effect size.

Deletion changes the duration target when interruption depends on performance. The twelve-record calculation gives 2.20 seconds after deletion, compared with a full-record upper bound of 2.00 seconds. In the probability laws, preferential interruption of long or short tasks gives complete-case limits of -3.43 or 3.43 seconds while the underlying contrast remains zero. A continuation model can produce a point estimate, but its restrictions on unobserved duration must be stated and defended separately.

The derivation depends on preservation of the task before an observation stop. A perturbation that changes stance violates that interpretation even if its time is known exactly. The evaluated laws use discrete durations and constant B scores; they test confidence coverage under specified distributions. The analysis does not estimate fall risk, clinical

importance, examiner agreement, or responsiveness. The confidence interval guarantees coverage under its assumptions but is not asserted to have minimum width among valid procedures.

CONCLUSION

Verified horizon completion fixes restricted single-leg stance duration. Observation interruption leaves an interval whose width is the remaining unobserved task time. Termination codes and timing evidence determine sharp mean and paired-difference endpoints; a bounded-mean interval adds finite-sample population coverage when participant independence, fixed sample size, and prespecified ranges hold. Exact enumeration demonstrates coverage loss for the normal approximation under rare interruption and quantifies the larger widths needed for range-based protection. Application to participant measurements requires validated event classification and a measurement procedure that preserves the defined stance task.

DECLARATIONS

AUTHOR CONTRIBUTIONS

All authors contributed equally and all authors read and approved the final version of this manuscript.

CONFLICT OF INTEREST

The authors declare 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 corresponding author upon reasonable request.

FUNDING

This research received no external funding.

REFERENCES

1. Choi YM, Dobson F, Martin J, Bennell KL, Hinman RS. Interrater and intrarater reliability of common clinical standing balance tests for people with hip osteoarthritis. *Phys Ther.* 2014;94(5):696–704.
2. Cole SR, Zivich PN, Edwards JK, Shook-Sa BE, Hudgens MG. Sensitivity analyses for means or proportions with missing outcome data. *Epidemiology.* 2023;34(5):645–651.
3. Condon C, Cremin K. Static balance norms in children. *Physiother Res Int.* 2014;19(1):1–7.
4. Conner SC, Trinquart L. Estimation and modeling of the restricted mean time lost in the presence of competing risks. *Stat Med.* 2021;40(9):2177–2196.
5. Genbäck M, Stanghellini E, de Luna X. Uncertainty intervals for regression parameters with non-ignorable missingness in the outcome. *Stat Pap.* 2015;56:829–847.
6. Howard SR, Ramdas A, McAuliffe J, Sekhon J. Time-uniform, nonparametric, nonasymptotic confidence sequences. *Ann Stat.* 2021;49(2):1055–1080.
7. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. E9(R1). 2019. ICH E9(R1) guideline.
8. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15(2):155–163.
9. Royston P, Parmar MKB. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Stat Med.* 2011;30(19):2409–2421.
10. Royston P, Parmar MKB. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol.* 2013;13:152.

11. Shen X, Mak MKY. Balance and gait training with augmented feedback improves balance confidence in people with Parkinson's disease: a randomized controlled trial. *Neurorehabil Neural Repair*. 2014;28(6):524–535.
12. Stoye J. A simple, short, but never-empty confidence interval for partially identified parameters. *CeMMAP Working Paper CWP23/21*. 2021.
13. White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. *BMJ*. 2011;342:d40.
14. Zhang C, Wu Y, Yin G. Restricted mean survival time for interval-censored data. *Stat Med*. 2020;39(26):3879–3895.