

ACUTE CONDITION CONTRASTS UNDER SHARED TEMPORAL DRIFT: POLYNOMIAL CANCELLATION, PRECISION, AND TIMING

Shin Min Young

Gyeongsang National University, Jinju, South Korea

Received: 09 January 2024 Revised: 25 April 2024 Accepted: 11 May 2024 Published: 20 June 2024

Correspondence: smjyoung@gnu.ac.kr

ABSTRACT

To determine when three neutral attempts improve estimation of an acute A-B motor-performance difference during a seven-attempt session. The schedules were R-A-B-R-B-A-R and R-B-A-R-A-B-R, where R is the same movement under a neutral instruction. The contrast subtracts one sixth of the signed reference second difference from the mirrored A-B comparison. Polynomial cancellation, measurement variance, carryover bias, and weights for unequal onset times were derived analytically. Ten drift and error conditions each used 20,000 simulated studies of 40 independently generated individuals. The timing analysis used 50,000 grids at each of five jitter levels. No participants were recruited. The fixed contrast removed shared polynomial drift through degree three and equalled the minimum-variance linear unbiased regression contrast under equal independent error variances. With condition and reference error standard deviations of 1 cm, quadratic-drift root mean squared error decreased from 1.449 to 1.080 cm; with flat drift it increased from 1.000 to 1.080 cm. A fatigue step increased error from 0.999 to 1.269 cm, and reference error of 3 cm reversed the benefit under quadratic drift. Differential carryover remained biased, with corrected population-mean interval coverage of 29.40%. Timestamp-aware weights restored polynomial cancellation on irregular grids but increased measurement variance. For shared cubic drift and independent errors, correction reduces individual mean squared error when the absolute drift bias exceeds the reference-error standard deviation divided by $\sqrt{6}$. Equal task definitions, qualified recovery intervals, recorded onset times, and complete cycles are required. Differential carryover and changing condition effects remain outside the constant-offset model.

Keywords: kinesiology; repeated movement assessment; crossover design; polynomial drift; within-participant contrast; measurement error.

INTRODUCTION

Repeated movement scores can change with familiarization, motor calibration, attention, recovery, and fatigue during a single session. Studies of the Star Excursion Balance Test and Lower Quarter Y-Balance Test document variation associated with measurement procedures, trial number, and learning (Powden et al., 2019; Linek et al., 2017; Kattilakoski et al., 2023; Foldager et al., 2023). When two conditions occupy different trial positions, their score difference can therefore contain both a condition effect and a temporal effect.

Acute instruction comparisons make this distinction especially relevant: an attentional cue may affect both immediate performance and learning (Wulf, 2013). A mirrored A-B-B-A arrangement cancels a linear position trend, but retains quadratic and cubic drift. Placing neutral attempts before, between, and after the four condition attempts yields R-A-B-R-B-A-R. The three R scores provide a second difference that can remove the remaining polynomial contribution when R, A, and B share the same background trajectory.

Acquisition sequences and drift correction have long been used to control uncertainty in comparative measurement (Beatrice et al., 2015). Crossover methods address treatment persistence, sequence, period, and washout (Wellek and Blettner, 2012; Dwan et al., 2019); switchback methods address repeated treatment changes with carryover (Bojinov et al., 2023). In a seven-attempt motor session, these issues enter through the neutral-reference second difference, the error covariance, the recovery interval, and the constancy of the A and B offsets. The resulting contrast can be expressed both as a weighted score difference and as a cubic-trend regression coefficient difference.

The analysis derives the seven-position weights, identifies the reference-error threshold for reducing mean squared error, and calculates the bias left by carryover and reference disturbances. Simulated scores then quantify individual estimation error and population-mean interval coverage for ten explicit drift and error specifications. A separate unequal-timing calculation tests weights derived from actual onset times.

DESIGN, ESTIMAND, AND OBSERVATION MODEL

ELIGIBLE COMPARISON AND TRIAL ARRANGEMENT

Let A and B be acute conditions applied to the same movement, and let R be that movement performed under a prespecified neutral instruction. All attempts produce one scalar response in the same unit. The participant-level target is

$$\Delta_i = \tau_{iA} - \tau_{iB}, \quad \mu = \mathbb{E}(\Delta_i), \quad (1)$$

where the immediate condition offsets are constant during a qualified cycle. A positive difference denotes a larger A score; its clinical or performance interpretation depends on the outcome. The population target requires an explicitly defined sampling population.

After familiarization and qualification, assign a participant with probability one half to

$$\begin{array}{l} s_i = +1 : \text{R A B R B A R,} \\ s_i = -1 : \text{R B A R A B R.} \end{array} \quad (2)$$

The nominal positions are $t = 0, \dots, 6$. For an elapsed-time model, equal spacing requires onset times of the form $\xi_t = \xi_0 + dt$ with constant $d > 0$. The schedule contains two A attempts, two B attempts, and three R attempts. Numbering seven attempts consecutively is insufficient if their onset intervals differ.

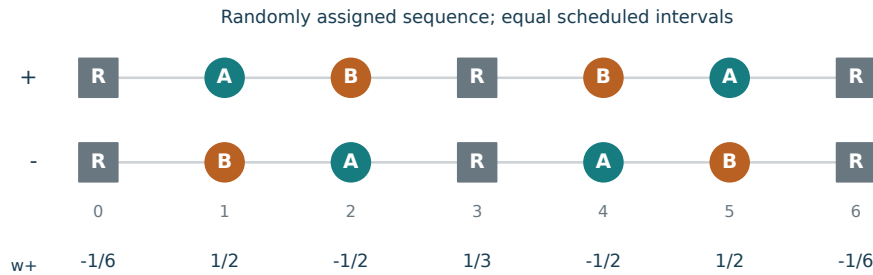


FIGURE 1. SEVEN-ATTEMPT SEQUENCES AND CONTRAST WEIGHTS. R IS A NEUTRAL ATTEMPT OF THE SAME TASK. THE R-B-A-R-A-B-R SEQUENCE USES THE NEGATIVE OF THE DISPLAYED WEIGHTS. POSITIONS 0, ..., 6 REQUIRE EQUAL INTERVALS IN THE SELECTED TIME OR EXPOSURE VARIABLE.

Every R attempt adds one movement to the exposure history. A and B must preserve the movement and scoring definition used for R, and their immediate effects must dissipate before subsequent scored attempts. Persistent interventions or unequal recovery violate these requirements. Reliability estimates for an existing balance-test score apply to that score and testing procedure; the A-B instruction difference requires its own assessment (Shaffer et al., 2013).

SHARED-DRIFT ASSUMPTION

For neutral and manipulated attempts, respectively, use

$$R_i(t) = a_i + g_i(t) + \tau_{iR} + u_i(t), \quad t \in \{0, 3, 6\}, \quad (3)$$

$$Y_i(t) = a_i + g_i(t) + \tau_{ic(t)} + e_i(t), \quad t \in \{1, 2, 4, 5\}. \quad (4)$$

The errors have zero conditional means given the sequence and underlying trajectory. They include trial-to-trial execution and scoring variation, rather than instrument error alone. The function g_i is shared by R, A, and B outside their immediate offsets. The model excludes differential after-effects and condition-dependent learning. Those exclusions are substantive requirements to be examined in task qualification, not consequences of the estimator.

Clock time and cumulative movement exposure need not describe the same biological process. Investigators must choose the drift variable before analysis. Reweighting by timestamps is appropriate for a polynomial trajectory in elapsed time; it does not correct a different, condition-dependent workload trajectory.

ANALYTICAL PROPERTIES

FIXED CONTRAST AND POLYNOMIAL CANCELLATION

Define the raw mirrored contrast and neutral-reference second difference by

$$C_i = \frac{s_i}{2} \{Y_i(1) - Y_i(2) - Y_i(4) + Y_i(5)\}, \quad (5)$$

$$Q_i = R_i(0) - 2R_i(3) + R_i(6). \quad (6)$$

The reference-bracketed estimator is

$$\hat{\Delta}_i = C_i - \frac{s_i Q_i}{6}. \quad (7)$$

The coefficient $1/6$ follows from the fixed grid. Linear interpolation inside the first bracket gives $L_i(1) = \{2R_i(0) + R_i(3)\}/3$ and $L_i(2) = \{R_i(0) + 2R_i(3)\}/3$. The second bracket gives $L_i(4) = \{2R_i(3) + R_i(6)\}/3$ and $L_i(5) = \{R_i(3) + 2R_i(6)\}/3$. Averaging the two order-signed differences after subtracting these interpolations gives Eq. (7). Interpolated values need not individually equal the latent drift for the combined contrast to cancel it.

Proposition 1 (Fixed-grid cancellation). *Under Equations (3)–(4), equally spaced attempts, and constant condition offsets, $\mathbb{E}(\hat{\Delta}_i \mid s_i, g_i, \Delta_i) = \Delta_i$ whenever g_i is a polynomial of degree at most three.*

Proof. In chronological order, the positive-sequence weight vector is

$$w_+ = (-1/6, 1/2, -1/2, 1/3, -1/2, 1/2, -1/6)^T. \quad (8)$$

It satisfies $\sum_{t=0}^6 w_{+t} t^k = 0$ for $k = 0, 1, 2, 3$. Its A weights sum to one, its B weights sum to minus one, and its R weights sum to zero. The negative sequence changes the weight sign and interchanges A and B, preserving the treatment contrast. Linearity and zero conditional error means establish the result. $\square$

TABLE 1. DETERMINISTIC DRIFT CONTRIBUTIONS FOR THE POSITIVE SEQUENCE. EACH ENTRY MULTIPLIES THE COEFFICIENT OF THE INDICATED POLYNOMIAL BASIS.

Basis	Raw contrast	Reference correction	Residual
1	0	0	0
t	0	0	0
t^2	3	3	0
t^3	27	27	0
t^4	177	189	-12

For $g_i(t) = \kappa_{4i} t^4$, residual bias is $-12s_i \kappa_{4i}$. General nonlinear change is therefore not removed. The second difference Q_i also mixes underlying curvature with reference noise and any reference-specific disturbance; it cannot independently establish the cause of temporal change.

EQUIVALENCE TO A REGRESSION CONTRAST

For the positive sequence, let X have columns $1, t, t^2, t^3, I_A, I_B$, where I_A and I_B identify the assigned conditions. The intercept incorporates $a_i + \tau_{iR}$; the two indicator coefficients are $\tau_{iA} - \tau_{iR}$ and $\tau_{iB} - \tau_{iR}$. With $c = (0, 0, 0, 0, 1, -1)^T$, the target is $c^T \theta_i$.

The matrix has rank six. The fixed weights satisfy $X^T w_+ = c$, so they define an unbiased linear contrast in this model. The null space of X^T is generated by $v = (1, -4, 5, 0, -5, 4, -1)^T$. Every unbiased weight vector is $w_+ + av$ for a scalar a . Since $w_+^T v = 0$,

$$\|w_+ + av\|^2 = \|w_+\|^2 + a^2 \|v\|^2. \quad (9)$$

Thus w_+ minimizes variance when the error covariance is $\sigma^2 I$. Equivalently,

$$\hat{\Delta}_i = c^T (X^T X)^{-1} X^T y_i = w_+^T y_i. \quad (10)$$

The reference correction is therefore exactly the least-squares A-B contrast for this design matrix. Seven observations and six regression parameters leave one residual degree of freedom. This limits checks of polynomial fit and prevents estimation of an unrestricted within-cycle error covariance from a single cycle.

PRECISION AND THE DECISION TO CORRECT

With independent errors, condition-attempt variance σ_T^2 , and reference-attempt variance σ_R^2 , the measurement-error variances are

$$\begin{aligned}\text{Var}(C_i | g_i, \Delta_i, s_i) &= \sigma_T^2, \\ \text{Var}(\hat{\Delta}_i | g_i, \Delta_i, s_i) &= \sigma_T^2 + \sigma_R^2/6.\end{aligned}\quad (11)$$

When both error variances equal σ^2 , correction adds 16.7% to measurement variance and increases its standard deviation by $\sqrt{7/6} - 1 = 8.0\%$. These are conditional measurement calculations; the variation of completed contrasts across individuals can also include true condition-effect heterogeneity and residual drift.

Let b_i be the raw contrast's drift bias under shared cubic drift. The same seven observations then yield

$$\text{MSE}(C_i) = \sigma_T^2 + b_i^2, \quad \text{MSE}(\hat{\Delta}_i) = \sigma_T^2 + \sigma_R^2/6. \quad (12)$$

Correction improves individual mean squared error exactly when

$$|b_i| > \sigma_R/\sqrt{6}. \quad (13)$$

This condition uses the underlying drift bias, not the noisy observed correction $Q_i/6$. Switching estimators after seeing a large Q_i changes the analysis and requires its own evaluation. The estimator and any switching rule should therefore be prespecified.

For a general covariance matrix Σ_i , replace the independent-error expression by $w_i^T \Sigma_i w_i$, with $w_i = s_i w_+$. If higher-order drift leaves corrected bias b_i^* , the improvement condition becomes $b_i^{*2} - (b_i^*)^2 > w_i^T \Sigma_i w_i - r_i^T \Sigma_i r_i$, where r_i is the signed raw weight vector. Reliability and agreement statistics address different aspects of repeated measurement and do not establish absence of bias in the condition contrast (Koo and Li, 2016; Giavarina, 2015).

INDIVIDUAL CORRECTION AND POPULATION RANDOMIZATION

If allocation is independent of a participant's trajectory and does not change it, both estimators have the form $\Delta_i + s_i D_i$ + error for an allocation-independent drift term D_i . Since $\mathbb{E}(s_i) = 0$, the raw and corrected contrasts are both unbiased for μ over random assignments, even for non-polynomial drift. This protection is not exact cancellation of individual drift, nor does it require every realized sample to contain equal sequence counts.

Reference correction can reduce individual error and the drift-related variance of a population estimate. It does not uniquely confer population unbiasedness. If treatment history changes the trajectory, assignment independence fails and the randomization argument no longer removes that history effect. This distinction is consistent with the central role of persistence in crossover and switchback designs (Wellek and Blettner, 2012; Bojinov et al., 2023).

IRREGULAR TIMING AND ESTIMABLE CONTRASTS

Let the actual onset times be $\xi_{i0}, \dots, \xi_{i6}$ and form X_i using $1, x, x^2, x^3, I_A, I_B$, where x is a prespecified affine scaling of elapsed time. Provided that X_i has full column rank and Σ_i is positive definite, the minimum-variance linear unbiased weights are

$$w_i^{\text{time}} = \Sigma_i^{-1} X_i (X_i^T \Sigma_i^{-1} X_i)^{-1} c. \quad (14)$$

They satisfy $X_i^T w_i^{\text{time}} = c$ and therefore remove the specified cubic temporal drift on the actual grid. Equal independent errors give the ordinary least-squares special case. Record onset times independently of scores, and select the time variable and polynomial degree before examining the A-B difference.

Timestamp weights cancel clock-time polynomial drift but retain any bias caused by altered recovery or carryover. Unequal intervals also change measurement variance and the conditioning of X_i . The onset times, weight vector, matrix rank, and condition number for a stated time scaling must accompany the contrast. A calculable contrast does not make an unqualified recovery interval acceptable.

COMPUTATIONAL EVALUATION

DETERMINISTIC REACH CALCULATION

For the deterministic reach calculation, the neutral trajectory and condition offsets are specified in *centimetres* as

$$\begin{cases} g(t) = 72 + 0.80t + 0.35t^2, \\ \tau_A = 3.00, \\ \tau_B = -2.00, \\ \tau_R = 0. \end{cases} \quad (15)$$

For R-A-B-R-B-A-R with zero measurement error, the seven scores are 72.00, 76.15, 73.00, 77.55, 78.80, 87.75, and 89.40 cm (Table 2). They give $C = 6.05$ cm and $Q = 6.30$ cm. Subtracting $Q/6 = 1.05$ cm gives $\hat{\Delta} = 5.00$ cm, equal to $\tau_A - \tau_B = 3.00 - (-2.00)$. The values follow directly from the specified function and offsets; they are not participant measurements.

TABLE 2. DETERMINISTIC REACH SCORES FOR $g(t) = 72 + 0.80t + 0.35t^2$, A OFFSET 3.00 CM, AND B OFFSET -2.00 CM. ALL RESPONSE COLUMNS ARE IN CENTIMETRES; t IS A DIMENSIONLESS SCHEDULED POSITION.

t	Trial	$g(t)$	Offset	Score	$L(t)$	Score $- L(t)$
0	R	72.00	0.00	72.00	—	—
1	A	73.15	3.00	76.15	73.85	2.30
2	B	75.00	-2.00	73.00	75.70	-2.70
3	R	77.55	0.00	77.55	—	—
4	B	80.80	-2.00	78.80	81.50	-2.70
5	A	84.75	3.00	87.75	85.45	2.30
6	R	89.40	0.00	89.40	—	—

SIMULATED STUDIES AND OPERATING CONDITIONS

For each condition in Table 3, 20,000 independent studies of $n = 40$ computational individuals were generated. The random generator was NumPy PCG64, with initial seed $20261010 + j$ for condition index $j = 0, \dots, 9$. Each individual had independently sampled sequence sign $s_i \in \{-1, +1\}$ and immediate offset $\Delta_i \sim N(5, 1.5^2)$ cm. The A contribution was Δ_i and the B and R contributions were zero. Scores were $72 + g(t) + I_A \Delta_i + \varepsilon_i$, with the additions specified below. Changing the constant B offset does not change the condition contrast.

TABLE 3. TEN DRIFT AND ERROR SPECIFICATIONS FOR SIMULATED SCORES. ADDITIVE RESPONSE QUANTITIES ARE IN CENTIMETRES AND $t = 0, \dots, 6$. ERRORS ARE INDEPENDENT WITH STANDARD DEVIATION 1 CM AT EVERY POSITION UNLESS A ROW STATES OTHERWISE. THE DRIFT FUNCTIONS ARE FIXED MATHEMATICAL INPUTS.

Condition	Drift and additional specification
Flat	$g(t) = 0$.
Quadratic	$g(t) = 0.35(t - 3)^2$.
Cubic	$g(t) = 0.20t + 0.20t^2 + 0.015t^3$.
Saturating	$g(t) = 4\{1 - \exp(-t/1.5)\}$.
Fatigue step	$g(t) = -4I(t \geq 4)$.
Noisy references	Quadratic drift; reference SD 3 cm, condition SD 1 cm.
Serial error	Quadratic drift; $\Sigma_{uv} = 0.6^{ u-v }$ cm ² .
Carryover	Quadratic drift; add 1.5 cm to an A/B attempt immediately following A, with neutral attempts unaffected.
Middle reference	Quadratic drift; add 3 cm to the middle R score.
Changing A effect	Quadratic drift; additionally add $0.4t$ to each A score.

Both estimators used the same seven scores, onset positions, and error realizations. For each generated individual, estimation error was the contrast minus Δ_i . Root mean squared error was calculated over 800,000 generated individuals per condition, and conditional bias was calculated separately for each sequence sign. The comparison holds session length and movement exposure fixed.

For each simulated study, the population-mean interval was $\bar{D} \pm t_{0.975, 39} s_D / \sqrt{40}$, calculated from 40 individual contrasts. Coverage was the fraction of 20,000 intervals containing the generating mean of 5 cm. The between-individual standard deviation was fixed at 1.5 cm; the drift functions and measurement-error magnitudes are given in Table 3. These inputs were not estimated from participants. In the changing-A condition, the added $0.4t$ cm response violates the constant-offset estimand. Error and coverage retain the nominal offset as the target, excluding the added time-dependent A response.

The largest Monte Carlo standard error for an individual RMSE estimate was 0.0013 cm. Coverage standard error is $\sqrt{\hat{p}(1-\hat{p})/20000}$, approximately 0.15 percentage points near 95% and at most 0.35 percentage points. Full generation and analysis code is embedded in Appendix .

COMPARATIVE RESULTS

TABLE 4. SIMULATED ESTIMATION ERROR AND POPULATION-MEAN INTERVAL COVERAGE. BIAS AND RMSE ARE IN CENTIMETRES. BIAS IS CONDITIONAL ON THE POSITIVE SEQUENCE; RMSE INCLUDES BOTH SEQUENCES. COVERAGE IS A PERCENTAGE OVER 20,000 STUDIES OF 40 GENERATED INDIVIDUALS. FOR THE CHANGING-A CONDITION, THE NOMINAL CONSTANT OFFSET IS THE REFERENCE TARGET.

Condition	Bias, $s = +1$		Individual RMSE		95% coverage	
	Raw	Corrected	Raw	Corrected	Raw	Corrected
Flat	0.000	0.000	1.000	1.080	95.12	95.03
Quadratic	1.051	0.002	1.449	1.080	95.01	94.88
Cubic	1.004	-0.001	1.419	1.082	94.92	95.03
Saturating	-0.430	0.068	1.090	1.083	95.22	94.96
Fatigue step	0.001	0.669	0.999	1.269	94.93	94.88
Noisy references	1.048	-0.002	1.449	1.579	95.12	95.30
Serial error	1.049	-0.001	1.238	0.573	94.95	95.12
Carryover	0.301	-0.748	1.632	1.314	40.16	29.40
Middle reference	1.048	0.999	1.448	1.470	95.25	95.19
Changing A effect	2.251	1.202	1.884	1.616	5.46	2.01

Under quadratic drift, correction decreased individual RMSE by 25.5%, from 1.449 to 1.080 cm. Under cubic drift it decreased RMSE from 1.419 to 1.082 cm; under saturating drift the change was 1.090 to 1.083 cm. With flat drift, reference noise increased RMSE from 1.000 to 1.080 cm. Increasing reference-error SD to 3 cm under quadratic drift raised corrected RMSE to 1.579 cm, compared with 1.449 cm for the raw contrast, consistent with Eq. (13).

For the fatigue step $g(t) = -4I(t \geq 4)$, the raw drift contribution is zero and the reference second difference is -4 cm. Correction therefore adds $2s_i/3$ cm and increases RMSE from 0.999 to 1.269 cm. Adding 3 cm to the middle R score adds exactly s_i cm to the corrected contrast. Three reference positions alone cannot establish smooth drift or distinguish curvature from a reference-specific disturbance.

Positive serial error reduced the variance of this alternating contrast. For the specified AR(1) covariance, the corrected measurement variance is 0.32734 cm^2 , giving a standard deviation of 0.572 cm. This is a property of the particular covariance pattern, not a general benefit of autocorrelation. A different covariance structure can change the direction of the precision effect.

Population-mean coverage ranged from 94.88% to 95.30% for allocation-independent drift and reference disturbances. Random sequence assignment preserved mean unbiasedness for both estimators despite their different individual errors. The specified carryover mechanism has a -0.75 -cm population bias: coverage fell to 40.16% for the raw contrast and 29.40% for the corrected contrast. Correction reduced the polynomial contribution without removing the carryover bias. Adding $0.4t$ cm to A scores reduced coverage for the nominal constant offset to 5.46% and 2.01%, respectively.

TIMING ANALYSIS

At each jitter level $a \in \{0, 0.05, 0.10, 0.20, 0.40\}$, 50,000 grids were generated with endpoints fixed at 0 and 6 and independent internal deviations $U(-a, a)$ from nominal positions. A level of $a = 0.40$ permits each internal onset to deviate by up to 40% of the nominal interval. The drift was $g(t) = 0.20t + 0.20t^2 + 0.015t^3$ cm and the measurement-error covariance was the identity matrix in cm^2 . Expected squared error on each grid was the squared deterministic residual plus analytical measurement variance. The grid average was then square-rooted to obtain expected RMSE. These jitter levels quantify numerical sensitivity; recovery tolerance requires task-specific assessment.

At $a = 0.40$, fixed-weight expected RMSE was 1.201 cm, compared with 1.091 cm for timestamp-aware weights. Polynomial residuals for the timestamp-aware contrasts remained below 8.5×10^{-15} cm. From $a = 0$ to $a = 0.40$, the 95th percentile of timestamp-weight variance rose from 1.167 to 1.355 cm^2 , and the 95th percentile of the scaled design-matrix condition number rose from 7.32 to 8.74. Actual-time weights thus removed the polynomial residual while retaining a timing-dependent precision cost. The calculation contains no physiological model of recovery.

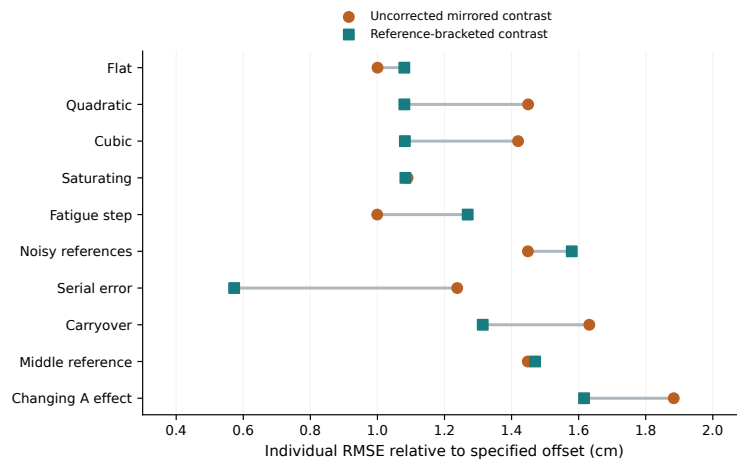


FIGURE 2. INDIVIDUAL RMSE FROM 800,000 GENERATED SEVEN-SCORE CYCLES PER DRIFT AND ERROR SPECIFICATION. THE CHANGING-A ROW USES THE CONSTANT OFFSET Δ_i AS THE TARGET AND EXCLUDES THE ADDED $0.4t$ CM A RESPONSE. MONTE CARLO STANDARD ERRORS ARE AT MOST 0.0013 CM.

TABLE 5. TIMING-GRID SENSITIVITY UNDER CUBIC ELAPSED-TIME DRIFT. RMSE IS IN CENTIMETRES AND VARIANCE IN CM^2 . THE CONDITION NUMBER USES THE DESIGN MATRIX WITH TIME SCALED AS $(\xi - 3)/3$. RESULTS AVERAGE 50,000 GENERATED GRIDS PER JITTER LEVEL.

Jitter fraction	Fixed-weight RMSE	Timestamp-aware RMSE	Timestamp variance 95th percentile	Condition number 95th percentile
0.00	1.080	1.080	1.167	7.32
0.05	1.082	1.080	1.184	7.43
0.10	1.088	1.081	1.203	7.55
0.20	1.111	1.083	1.245	7.85
0.40	1.201	1.091	1.355	8.74

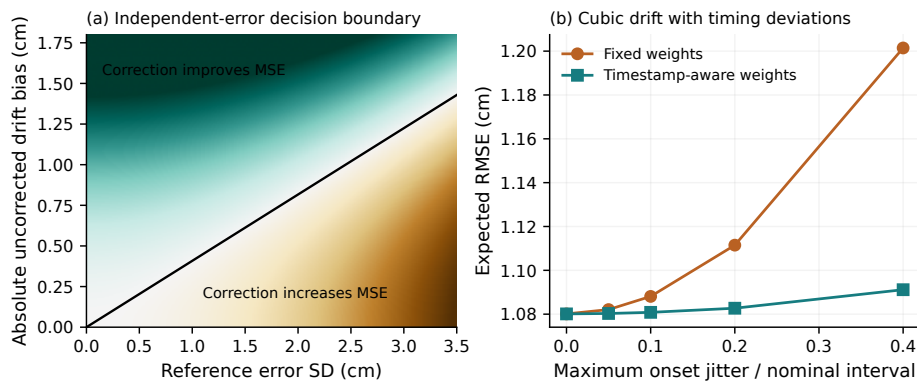


FIGURE 3. REFERENCE-ERROR THRESHOLD AND ONSET-TIME SENSITIVITY. (A) COLOUR GIVES RAW MINUS CORRECTED MSE UNDER SHARED CUBIC DRIFT AND INDEPENDENT ERRORS; $|b| = \sigma_R/\sqrt{6}$ IS THE ZERO-DIFFERENCE BOUNDARY. (B) EXPECTED RMSE ON 50,000 GRIDS PER JITTER LEVEL FOR $g(t) = 0.20t + 0.20t^2 + 0.015t^3$ CM AND IDENTITY ERROR COVARIANCE.

MOTOR-TASK QUALIFICATION AND ANALYSIS RULES

TASK AND RECOVERY QUALIFICATION

A neutral reference must retain the same stance, target direction, footwear, score definition, and validity criteria as A and B. For an anterior reach, the protocol must state the stance limb, limb-length normalization policy, displacement rule, and assessor procedure. Y-Balance studies document trial-number and practice effects, so familiarization requirements depend on the task and population (Linek et al., 2017; Kattilakoski et al., 2023; Foldager et al., 2023). Familiarization ends according to a prespecified stability criterion; it cannot continue until the A–B difference becomes favourable.

Task qualification must distinguish an immediate offset from a change in subsequent learning or readiness. Neutral-only repetitions measure background variation. Neutral performance after A and B at prespecified recovery intervals measures condition-specific after-effects. Under the first-order mechanism in Eq. (16), a maximum acceptable contrast bias ϵ requires $|\rho_B - \rho_A| \leq 2\epsilon$. Qualification therefore concerns the size of the residual effect, rather than whether a

carryover test reaches statistical significance.

For onset interval d and attempt duration ℓ_{it} , the rest before the next onset is $d - \ell_{it}$. Equal onsets thus give unequal rest if durations differ. Record both onset and completion times and require every resulting rest interval to satisfy the task's recovery criterion. The analysis establishes no numerical recovery duration. When completion time is the outcome, the duration and recovery process must be included in the observation model or controlled by a different acquisition design.

CARRYOVER AND REFERENCE DISTURBANCE

Suppose a manipulated attempt immediately after A receives an additive after-effect ρ_A , and one immediately after B receives ρ_B , while R attempts remain uncontaminated. Substitution into the contrast gives, for either order,

$$\text{Bias}_{\text{carryover}} = (\rho_B - \rho_A)/2. \quad (16)$$

If neutral references also retain an after-effect, their contributions enter through Q_i and Eq. (16) is insufficient. Longer rest, a longer design, or a different estimand is then needed. General treatment persistence is not removed by a polynomial correction.

If the middle R score alone is raised by h , the corrected contrast changes by $s_i h/3$. Duplicate anchors might reduce reference uncertainty, but they also change movement dose and spacing. They therefore require a separately defined acquisition and contrast rather than being silently added to a seven-attempt cycle.

INVALID ATTEMPTS, MISSINGNESS, AND DEVIATIONS

An invalid scheduled attempt is recorded as missing with its reason and timestamp. Inserting another attempt changes exposure history and invalidates use of the original seven-position weights. Prespecified stopping events include pain, a fall or near fall, and inability to maintain the standardized movement.

The primary fixed-grid contrast requires all seven eligible scores. The analysis record must contain assigned and completed cycles, completion by sequence, invalidity reasons, and timing deviations. A later complete cycle has a separate exposure history and cannot fill a missing position in an earlier cycle. Any rule for selecting among repeated cycles must be set before scores are inspected.

Population inference from complete cycles requires an assumption about completion, particularly when pain or fatigue predicts missingness. Without an analysis of that selection process, the estimate describes individuals who completed the cycle. A timestamp model can use an incomplete record only if the A-B contrast remains estimable; its rank, weights, and variance must accompany the estimate. All numerical calculations here use complete cycles and therefore provide no estimate of bias from selective dropout.

PARTICIPANT-LEVEL INFERENCE AND PLANNING

One complete cycle yields one primary contrast per individual. An interval for a group mean uses variation among independent individuals, not seven times the participant count. Repeated cycles or test days require participant clustering; averaging all cycles is appropriate only when it matches the declared target and weighting. More extensive individual trajectories may be better addressed by an explicitly repeated single-case design (Kwasnicka and Naughton, 2020).

For a planned 95% interval half-width h , a normal approximation is $n \geq (1.96 s_D / h)^2$, where s_D is the anticipated SD of completed individual contrasts. Setting $s_D = 4$ cm and $h = 1.5$ cm gives 28 complete individuals after rounding. A finite-sample t calculation gives 30: $t_{0.975, 29}(4/\sqrt{30}) = 1.494$ cm. The 4-cm SD is a stipulated planning input, not a participant estimate. The required count changes with the target half-width, uncertainty in s_D , and incomplete cycles (Lakens, 2022).

DISCUSSION

The seven-position contrast cancels shared cubic drift because its weights are orthogonal to $1, t, t^2, t^3$ while summing to one for A and minus one for B. On an equally spaced grid with equal independent errors, those weights are also the least-squares treatment contrast. In the deterministic reach calculation, the reference second difference removes a 1.05-cm drift contribution and recovers the specified 5.00-cm A-B offset. This identity connects the three neutral scores directly to the condition comparison.

The correction's precision cost is $\sigma_R^2/6$ under independent errors. With reference-error SD 1 cm, the drift-bias threshold for reducing individual mean squared error is $1/\sqrt{6} = 0.408$ cm. Increasing that SD to 3 cm raises the threshold to 1.225 cm, above the 1.05-cm raw bias under the quadratic drift specification. The simulated RMSE consequently increases from 1.449 to 1.579 cm after correction. With allocation-independent drift, random sequence assignment keeps both population means unbiased; correction changes individual error and population precision rather than creating that unbiasedness.

The carryover specification sets $\rho_A = 1.5$ cm and $\rho_B = 0$, leaving bias $(\rho_B - \rho_A)/2 = -0.75$ cm in both contrasts. Removing polynomial drift narrows intervals around that biased mean: corrected 95% interval coverage is 29.40%, compared with 40.16% for the raw contrast. Adding $0.4t$ cm to A creates a different problem. The A effect then changes during the cycle, so a constant-offset target no longer represents the full response. A study of differential learning must define and estimate that time-dependent effect explicitly.

The numerical inputs specify five drift functions, two error structures, a 1.5-cm carryover addition, a 3-cm middle-reference disturbance, and a $0.4t$ cm A addition. They quantify the estimator's response to those inputs; their prevalence in athletes or patients was not estimated. Three reference scores cannot distinguish biological curvature from reference error, and one residual degree of freedom limits checks of the cubic model. Physiological validity, safe movement exposure, adequate recovery, and between-session agreement require observations from the intended motor task and population.

Motor-task assessment requires cycle-completion counts, adverse events, onset and rest durations, reference-error estimates, and agreement of repeated A-B contrasts. Agreement limits and measurement error should accompany reliability coefficients (Koo and Li, 2016; Giavarina, 2015). Comparisons with shorter schedules must account for differences in movement exposure and duration. Prespecifying the estimand, time variable, drift degree, sequence allocation, score validity, missingness rules, and primary contrast fixes the analysis before outcome inspection (Dwan et al., 2019).

CONCLUSION

For R-A-B-R-B-A-R and its condition-reversed sequence, $\hat{\Delta} = C - sQ/6$ removes shared drift through polynomial degree three. Under independent errors, it reduces individual mean squared error when $|b| > \sigma_R/\sqrt{6}$. In the quadratic reach calculation, the correction changes 6.05 cm to the specified 5.00-cm condition difference. Actual-time regression weights preserve polynomial cancellation on full-rank unequal grids. The calculation requires common drift, constant condition offsets, comparable neutral attempts, and qualified recovery; carryover and changing condition effects require separate treatment in the observation model.

DECLARATIONS

CONFLICT OF INTEREST

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

DATA AVAILABILITY

Table 2 contains the seven deterministic reach scores. Table 3 contains every drift and error input. Appendix contains executable Python code, the PCG64 seeds, and the calculations for the unrounded summaries in Tables 4 and 5. All scores are calculated or simulated; no participant measurements were collected.

Software. The computations used Python 3, NumPy 2.3.5, and SciPy 1.17.0. Figures were generated with Matplotlib 3.10.8.

FUNDING

This research received no external funding.

REFERENCES

1. Beatrici A, Rebelo A, Quintão D, Cacaís FL, Loayza VM. Analysis of drift correction in different simulated weighing schemes. J Phys Conf Ser. 2015;648:012025.
2. Bojinov I, Simchi-Levi D, Zhao J. Design and analysis of switchback experiments. Manage Sci. 2023;69(7):3759–3777.

3. Dwan K, Li T, Altman DG, Elbourne D. CONSORT 2010 statement: extension to randomised crossover trials. *BMJ*. 2019;366:l4378.
4. Foldager FN, Aslerin S, Bækdaahl S, Tønning LU, Mechlenburg I. Interrater, test-retest reliability of the Y Balance Test: a reliability study including 51 healthy participants. *Int J Exerc Sci*. 2023;16(4):182–192.
5. Giavarina D. Understanding Bland Altman analysis. *Biochem Med (Zagreb)*. 2015;25(2):141–151.
6. Kattilakoski O, Kauranen N, Leppänen M, Kannus P, Pasanen K, Vasankari T, Parkkari J. Intrarater reliability and analysis of learning effects in the Y Balance Test. *Methods Protoc*. 2023;6(2):41.
7. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155–163.
8. Kwasnicka D, Naughton F. N-of-1 methods: a practical guide to exploring trajectories of behaviour change and designing precision behaviour change interventions. *Psychol Sport Exerc*. 2020;47:101570.
9. Lakens D. Sample size justification. *Collabra Psychol*. 2022;8(1):33267.
10. Linek P, Sikora D, Wolny T, Saulicz E. Reliability and number of trials of Y Balance Test in adolescent athletes. *Musculoskelet Sci Pract*. 2017;31:72–75.
11. Powden CJ, Dodds TK, Gabriel EH. The reliability of the Star Excursion Balance Test and Lower Quarter Y-Balance Test in healthy adults: a systematic review. *Int J Sports Phys Ther*. 2019;14(5):683–694.
12. Shaffer SW, Teyhen DS, Lorensen CL, Warren RL, Koreerat CM, Straseske CA, Childs JD. Y-balance test: a reliability study involving multiple raters. *Mil Med*. 2013;178(11):1264–1270.
13. Wellek S, Blettner M. On the proper use of the crossover design in clinical trials. *Dtsch Arztebl Int*. 2012;109(15):276–281.
14. Wulf G. Attentional focus and motor learning: a review of 15 years. *Int Rev Sport Exerc Psychol*. 2013;6(1):77–104.

EXECUTABLE VERIFICATION AND NUMERICAL EVALUATION

The three blocks form one Python program when executed consecutively with NumPy and SciPy. Block A.1 calculates all seven reach scores, interpolations, and residuals in Table 2, then checks the polynomial, sequence, carryover, variance, and regression identities. Block A.2 uses 20,000 studies of 40 generated individuals for each drift and error specification and prints the unrounded values for Table 4. Block A.3 uses 50,000 grids per jitter level and prints the unrounded values for Table 5. Seeds and sampling order are fixed in the code.

DETERMINISTIC VALUES AND ALGEBRAIC CHECKS

```
import numpy as np
from scipy.stats import t as student_t
t = np.arange(7, dtype=float)
raw_w = np.array([0, .5, -.5, 0, -.5, .5, 0])
cor_w = np.array([-1/6, .5, -.5, 1/3, -.5, .5, -1/6])
refs = np.array([0, 3, 6])
labels_p = np.array(list('RABRBAR'))
labels_m = np.array(list('RBARABR'))
def verify():
    from fractions import Fraction as F
    w = list(map(F, ['-1/6', '1/2', '-1/2', '1/3', '-1/2', '1/2', '-1/6']))
    assert [sum(w[j]*j**p for j in range(7)) for p in range(5)] == [0,0,0,0,-12]
    for sequence, s in [(('RABRBAR',1), ('RBARABR',-1))]:
        y = np.array([72+.8*j+.35*j*j+{'R':0, 'A':3, 'B':-2}[c] for j,c in enumerate(sequence)])
        assert np.isclose(s*(y@cor_w),5)
        for previous, expected in [(('A',-.5), ('B',.5))]:
            z = np.array([int(j>0 and c!='R' and sequence[j-1]==previous) for j,c in enumerate(sequence)])
            assert np.isclose(s*(z@cor_w),expected)
    X = np.column_stack([t**j for j in range(4)]+[(labels_p==c).astype(float) for c in ['A','B']])
    assert np.allclose(np.array([0,0,0,0,1,-1])@np.linalg.pinv(X),cor_w)
    assert np.isclose(cor_w[refs]@cor_w[refs],1/6)
    assert np.isclose(cor_w@cor_w,7/6)
    print('Analytical verification passed.')
verify()
for sequence, s in [(('RABRBAR',1), ('RBARABR',-1))]:
    g = 72 + .8*t + .35*t*t
    offsets = np.array([{'R':0, 'A':3, 'B':-2}[c]
                        for c in sequence])
    scores = g + offsets
    interpolated = np.interp(t, t[refs], scores[refs])
```

```
print(sequence, 'raw/corrected',
      s*(scores@raw_w), s*(scores@cor_w))
for j in range(7):
    lt = None if j in refs else interpolated[j]
    residual = None if j in refs else scores[j]-lt
    print(j, sequence[j], g[j], offsets[j],
          scores[j], lt, residual)
```

COMPARATIVE STUDY GENERATION AND INFERENCE

```
SEED = 20261010
M, N = 20000, 40
CASES = [
    ('Flat', 'flat', 1., 0., 'none'),
    ('Quadratic', 'quadratic', 1., 0., 'none'),
    ('Cubic', 'cubic', 1., 0., 'none'),
    ('Saturating', 'saturating', 1., 0., 'none'),
    ('Fatigue step', 'step', 1., 0., 'none'),
    ('Noisy references', 'quadratic', 3., 0., 'none'),
    ('Serial error', 'quadratic', 1., .6, 'none'),
    ('Carryover', 'quadratic', 1., 0., 'carry'),
    ('Middle reference', 'quadratic', 1., 0., 'middle'),
    ('Changing A effect', 'quadratic', 1., 0., 'slope'),
]
def drift(name, time):
    if name == 'flat': return np.zeros_like(time)
    if name == 'quadratic': return .35*(time-3)**2
    if name == 'cubic': return .2*time + .2*time**2 + .015*time**3
    if name == 'saturating': return 4*(1-np.exp(-time/1.5))
    if name == 'step': return -4*(time >= 4).astype(float)
    raise ValueError(name)
def study(j, case):
    name, shape, sig_r, phi, violation = case
    rng = np.random.Generator(np.random.PCG64(SEED+j))
    total = M*N
    s = rng.choice([-1., 1.], size=total)
    delta = rng.normal(5., 1.5, size=total)
    labels = np.where((s == 1)[:], None, labels_p, labels_m)
    scales = np.ones(7); scales[refs] = sig_r
    cov = np.outer(scales, scales)*phi*np.abs(t[:, None]-t[None, :])
    errors = rng.normal(size=(total, 7)) @ np.linalg.cholesky(cov).T
    scores = 72 + drift(shape, t)[None, :] + (labels == 'A')*delta[:, None] + errors
    if violation == 'carry':
        scores[:, 1:] += 1.5*((labels[:, :-1] == 'A') & (labels[:, 1:] != 'R'))
    if violation == 'middle': scores[:, 3] += 3.
    if violation == 'slope': scores += .4*t[None, :]*(labels == 'A')
    estimates = np.column_stack([s*(scores @ raw_w), s*(scores @ cor_w)])
    error = estimates-delta[:, None]
    groups = estimates.reshape(M, N, 2)
    means = groups.mean(axis=1)
    sem = groups.std(axis=1, ddof=1)/np.sqrt(N)
    covered = np.abs(means-5.) <= student_t.ppf(.975, N-1)*sem
    mse = np.mean(error**2, axis=0)
    rmse = np.sqrt(mse)
    mcse_rmse = (error**2).std(axis=0, ddof=1)/(2*rmse*np.sqrt(total))
    return {'name': name, 'rmse': rmse.tolist(), 'mcse_rmse': mcse_rmse.tolist(),
            'bias_plus': error[s == 1].mean(axis=0).tolist(),
            'bias_minus': error[s == -1].mean(axis=0).tolist(),
            'group_bias': (means-5).mean(axis=0).tolist(),
            'group_mcse': (means.std(axis=0, ddof=1)/np.sqrt(M)).tolist(),
            'coverage': covered.mean(axis=0).tolist()}
results = [study(j, case) for j, case in enumerate(CASES)]
for result in results:
    print(result)
print('maximum RMSE Monte Carlo SE',
      max(max(r['mcse_rmse']) for r in results))
```

ACTUAL-TIME CONTRAST EVALUATION

```
def timing(level, j, k=50000):
    rng = np.random.Generator(np.random.PCG64(SEED+100+j))
    time = np.tile(t, (k, 1))
    time[:, 1:6] += rng.uniform(-level, level, size=(k, 5))
    x = (time-3)/3
    ap = (labels_p == 'A').astype(float); bp = (labels_p == 'B').astype(float)
    X = np.stack([np.ones_like(x), x, x*x, x**3,
                  np.broadcast_to(ap, x.shape), np.broadcast_to(bp, x.shape)], axis=2)
    c = np.broadcast_to(np.array([0,0,0,0,1,-1]), (k,6))
    coeff = np.linalg.solve(np.swapaxes(X,1,2) @ X, c[:, None])[..., 0]
    w = (X @ coeff[:, None])[..., 0]
    g = drift('cubic', time)
    residual_fixed = g @ cor_w
    residual_adapt = np.sum(g*w, axis=1)
    mse_fixed = np.sum(cor_w**2)+np.mean(residual_fixed**2)
    mse_adapt = np.mean(np.sum(w*w, axis=1)+residual_adapt**2)
    return {'level': level, 'fixed_rmse': float(np.sqrt(mse_fixed)),
            'adapt_rmse': float(np.sqrt(mse_adapt)),
            'adapt_var95': float(np.quantile(np.sum(w*w,axis=1), .95)),
            'max_abs_residual': float(np.max(np.abs(residual_adapt))),
            'condition95': float(np.quantile(np.linalg.cond(X), .95))}

timing_results = [timing(a,j) for j,a in
                  enumerate([0., .05, .1, .2, .4])]

for result in timing_results:
    print(result)
```