Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Trial-count estimates for table tennis displacement waveforms vary with action, representation and analytical assumptions

Mingke Han
South China normal university
Yupeng Shen
South China normal university
Lianzhen Ma
South China normal university
Pengfei Jin
Shanghai Ocean University
Chen Yang
Southwest Jiaotong University
Chaochong Deng
South China normal university
Liang Wang
20051116@m.scnu.edu.cn

South China normal university

Article

Keywords: table tennis, functional data, repeated measures, relative reliability, trial count, motion capture

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10541150/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Trial-count estimates for table tennis displacement waveforms vary with action, representation and analytical assumptions

Mingke Han[1†], Yupeng Shen[1†], Lianzhen Ma[1], Pengfei Jin[2], Chen Yang[3], Chaochong Deng[1], Liang Wang[1*]

[1] School of Physical Education and Sports Science, South China Normal University, Guangzhou 510006, China

[2] Department of Physical Education and Sport, Shanghai Ocean University, Shanghai 201306, China

[3] SWJTU-Leeds Joint School and College of Sports and Health, Southwest Jiaotong University, Chengdu 611756, China

[†] Mingke Han and Yupeng Shen contributed equally to this work and share first authorship.

[*] Correspondence: Liang Wang, School of Physical Education and Sports Science, South China Normal University, 378 Outer Ring Road, Guangzhou University Town, Panyu District, Guangzhou, Guangdong 510006, China. Email: 20051116@m.scnu.edu.cn. ORCID: https://orcid.org/0009-0003-9906-0803

Abstract

Repeated-trial mean waveforms often characterise individuals, but the required trial count is not fixed. We reanalysed the public TTMD6 archive (30 archive codes, six fixed English action labels and 9,000 trials), constructed racket and aggregated-body displacement waveforms, and used 5,000 archive-code cluster bootstrap replicates to estimate the number of trials required to reach a relative-reliability reference level of 0.90. For event-unregistered waveforms under negligible residual dependence, minimum integer estimates from the L^2 trace ratio, which integrates variance across the full waveform, were 6–27 trials for the racket and 4–8 for the aggregated body. For backhand drive (BD), the racket estimate was 27 trials, compared with 17 from the phase-averaged pointwise reliability. The continuous BD threshold was 26.32 (95% bootstrap percentile interval, 17.65–43.48). Estimates were sensitive to assumptions: a Hampel-type local high-value rule reduced the BD racket estimate to 18; a positive trial-dependence scenario (AR(1), φ = 0.30) gave 49 with fixed variance components; and single proxy-peak registration yielded 3–31 trials across conditions. These estimates apply to the specified TTMD6 actions, representations and within-session assumptions and should not be interpreted as a general acquisition prescription.

Keywords: table tennis; functional data; repeated measures; relative reliability; trial count; motion capture

Introduction

Movement biomechanics commonly characterises an individual using a mean waveform calculated from repeated trials, yet the question of how many repetitions are sufficient is often reduced to a fixed

acquisition count. The same count may be intended to estimate a single-trial peak, a time-normalised mean waveform, day-to-day change or relative differences between individuals, and each purpose has a different signal and error structure [1–4]. This distinction is especially important in dense movement archives. Thousands of files do not increase the number of independent units at the population level, which remains determined by the number of participants. Treating repeated trials from one participant as independent observations creates pseudoreplication and systematically understates population-level uncertainty [5].

Continuous waveforms retain more information about the course of an action than a single peak or mean, but they also require a more explicit estimand. Between-waveform variation may combine amplitude differences, temporal displacement of internal events and segmentation error; percentage time normalisation does not automatically remove these sources [6–9]. Any claim about waveform reliability must therefore specify how the waveform was constructed, whether it was registered, the target population and time scale, and whether the objective is relative discrimination or absolute agreement [10].

Generalizability theory explains how replication reduces error, but a traditional decision study can be transferred to functional data only when the object of measurement, sources of error and intended universe of inference are defined [11,12]. If a time-normalised waveform is treated as a random element in L^2 over the unit interval, the trace of its covariance operator summarises pointwise variance over the action cycle, allowing relative reliability to be expressed as a ratio of hierarchical covariance traces [13,14]. This framework compresses between-archive-code and within-archive-code trial variation into an overall ratio, while exposing a crucial assumption: the mean error decreases as W/n only when residual dependence between trials is negligible. Positive dependence during sequential acquisition would make an independence-based trial count too small [15,16].

Action conditions should not be collapsed into a single global reliability coefficient. Table-tennis actions differ in racket trajectories and joint motions, and both within-participant and between-participant variation occur within an action [17,18]. Counting action differences as signal may make archive-code–action combinations easy to distinguish, but does not answer whether archive codes can be separated reliably within a fixed action. Moreover, the six action labels form a finite fixed set. Their dispersion describes only these conditions and is not a random action variance sampled from all table-tennis techniques [12].

The TTMD6 source article reports 30 athletes, six action labels and 9,000 actions, with 55 consecutive repetitions recorded per acquisition [19]. The public archive, however, contains 12,000 matched racket and body files assigned to codes 1–40, with 50 trials in each code–action cell. This structure requires audit before inference. No public mapping links the archive codes to the 30 reported participants; the archive also contains duplicated storage blocks, records whose nominal length differs from the matrix row count, and structural zero body landmarks. The selection of the public 50 trials from the reported 55 cannot be verified. Direct label validation in the source paper was limited to 30 separate racket-trajectory animations,

five per class, identified by 40 table-tennis specialists with a mean accuracy of 92.63%; most errors occurred between attack and drive [19]. We therefore treat forehand attack, forehand drive, forehand push, backhand attack, backhand drive and backhand push only as fixed archive labels. That small animation study cannot replace trial-level verification of 9,000 records or validate the body-coordinate labels or the code-to-participant mapping.

We consequently performed an estimand-led, conditional analysis of trial count. We asked: within the six fixed labels, how much archive-code variation is consistent across actions and how much is action dependent; how many event-unregistered, percentage-normalised trials are required to reach a relative-reliability reference level for each action and representation; and how sensitive the estimates are to quality rules, archive-code influence, residual dependence, peak registration, archive boundaries and bootstrap method. Our aim was not a transferable acquisition prescription, but a reproducible framework that reports trial counts together with their assumptions and inferential limits.

Results

Archive audit defined the analysis cohort and estimand

The size and MD5/ETag identifiers of the local TTMD6 archive matched the Springer supplementary object linked to the original Scientific Reports article. The complete archive contained 12,000 matched racket–body files assigned to codes 1–40. Because the source article reported 30 participants and codes 1–30 formed the first complete contiguous block, the primary analysis used a deterministic rule to include the 9,000 pairs under codes 1–30 and quarantine the 3,000 pairs under codes 31–40, whose provenance was not explained. This rule does not imply that the source authors confirmed a one-to-one mapping between codes 1–30 and the reported participants. The primary cohort therefore comprised 30 archive codes × six fixed English labels × 50 trials. The labels were FA (forehand attack), FD (forehand drive), FP (forehand push), BA (backhand attack), BD (backhand drive) and BP (backhand push) (Table 1; Supplementary Table S1).

Fifty racket files contained two element-wise identical 200-row blocks; only the first block was retained under a deterministic rule. The nominal filename length exceeded 200 frames in 1,180/9,000 trials. Because the transformation from the complete action to the public 200-row matrix cannot be reconstructed, 200 rows were defined as an archive storage/transformation boundary rather than a verified truncation at the first 200 frames. A total of 190 body trials contained at least one exact (0, 0, 0) landmark triplet, giving 9,594 landmark–frame triplets. After landmarks were excluded whenever either endpoint of a frame interval contained a structural-zero candidate, at least 10 landmarks remained available for every interval. No complete racket–body trial was duplicated within the primary cohort, although eight complete duplicates linked the primary cohort to additional codes. Because no public code-to-participant mapping

exists, the 30 archive codes were treated as independent clusters under an explicit but unverified working assumption, with 9,000 trials nested within clusters (Fig. 1; Table 1).

Three alternative analyses were used only as robustness checks: the equally wide window of codes 11–40, all codes 1–40, and unbalanced REML after removing the codes 31–40 member of each of the eight exact duplicate pairs. The BD racket L^2 trace-ratio integer threshold was 27, 27, 25 and 25 across the four cohort scenarios; the corresponding phase-averaged pointwise thresholds were 17, 20, 17 and 17. Across all 12 action–representation conditions, the alternative window changed the L^2 threshold by 0–4 trials, and exact de-duplication did not change any integer result from the full 40-code analysis. The main findings were therefore not driven by the eight duplicates, but the 40 codes still cannot be interpreted as 40 confirmed independent participants. The pre-specified operational cohort of codes 1–30 remained primary (Supplementary Table S2).

Archive-code differences within the six fixed labels were predominantly condition dependent

The across-action background model treated the six labels as fixed conditions and used the noise-corrected finite-set action-contrast dispersion D(A) to describe separation among their means, rather than calling it a random action variance. After D(A), the across-action archive-code component, the archive-code × fixed-action component and unexplained within-cell variation were integrated over normalised time, their racket-waveform proportions were 31.2%, 1.5%, 29.0% and 38.3%. Corresponding medians from 5,000 archive-code cluster bootstrap replicates (95% percentile intervals) were 32.1% (27.6–37.4%), 1.5% (0.6–3.0%), 28.0% (23.6–32.0%) and 38.3% (33.9–42.7%).

For the aggregated-body waveform, point estimates were 50.0%, 4.4%, 27.1% and 18.5%; bootstrap medians (95% percentile intervals) were 50.9% (45.6–56.1%), 4.3% (2.4–6.6%), 26.1% (21.5–30.8%) and 18.4% (16.0–21.7%). In both representations, the archive-code × fixed-action component greatly exceeded the across-action archive-code component. Relative differences between archive codes therefore changed with action rather than persisting across labels as a common effect (Fig. 2). This matters directly for trial-count estimation. Each action has its own between-code signal relative to within-cell error, whereas pooling labels would count action contrasts as signal and answer a different question. We did not test broader identity classification or a ‘movement fingerprint’. Pointwise curves and marginal bootstrap intervals show only how dispersion is distributed over normalised time and are not local significance tests (Supplementary Fig. S3).

Action and representation jointly determined relative reliability and trial count

All archive-code variance estimates were interior solutions at the 200 nodes of the 12 action-specific models. Under the primary scenario of negligible trial-residual dependence, $R(L^2,a;10)$ for the racket

ranged from 0.774 for BD to 0.939 for FA; aggregated-body values ranged from 0.919 for BD to 0.965 for FP. The minimum integer trial counts satisfying $R(L^2,a;n) = B_a/[B_a + W_a/n] \geq 0.90$ were 6, 7, 8, 17, 27 and 10 for the racket, and 6, 6, 4, 7, 8 and 4 for the aggregated body, in FA-to-BP order (Fig. 3; Table 2; Supplementary Table S4). These are action-specific descriptive estimates, not multiply adjusted tests of between-action differences.

The functional summary changed some conclusions. When relative reliability was first calculated at each phase node and then averaged over phase by trapezoidal integration, the 0.90 integer thresholds were 7, 8, 9, 16, 17 and 11 for the racket, and 6, 6, 4, 7, 8 and 5 for the body (Table 2). Most conditions differed by 0–1 trial, but the BD racket threshold changed from 27 for the L^2 trace ratio to 17 for the phase-averaged pointwise ratio. The two estimates reflect different phase-weighting schemes and therefore represent distinct functional summaries.

Trial count was determined by W_a/B_a rather than W_a alone. For the BD racket, $B_a = 72.758$, $W_a = 212.798$ and $W_a/B_a = 2.925$, giving a continuous threshold of 26.32 and an integer threshold of 27. For FA, W_a/B_a was 0.650 and the continuous threshold was 5.85. A larger trial count therefore means that between-code signal is weaker relative to within-cell trial error in the current population and representation; it does not by itself show that an action is more difficult, unstable or technically inferior.

The 95% percentile intervals from 5,000 archive-code cluster bootstrap replicates ranged from 3.80–10.29 for FA to 17.65–43.48 for BD in the racket, and from 2.40–5.21 for FP to 6.31–11.07 for BD in the aggregated body (Table 2). Percentile intervals were primary because they preserve the positive ratio scale and are invariant under monotone transformation. Basic intervals and BCa intervals, which adjust bootstrap limits for bias and skewness, assessed method sensitivity. None of 60,000 fits produced $B_a \leq 0$ or an undefined threshold, and only two FD racket replicates contained a boundary solution at any node. For the BD racket, the BCa interval was 16.47–40.04 and the basic interval was 9.17–35.00. These differences show that the threshold distribution is skewed with 30 independent clusters, so point estimates require interval-based interpretation. Bootstrap diagnostics are provided in Supplementary Table S8.

Sensitivity of the BD racket estimate to the quality rule was distributed

The primary rule treated exact structural zeros as missingness candidates and retained all finite non-zero coordinates. A Hampel-type local high-value rule, defined in Methods from a local median and scaled median absolute deviation, flagged 423/1,647,435 racket frame-interval values (0.0257%). Its parameters were fixed in code before this sensitivity run and were not tuned by action. Of the flagged values, 309 (73.0%) occurred in BD and involved 292/1,500 BD trials and 27/30 archive codes (Supplementary Fig. S2). For the BD racket, the rule reduced B_a from 72.758 to 67.456 (−7.3%) and W_a from 212.798 to 134.625 (−36.7%); the continuous threshold fell from 26.32 to 17.96 and the integer estimate from 27 to 18.

The corresponding 95% bootstrap percentile interval was 12.87–29.01. The other five racket thresholds remained 6, 7, 8, 17 and 10; for the body, only FA and FD changed, from six to five trials.

Leave-one-archive-code-out analysis gave continuous BD racket thresholds of 23.38–28.79 (integers 24–29) under the primary rule and 16.66–19.59 (integers 17–20) under the Hampel-type rule. The reduction from 27 to 18 was therefore not removed by excluding one anomalous archive code, but arose from the cumulative influence of high values across multiple codes and trials (Fig. 4d). Without device-level error limits, framewise tracking confidence or externally synchronised ground truth, neither rule can be designated uniquely correct. The BD estimates should therefore be reported separately as 27 under the primary rule and 18 under the Hampel-type sensitivity rule.

Archive-order structure did not validate the independence assumption

Each archive-code–action cell contained 50 files with consecutive global indices. We defined within-cell archive ranks 1–50 by ascending global index. These ranks were not independently supplied trial numbers, and no timestamps, continuous-file mapping or acquisition logs establish that they represent acquisition order. Residual L^2 trace-type lag-1 correlations calculated by archive rank were positive in all 12 action–representation conditions, ranging from 0.075 to 0.328. Every observed value exceeded the condition-specific 97.5% reference bound from 5,000 within-code random order permutations (Fig. 4a). This permutation analysis only characterises how unusual the numeric archive order is relative to random ordering; it is not a confirmatory test of an acquisition time series. The result indicates order structure in file indices, not verified temporal autocorrelation, fatigue or learning. Condition-specific values are reported in Supplementary Table S5.

Because acquisition order could not be independently verified, results were reported in two layers. The primary layer assumed negligible residual dependence. The second retained the primary B_a and W_a values but changed the mean-error reduction term under a positive AR(1) working scenario, in which trial correlation decays geometrically with lag. For the BD racket, $\varphi = 0$, 0.10, 0.20 and 0.30 gave 27, 32, 40 and 49 trials; corresponding BA values were 17, 21, 25 and 31 (Fig. 4b). The complete scenario results are provided in Supplementary Table S6. Because φ was not estimated from verified acquisition times and the variance components were held fixed, these results are scenario-specific sensitivity estimates. They should not be interpreted as empirical bounds or acquisition recommendations.

Registration and archive sensitivities further constrained waveform meaning

A single robust proxy peak within 10–90% of each trial's racket waveform was aligned to the action-specific median peak location. The racket integer thresholds changed from 6, 7, 8, 17, 27 and 10 before registration to 4, 5, 4, 6, 31 and 4 after registration. Applying the same time transformation to paired body

waveforms changed their thresholds from 6, 6, 4, 7, 8 and 4 to 6, 4, 3, 4, 5 and 3 (Fig. 4c). Most estimates fell, but the BD racket increased from 27 to 31. Registration neither reduced error consistently nor supplied a more correct answer. It changed the estimand and demonstrated that the primary estimate combined amplitude dispersion with internal temporal displacement. The primary results therefore refer explicitly to event-unregistered, time-normalised archive waveforms. Full registration results are shown in Supplementary Table S7.

After excluding 1,180 trials whose nominal length exceeded 200 frames, racket thresholds were 5, 6, 7, 16, 23 and 10, and body thresholds were 5, 5, 3, 7, 7 and 4. Across 1,000 action-wise balanced subsamples of this exclusion set, the median $R(L^2,a;50)$ differed from the full unbalanced REML estimate by at most 0.0021 (Supplementary Fig. S1). Complete deletion of 190 body trials containing structural zeros gave body thresholds of 5, 6, 4, 7, 9 and 4; fixed 13-landmark and fixed 8-landmark analyses retained the action-specific pattern. These results support the conclusion that labels and representations produce different conditional estimates, but they do not remove uncertainty associated with the BD quality rule, residual dependence or registration.

Discussion

The archive did not support a single trial-count recommendation. Estimates varied systematically across actions, waveform representations, functional summaries, quality rules, registration choices and residual-dependence scenarios. For event-unregistered waveforms under negligible residual dependence and with all finite non-zero coordinates retained, L^2 trace-ratio estimates were 6–27 trials for the racket and 4–8 for the aggregated body. The phase-averaged pointwise summary changed the BD racket estimate from 27 to 17. The BD racket estimate was 18 under the Hampel-type sensitivity rule and 49 in the $\varphi = 0.30$ AR(1) scenario with fixed variance components. Single proxy-peak registration produced estimates of 3–31 trials across conditions. Together, these findings show that trial-count estimates depend on both the measurement target and the analytical choices used to define and process the waveform. Earlier repeated-trial studies estimated the required count for a selected scalar or trajectory outcome [1–4]. Fixed-facet decomposition and the amplitude–phase distinction help explain why that count changes when the action, representation or registration target changes [6–8,11,12].

The statistical status of the six labels determines the interpretation of the across-action results. D(A) quantified the noise-corrected contrast dispersion among the six fixed-label means. The 31.2% racket and 50.0% body estimates therefore describe this finite set rather than a random action variance generalisable to all table-tennis actions. The across-action archive-code component accounted for 1.5% of integrated racket dispersion and 4.4% of integrated body dispersion. The corresponding archive-code × fixed-action proportions were 29.0% and 27.1%. Within these six labels, relative archive-code differences were

predominantly action dependent. That pattern matters for study design because adding trials reduces within-cell error but does not make between-code differences consistent across actions. A pooled count would therefore hide the action-specific balance of signal and error that produced thresholds from 4 to 27 trials. Broader individualised comparisons, including nonlinear or multivariate identity features, require a fixed action and representation and were beyond the scope of this analysis.

Under negligible residual dependence, an action-specific threshold follows directly from $n^*_{a,0.90} = 9W_a/B_a$. The larger BD racket threshold reflected substantial within-cell dispersion relative to the between-code signal, whereas the smaller FA threshold reflected a lower W_a/B_a ratio. In biomechanical terms, the BD pattern may reflect trial-to-trial changes in stroke initiation or impact timing under machine delivery [18,20,21]. Variation within archive codes in racket-path shape, preparation or follow-through may also increase W_a. High values were distributed across 27 archive codes and 292 trials, so no single code explained the result. Their concentration in BD, and the fall from 27 to 18 after filtering, indicate that short high-displacement intervals contributed to W_a. However, proxy-peak registration increased the estimate to 31, which argues against misalignment of one prominent peak as the sole explanation. TTMD6 provides no ball speed, spin, placement, outcome, racket-face orientation, verified impact event or kinetic variable with which to distinguish timing, technique and measurement-error contributions [22]. These mechanisms remain hypotheses, and the threshold should not be read as action difficulty, technical quality or training dose.

The aggregated-body waveform was included chiefly as a representation contrast, not as an independent measure of whole-body control or coordination. We interpreted the 42 body columns as 14 consecutive three-coordinate blocks in the label order shown in Fig. 1 of the source paper: Hips, Head, LeftShoulder, LeftArm, LeftForeArm, RightShoulder, RightArm, RightForeArm, LeftUpLeg, LeftLeg, LeftFoot, RightUpLeg, RightLeg and RightFoot [19]. The body measure equally averaged Euclidean displacement magnitudes across available landmarks for each frame interval. Equal weighting was chosen because no segment mass or inertia parameters were available and avoided introducing unverifiable mass-based weights. Euclidean magnitudes avoided cancellation of signed coordinates, but the construction aggregates and smooths while discarding direction, local segment behaviour and joint configuration. Biomechanically, it describes only the overall amount of landmark displacement. It is not a centre of mass, joint angle or coordination metric. The lower body trial counts show that this aggregate had a larger between-code-to-within-cell variance ratio. They do not show that body movement was biomechanically more stable. A study of centre-of-mass motion, segmental control or joint coordination would require a representation designed for that purpose and a new trial-count estimate.

The primary and Hampel-type rules produced different BD racket estimates because the latter reduced within-cell dispersion. Although the sensitivity rule modified only 0.0257% of racket frame-interval values,

it reduced the threshold mainly through W_a. Flags occurred in 27 codes and 292 trials, and leave-one-code-out analysis preserved the separation between rules, indicating that no single archive code drove the result. The archive lacks device error limits or original tracking-quality indicators needed to establish whether every flagged value was artefactual. Retrospective selection of a preprocessing threshold to obtain a preferred reliability would introduce hindsight bias. We therefore report 27 trials under the primary rule and 18 under the pre-specified Hampel-type sensitivity rule as rule-dependent estimates.

Residual dependence and registration define separate inferential boundaries. The positive lag association in archive rank left the W_a/n reduction unverified because archive order alone was insufficient to establish an AR(1) acquisition time series. We therefore treated $\varphi = 0.10–0.30$ as working scenarios. Proxy-peak registration changed W_a/B_a in both directions, and the BD racket required more trials after registration. The unregistered estimate thus describes the relative distinguishability of whole waveforms after archive segmentation and percentage normalisation, rather than pure amplitude reliability. Future studies should preserve acquisition order, impact events, original time stamps and trajectory-quality information. Those data would support block resampling, correlated-residual modelling and pre-registered functional registration with clearer mechanistic interpretation.

Strengths include use of the complete $30 \times 6 \times 50$ operationally balanced structure, retention of archive code as the highest independent level, and integration of archive audit, alternative code windows, exact de-duplication, functional summaries, continuous thresholds, cluster bootstrap, leave-one-archive-code-out analysis, AR(1) scenarios, registration and quality rules in one inferential framework. Several limitations constrain interpretation. The source study came from one institution and reported 30 right-handed participants; the public archive provided 50 repetitions per archive-code–action cell. Repetition improves estimation of W_a but not cluster-level information. The source article reports ages 14–24 years and written informed consent, but the public materials do not allow verification of guardian-consent procedures for minors. The selection of 50 trials from the reported 55, the one-to-one mapping of codes 1–30 to participants, the origin of codes 31–40, the 200-row transformation and the official meaning of structural zeros remain unconfirmed [19]. Alternative-window and de-duplication analyses reduce concern about their numerical impact but cannot reconstruct a missing identity crosswalk. This outcome-informed secondary analysis was not preregistered. It included neither between-day repetition nor left-handed or novice participants and cannot assess absolute agreement, responsiveness or identity-classification performance.

The current estimates are most useful for initial planning in studies that closely match the present estimand. Here, the target was within-session relative distinguishability of event-unregistered, percentage-normalised mean displacement waveforms within a fixed action. With comparable participants, ball delivery, capture procedures, waveform construction and the primary quality rule, a pilot could start near 27 usable BD racket trials. This anchor assumes negligible residual dependence and should not be treated as the final

requirement. The estimates of 18 under the Hampel-type rule and 31 after proxy-peak registration show why the quality rule and registration policy should be decided before the definitive count is set. Where both quality rules are scientifically plausible, reporting both estimates preserves that design uncertainty.

When the target population, ball delivery or capture system differs materially from TTMD6, the numerical estimate should be recomputed from local repeated-trial data. The TTMD6 values can help size that pilot, but they cannot replace local estimates of B_a and W_a or a check of trial-order dependence. Resampling should preserve participants as independent clusters; fixed actions should be analysed separately when participant differences vary by action, and actual acquisition order should be recorded [11,12,15,16]. The waveform, success criteria, quality rule and registration policy should be specified before estimation. The number of independent participants determines the precision of the between-participant component. Even 30 archive-code clusters produced a wide, skewed interval for BD, so this archive does not establish a universal minimum cluster count. Cluster adequacy should instead be assessed from boundary solutions, bootstrap stability and leave-one-cluster influence rather than from a fixed numerical cutoff.

The transferable part of the framework is this sequence of design decisions rather than the TTMD6 counts. It applies when repeated continuous measurements can be placed on a common phase domain and their squared differences have a meaningful interpretation. It also requires local estimates of between- and within-participant variation [6,11–14]. The six action labels, machine delivery, right-handed source cohort and 200-node displacement construction are specific to TTMD6. Other racket sports can use the same workflow after redefining the task and waveform, preserving acquisition logs, and examining quality processing, registration and residual dependence for the new setting. The resulting trial count remains task-, population- and system-specific.

Methods

Study design and data source

This was a secondary analysis of a public, de-identified coordinate archive; no participants were newly recruited or contacted. The TTMD6 source study reported 30 right-handed athletes (17 male and 13 female; 14–24 years) with Chinese national second-class, first-class or master-level classifications, all described as undergraduates at the China Table Tennis College [19]. We use the reported sporting classifications to describe the source cohort and do not equate the article's term 'professional players' with professional employment or national-team status. Participants used the same racket and performed diagonal strokes against balls delivered by a machine at the reported 6.5 m/s; motion capture was sampled at 120 Hz [19]. No mapping links individual-level sex, age or classification to archive codes, so these variables were neither stratified nor modelled. Motion-capture coordinates are generally affected by calibration, occlusion,

marker definition and task environment [22]; this reanalysis could audit public coordinates but could not recover device-level quality records.

Ethics approval and consent to participate

This study was a secondary analysis of the publicly available, de-identified TTMD6 dataset. The authors did not recruit or interact with participants, collect new human-participant data, or have access to direct identifiers or a re-identification key. No additional ethics approval was sought for the present analysis. The original data collection was approved by the Ethics Committee of Shanghai University of Sport (approval no. 102772019RT030). The source article reports that participants provided written informed consent and that all methods were performed in accordance with the approved protocol [19].

The complete block of codes 1–30 was used under the deterministic rule described below. The number of independent archive codes was fixed by the data source and no post hoc power calculation was performed. Fifty repeated trials per code–action cell improved estimation of within-cell dispersion, whereas cross-code precision remained limited by N = 30. Uncertainty was therefore quantified with archive-code cluster bootstrap, leave-one-archive-code-out analysis and comparisons among interval methods.

Archive acquisition, cohort definition and quality audit

The analysed object was the local file TTMD6.rar (341,074,031 bytes; MD5 1c9ce9cbf79dd35dd22f16a7199e2a8c; SHA-256 93d1b52a470f14b9dc0ba0600959bff921be891a3da1b71e609bd328224b354d). On 15 July 2026, the Springer Nature supplementary object linked to the source article was downloaded in full. Its byte count and both hashes matched the local archive, and direct bytewise comparison found no difference. The input analysed here was therefore identical to the publisher-hosted supplementary archive. The later figshare record [23] and version differences are described separately under Data Availability.

After complete extraction, we built a manifest, paired racket and body files, and parsed archive code, action label, global index and nominal filename length. Archive ranks 1–50 were then assigned within each code–action cell by ascending global index. Because the source article reported 30 participants and codes 1–30 formed the first complete contiguous block, the primary cohort was fixed deterministically as $30 \times 6 \times 50$; all codes 31–40 were quarantined because their provenance was unknown. This operational rule does not establish participant identity. Element-wise hashes and array comparison identified duplicated 400-row blocks and complete duplicate records. When nominal length was below 200, trailing all-zero padding rows were not used. For nominal lengths of 200 or more, all 200 public matrix rows were used without claiming to reconstruct the nominal full duration.

Body files contained 42 columns representing 14 three-dimensional landmarks; racket files contained three coordinates for the racket centre. Exact (0, 0, 0) triplets were treated operationally as structural-zero or missingness candidates, without asserting an official meaning. If either endpoint of a frame interval was zero for a landmark, that landmark did not contribute to the aggregated-body displacement for that interval. The primary analysis did not remove trials according to outcome magnitude, impute raw coordinates or winsorise non-zero high values. All deterministic audit rules were applied before model fitting and recorded in a file-level manifest.

Action labels, waveform construction and primary estimand

The original labels and codes were FA (forehand attack), FD (forehand drive), FP (forehand push), BA (backhand attack), BD (backhand drive) and BP (backhand push). In the source paper, 40 specialists classified a separate sample of 30 racket-trajectory animations, five per class, with a mean accuracy of 92.63%; push had the highest recognition and the main confusion was between attack and drive [19]. This evidence supports coarse recognition for a small sample of racket trajectories, not trial-level adjudication of 9,000 records, validation of body labels or identification of archive codes. Because no trial-level ball speed, spin, placement, impact video or expert adjudication was available, modelling and conclusions used the English archive labels and codes.

The racket waveform was the Euclidean norm of the three-dimensional change in the racket-centre coordinate between adjacent valid frames:

$$d_{par}^{R}(t) = \left\| x_{par}^{R}(t+1) - x_{par}^{R}(t) \right\|_2 \quad (1)$$

The aggregated-body waveform was the mean landmark displacement magnitude across landmarks that were available at both adjacent frames:

$$d_{par}^{H}(t) = \left| V_{par,t} \right|^{-1} \sum_{j \in V_{par,t}} \left\| x_{jpar}(t+1) - x_{jpar}(t) \right\|_2 \quad (2)$$

Here p, a, r and t denote archive code, fixed action label, nested trial and original frame interval, and $V_{par,t}$ is the set of landmarks that were non-zero at both endpoints. The 14 labels shown in Fig. 1 of the source paper were Hips, Head, LeftShoulder, LeftArm, LeftForeArm, RightShoulder, RightArm, RightForeArm, LeftUpLeg, LeftLeg, LeftFoot, RightUpLeg, RightLeg and RightFoot. The 42 unlabelled body CSV columns were mapped to consecutive three-coordinate blocks in that order [19]. The operational landmark mapping is provided in Supplementary Table S3. Because no finer marker-centre definitions, segment masses or inertia parameters were available, displacement magnitudes were averaged equally across available landmarks. Euclidean magnitudes also avoided cancellation of signed coordinates. Both measures are frame-interval displacement magnitudes in archive coordinate units. They are not velocity, racket-face orientation, centre of mass, joint angle, coordination, torque or power. Each waveform was linearly interpolated to 200 equally spaced nodes over 0–100% of its observable interval [6,7].

The primary estimand was the relative distinguishability of event-unregistered, percentage-normalised mean displacement-magnitude waveforms within a fixed original English action label, among the included TTMD6 archive-code clusters and within the same acquisition session. It does not include between-day retesting, device recalibration, absolute agreement or performance prediction.

Finite-set background decomposition of the six fixed labels

The across-action model was used only to describe the dispersion background within the six fixed labels:

$$Y_{par}(u) = \mu(u) + \alpha_a(u) + P_p(u) + (P\alpha)_{pa}(u) + \varepsilon_{par}(u), \quad \sum_a \alpha_a(u) = 0 \tag{3}$$

Here A = 6, $\alpha_a(u)$ is a fixed action contrast, $P_p(u)$ is an archive-code random effect that is consistent across the six labels, $(P\alpha)_{pa}(u)$ is the random code-specific deviation for a fixed action, and $\varepsilon_{par}(u)$ is unexplained within-cell dispersion. The action term was defined as finite-set action-contrast dispersion

$$D(A)(u) = (A-1)^{-1} \sum_a \alpha_a^2(u) \tag{4}$$

and was estimated in the balanced design by $\max\{[MS(A)(u) - MS(P \times A)(u)]/(PR), 0\}$, where P = 30 and R = 50. D(A) is a finite-set descriptive quantity defined here for the six specified labels, not a random action variance sampled from all table-tennis techniques [12]. The archive-code, archive-code × fixed-action and within-cell components were estimated from balanced mean squares and constrained to be non-negative. All four components were trapezoidally integrated over the unit interval across 200 nodes and divided by their integrated total. Pointwise proportions were descriptive. The 11-node moving average in Fig. 2 was used only for display; all estimates and integrals used unsmoothed node values.

Action-specific L^2 trace-type relative reliability

A separate random-intercept model was fitted for each action–representation combination:

$$Y_{pr}^{(a)}(u) = \mu_a(u) + P_p^{(a)}(u) + \varepsilon_{pr}^{(a)}(u) \tag{5}$$

The variance functions of $P_p^{(a)}(u)$ and $\varepsilon_{pr}^{(a)}(u)$ were $\sigma^2 P_{,a}(u)$ and $\sigma^2 e_{,a}(u)$. Node-level Gaussianity was a working assumption for REML. Variance components were estimated at each node by non-negative constrained REML; an interior solution in the balanced one-way random-effects model agrees with analytical REML/ANOVA [16,24,25]. When the archive-code component lay on the non-negative boundary, the constrained likelihood solution was retained and no conventional interior-point likelihood-ratio test was used [26].

Trapezoidal integration over the 200 nodes gave

$$B_a = \int_0^1 \sigma_{P,a}^2(u)\mathrm{d}u, \quad W_a = \int_0^1 \sigma_{e,a}^2(u)\mathrm{d}u \tag{6}$$

When trial residuals within an archive code have the same marginal covariance and zero or negligible cross-trial correlation, the L^2 trace-type relative-reliability coefficient for a mean of n trials is

$$R(L^2, a; n) = \frac{B_a}{B_a + \frac{W_a}{n}} \quad (7)$$

The trace of a covariance operator equals the expected squared L^2 distance of a random waveform from its mean, so this ratio is invariant to orthogonal basis transformations [13,14]. It adopts the error-reduction logic of a generalizability-theory decision study [12], but is not a standard 'functional generalizability coefficient' with an established set of random facets. It summarises the proportion of total L^2 variation in the fixed-action mean waveform attributable to between-code variation. Nodewise REML estimates variance functions and a trace summary; it is not a complete functional mixed model for curve prediction, functional principal-component inference or simultaneous bands [27]. The 0.80 and 0.90 values were post hoc reporting references, not universal technical standards. For target g, the continuous threshold is

$$n^*_{a,g} = \frac{gW_a}{(1-g)B_a} \quad (8)$$

The design count was the smallest positive integer satisfying the reference level; primary reporting used g = 0.90. Integrating B_a and W_a before taking their ratio is not equivalent to averaging pointwise reliabilities. As an estimand sensitivity analysis, relative reliability was also calculated at each phase node and then trapezoidally averaged over phase while integer n was searched. The trace ratio weights phases with greater variance more heavily, whereas the phase average gives equal weight to phase intervals. They answer different questions, as illustrated by the BD racket estimates of 27 and 17 trials.

Numeric archive-order diagnostic and residual-dependence scenarios

We verified that each archive-code–action cell contained 50 files with consecutive global indices, assigned archive ranks 1–50 by ascending index and searched for timestamps or mappings that could establish acquisition order. None were available, so archive rank was treated as an index proxy rather than an independently supplied trial number [15]. For each action and representation, residual waveforms were obtained by subtracting the within-code mean of 50 trials. L^2 trace-type correlation at lag k was then calculated as

$$\hat{\rho}_{\text{archive}}(a, k) = \frac{S(a,k)}{\sqrt{Q(a,0)Q(a,k)}} \quad (9)$$

where S(a,k) is the sum of integrated residual-waveform cross-products over all valid lagged pairs, and Q(a,0) and Q(a,k) are the corresponding sums of integrated squares at the two endpoints. Five thousand random permutations of trial order within each archive code placed the lag-1 value relative to a random archive ordering. This was a diagnostic of archive-order structure, not an estimator of temporal dependence.

Because actual acquisition order could not be recovered, we used AR(1) working scenarios with φ = 0, 0.10, 0.20 and 0.30. On the trace scale, the working model assumed $\text{trCov}(\varepsilon_r, \varepsilon_{r+k}) = \varphi^k W_a$, so each lag-covariance trace was proportional to the marginal error trace and shared φ. The trace of the mean residual covariance was

$$\mathrm{tr}\{\mathrm{Cov}(\bar{\varepsilon}_a)\} = \frac{W_a}{n}\left[1 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\varphi^k\right] \quad (10)$$

The finite-sample design effect in brackets replaced W_a/n in equation (7), and the smallest integer reaching 0.90 was found by search. The grid $\varphi = 0$–0.30 represented sensitivity from no to moderate positive dependence. It was informed by, but not fitted to, the archive-rank diagnostic. B_a and W_a were held fixed, so this scenario does not quantify any effect of correlation on variance-component estimation itself [15,16].

Single proxy-peak registration sensitivity

To examine internal temporal displacement, we applied endpoint-preserving piecewise-linear registration using one landmark. An 11-node moving average was applied to a Hampel-type quality-control copy of the racket waveform, and its maximum was located only within 10–90% of phase. The median peak node across 1,500 trials of an action was the reference. Each trial peak was mapped to the reference while preserving 0% and 100%, and the same time transformation was applied to the original racket and paired body waveforms without Hampel replacement. The landmark was therefore protected from a single retained high value, while registration and coordinate-quality sensitivities remained separate. All 12 models were refitted after registration. This analysis did not claim that the proxy peak was ball impact or designate the registered estimate as primary [7,8,28].

Bootstrap, archive-code influence and other sensitivity analyses

Primary uncertainty was assessed with 5,000 archive-code cluster bootstrap replicates. Each replicate sampled 30 archive-code clusters with replacement and retained all six actions and 50 trials per cell from each selected cluster, respecting the nested design [29,30]. We primarily reported 95% percentile intervals for continuous $n^*_{a,0.90}$ because they preserve the positive ratio scale and are invariant to monotone transformation. Basic and bias-corrected and accelerated (BCa) intervals based on an archive-code jackknife, quantile Monte Carlo standard errors, ranges across five non-overlapping 1,000-replicate blocks, counts of $B_a \leq 0$, node boundary estimates and undefined thresholds were also calculated [31–33]. Integer thresholds were ceiling values for design use.

Leave-one-archive-code-out analyses were performed under both the primary and Hampel-type rules, and each code's contribution to integrated within-cell sums of squares and between-code mean distance was recorded. The Hampel-type rule operated on derived frame-interval displacements before phase normalisation. For an interior node, at most five neighbouring values on each side, excluding the centre, formed the reference set; $m = \mathrm{median}(x_{ref})$ and $s = 1.4826\,\mathrm{median}|x_{ref} - m|$. A value was flagged only when $x_i > \max(m + 8s, 4m)$, then linearly interpolated from adjacent unflagged displacement values. Parameters were fixed before the sensitivity run and not tuned by action. The rule is not a device-validated artefact classifier [34–36].

Other analyses were: (1) unbalanced profile REML after excluding 1,180 trials whose nominal filename length exceeded 200; (2) 1,000 action-wise balanced samples without replacement from that exclusion set, using the minimum cell size; (3) complete deletion of body trials containing any structural zero; (4) fixed 13-landmark and fixed 8-landmark body waveforms; (5) the alternative 30-code window 11–40, all codes 1–40, and unbalanced REML after removing the codes 31–40 member of each of eight exact duplicate pairs; (6) parallel thresholds for the L^2 trace ratio and phase-averaged pointwise reliability; and (7) paired archive-code cluster bootstrap contrasts of continuous thresholds between actions. Code sensitivities were not used to claim that 40 codes represented 40 independent participants. Action contrasts were unadjusted for multiplicity and were retained as descriptive machine-readable supplements.

Software, reproducibility and generative-AI assistance

Analyses were implemented in Python, principally using pandas, Matplotlib, NumPy and SciPy [37–40]. The random seed was 20260712 for primary analyses and 20260713 for added assumption and influence diagnostics. Custom variance-component code passed dense-likelihood equivalence tests on synthetic data and analytical-solution checks for the complete balanced models; all 13 numerical self-checks in the original workflow passed. The cold-reproduction environment was Python 3.12.13, NumPy 2.3.5, pandas 2.2.3, SciPy 1.18.0, Matplotlib 3.11.0 and Pillow 12.2.0. Every figure is accompanied by separate CSV source data and Python plotting code, with editable SVG, PDF, 300-dpi PNG and 600-dpi LZW-TIFF exports.

OpenAI Codex was used in July 2026 to edit language and formatting and to help revise explanatory passages. The authors defined the scientific claims and checked the revised passages against the reported analyses, source materials and cited literature. All methodological decisions, code execution, numerical verification, interpretation and final approval remained the responsibility of the authors. Generative AI was not listed as an author and did not replace the authors' scientific judgement.

Code availability

All custom analysis and plotting code, input-identity checks, environment lock files, archive-boundary and exact-de-duplication audits, 14-label mapping, functional-summary sensitivity analyses and publishable derived results are available at https://github.com/hmk-228674/TTMD6-functional-G-theory-analysis. The frozen release corresponding to the quantitative results and figures in this manuscript is v1.0.2, archived at Zenodo under version DOI https://doi.org/10.5281/zenodo.21599397; the concept DOI for all versions is https://doi.org/10.5281/zenodo.21382966 [41]. Historical releases v1.0.0 and v1.0.1 remain immutable, but their visible manuscript terminology or figure exports do not correspond to this final version. The repository does not redistribute raw TTMD6 coordinates. Reproduction requires the byte-identical archive

from the Springer Nature address in Data Availability and verification of its byte count, MD5 and SHA-256 before analysis.

Data availability

TTMD6 is a third-party dataset published by Zhang et al. [19]. The exact archive analysed here is available from the supplementary material hosted by Springer Nature for the source article: https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-024-54150-5/MediaObjects/41598_2024_54150_MOESM1_ESM.rar. The official file is 341,074,031 bytes; MD5 1c9ce9cbf79dd35dd22f16a7199e2a8c; SHA-256 93d1b52a470f14b9dc0ba0600959bff921be891a3da1b71e609bd328224b354d. We downloaded the complete file and confirmed bytewise identity with the analysis input. The third-party raw coordinates are not redistributed. A later figshare v2 record [23] differs in byte count and was not analysed. Analysis code, identity checks, archive-code boundary and de-duplication audits, 14-label mapping, functional-summary sensitivities, derived results, bootstrap diagnostics, figure source data and the exact software environment are available at https://github.com/hmk-228674/TTMD6-functional-G-theory-analysis. The frozen release for this manuscript is v1.0.2 (https://doi.org/10.5281/zenodo.21599397); the DOI for all versions is https://doi.org/10.5281/zenodo.21382966. The byte-identical source archive should be obtained from the publisher address above.

References

1. Forrester, S. E. Selecting the number of trials in experimental biomechanics studies. *Int. Biomech.* **2**, 62–72 (2015). https://doi.org/10.1080/23335432.2015.1049296

2. Blinch, J., Kim, Y. & Chua, R. Trajectory analysis of discrete goal-directed pointing movements: how many trials are needed for reliable data? *Behav. Res. Methods* **50**, 2162–2172 (2018). https://doi.org/10.3758/s13428-017-0983-6

3. Oliveira, A. S. & Pirscoveanu, C. I. Implications of sample size and acquired number of steps to investigate running biomechanics. *Sci. Rep.* **11**, 3083 (2021). https://doi.org/10.1038/s41598-021-82876-z

4. Severin, A. C., Barnes, S. G., Tackett, S. A., Barnes, C. L. & Mannen, E. M. The required number of trials for biomechanical analysis of a golf swing. *Sports Biomech.* **20**, 238–246 (2021). https://doi.org/10.1080/14763141.2018.1554085

5. Lazic, S. E. The problem of pseudoreplication in neuroscientific studies: is it affecting your analysis? *BMC Neurosci.* **11**, 5 (2010). https://doi.org/10.1186/1471-2202-11-5

6. Ramsay, J. O. & Silverman, B. W. *Functional Data Analysis 2nd edn* (Springer, 2005). https://doi.org/10.1007/b98888

7. Helwig, N. E., Hong, S., Hsiao-Wecksler, E. T. & Polk, J. D. Methods to temporally align gait cycle data. *J. Biomech.* **44**, 561–566 (2011). https://doi.org/10.1016/j.jbiomech.2010.09.015

8. Marron, J. S., Ramsay, J. O., Sangalli, L. M. & Srivastava, A. Functional data analysis of amplitude and phase variation. *Stat. Sci.* **30**, 468–484 (2015). https://doi.org/10.1214/15-STS524

9. Warmenhoven, J. et al. PCA of waveforms and functional PCA: a primer for biomechanics. *J. Biomech.* **116**, 110106 (2021). https://doi.org/10.1016/j.jbiomech.2020.110106

10. Atkinson, G. & Nevill, A. M. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med.* **26**, 217–238 (1998). https://doi.org/10.2165/00007256-199826040-00002

11. Cronbach, L. J., Rajaratnam, N. & Gleser, G. C. Theory of generalizability: a liberalization of reliability theory. *Br. J. Stat. Psychol.* **16**, 137–163 (1963). https://doi.org/10.1111/j.2044-8317.1963.tb00206.x

12. Brennan, R. L. *Generalizability Theory* (Springer, 2001). https://doi.org/10.1007/978-1-4757-3456-0

13. Di, C.-Z., Crainiceanu, C. M., Caffo, B. S. & Punjabi, N. M. Multilevel functional principal component analysis. *Ann. Appl. Stat.* **3**, 458–488 (2009). https://doi.org/10.1214/08-AOAS206

14. Hsing, T. & Eubank, R. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators* (Wiley, 2015). https://doi.org/10.1002/9781118762547

15. Brockwell, P. J. & Davis, R. A. *Introduction to Time Series and Forecasting 3rd edn* (Springer, 2016). https://doi.org/10.1007/978-3-319-29854-2

16. Pinheiro, J. C. & Bates, D. M. *Mixed-Effects Models in S and S-PLUS* (Springer, 2000). https://doi.org/10.1007/b98882

17. Bańkosz, Z. & Winiarski, S. Kinematic parameters of topspin forehand in table tennis and their inter- and intra-individual variability. *J. Sports Sci. Med.* **19**, 138–148 (2020).

18. Wong, D. W.-C., Lee, W. C.-C. & Lam, W.-K. Biomechanics of table tennis: a systematic scoping review of playing levels and maneuvers. *Appl. Sci.* **10**, 5203 (2020). https://doi.org/10.3390/app10155203

19. Zhang, J., Ren, Y., Lin, L., Xing, Y. & Ren, J. Table tennis motion recognition based on the bat trajectory using varying-length-input convolution neural networks. *Sci. Rep.* **14**, 3549 (2024). https://doi.org/10.1038/s41598-024-54150-5

20. Bartlett, R., Wheat, J. & Robins, M. Is movement variability important for sports biomechanists? *Sports Biomech.* **6**, 224–243 (2007). https://doi.org/10.1080/14763140701322994

21. Preatoni, E. et al. Movement variability and skills monitoring in sports. *Sports Biomech.* **12**, 69–92 (2013). https://doi.org/10.1080/14763141.2012.738700

22. van der Kruk, E. & Reijne, M. M. Accuracy of human motion capture systems for sport applications: state-of-the-art review. *Eur. J. Sport Sci.* **18**, 806–819 (2018). https://doi.org/10.1080/17461391.2018.1463397

23. Zhang, J. Table Tennis MOCAP Dataset, version 2. *figshare* (2026). https://doi.org/10.6084/m9.figshare.31746358.v2

24. Patterson, H. D. & Thompson, R. Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554 (1971). https://doi.org/10.1093/biomet/58.3.545

25. Harville, D. A. Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Stat. Assoc.* **72**, 320–338 (1977). https://doi.org/10.1080/01621459.1977.10480998

26. Self, S. G. & Liang, K.-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Stat. Assoc.* **82**, 605–610 (1987). https://doi.org/10.1080/01621459.1987.10478472

27. Morris, J. S. & Carroll, R. J. Wavelet-based functional mixed models. *J. R. Stat. Soc. B* **68**, 179–199 (2006). https://doi.org/10.1111/j.1467-9868.2006.00539.x

28. Ramsay, J. O. & Li, X. Curve registration. *J. R. Stat. Soc. B* **60**, 351–363 (1998). https://doi.org/10.1111/1467-9868.00129

29. Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap* (Chapman & Hall, 1993).

30. Field, C. A. & Welsh, A. H. Bootstrapping clustered data. *J. R. Stat. Soc. B* **69**, 369–390 (2007). https://doi.org/10.1111/j.1467-9868.2007.00593.x

31. Efron, B. Better bootstrap confidence intervals. *J. Am. Stat. Assoc.* **82**, 171–185 (1987). https://doi.org/10.1080/01621459.1987.10478410

32. Ukoumunne, O. C., Davison, A. C., Gulliford, M. C. & Chinn, S. Non-parametric bootstrap confidence intervals for the intraclass correlation coefficient. *Stat. Med.* **22**, 3805–3821 (2003). https://doi.org/10.1002/sim.1643

33. Andrews, D. W. K. & Buchinsky, M. A three-step method for choosing the number of bootstrap repetitions. *Econometrica* **68**, 23–51 (2000). https://doi.org/10.1111/1468-0262.00092

34. Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. & Stahel, W. A. *Robust Statistics: The Approach Based on Influence Functions* (Wiley, 2005). https://doi.org/10.1002/9781118186435

599 35. Rousseeuw, P. J. & Croux, C. Alternatives to the median absolute deviation. *J. Am. Stat. Assoc.* **88**,
600 1273–1283 (1993). https://doi.org/10.1080/01621459.1993.10476408

601 36. Pearson, R. K. *Mining Imperfect Data: Dealing with Contamination and Incomplete Records* (SIAM,
602 2005). https://doi.org/10.1137/1.9780898717884

603 37. McKinney, W. Data structures for statistical computing in Python. In *Proc. 9th Python in Science*
604 *Conference* 56–61 (2010). https://doi.org/10.25080/Majora-92bf1922-00a

605 38. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
606 https://doi.org/10.1109/MCSE.2007.55

607 39. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).
608 https://doi.org/10.1038/s41586-020-2649-2

609 40. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods*
610 **17**, 261–272 (2020). https://doi.org/10.1038/s41592-019-0686-2

611 41. Han, M. TTMD6 waveform relative-reliability analysis, version 1.0.2. *Zenodo* (2026).
612 https://doi.org/10.5281/zenodo.21599397

613 Acknowledge

614 The authors acknowledge the researchers who developed, curated, and made the TTMD6 dataset publicly
615 available. The availability of this dataset enabled the present secondary analysis and supports cumulative
616 and reproducible research in table tennis biomechanics.

617 Funding

618 This research received no specific funding.

619 Author contributions

620 M.H. and Y.S. jointly conceived and advanced the study, contributed equally to the work and share first
621 authorship. Y.S. and L.W. formulated the research question. M.H. designed and performed the statistical
622 analyses and drafted the manuscript. Y.S. contributed to refinement of the analytical framework,
623 interpretation of the results and substantive revision of the manuscript. C.D. processed the data and verified
624 the code and results. L.M., P.J., C.Y. and L.W. contributed to interpretation of the results and critically
625 revised the manuscript. All authors reviewed and approved the final manuscript and agree to be accountable
626 for the work.

Additional Information

Competing interests

The authors declare no competing interests.

Figure legends

Figure 1 | Study design, archive audit, estimand and inferential boundaries. Thirty archive codes were treated as independent clusters under an explicit but unverified working assumption; the six English action labels were fixed conditions, and each archive-code–action cell contained 50 archive trials. The diagram follows the analysis from archive audit and operational event-unregistered waveforms to the finite-set background decomposition and 12 action-specific models. Cluster bootstrap, residual-dependence scenarios, registration and quality-rule sensitivities are shown in parallel. All trial counts are restricted to within-session relative distinguishability.

Figure 2 | Pointwise dispersion profiles and integrated proportions for the six fixed labels. a, Racket displacement magnitude. b, Aggregated-body displacement magnitude. c, Integrated proportions of the four dispersion components. Blue denotes the noise-corrected finite-set action-contrast dispersion D(A), yellow the across-action archive-code component, orange the archive-code × fixed-action component and grey unexplained within-cell dispersion. The 11-node moving averages in a and b are for display only; all integrals and statistics use unsmoothed node estimates. Curves do not represent local significance.

Figure 3 | Action-specific L^2 trace-type relative reliability and trial count. a,b, $R(L^2,a;n) = B_a/[B_a + W_a/n]$ for racket and aggregated-body waveforms under negligible trial-residual dependence. Horizontal lines mark the 0.80 and 0.90 summary references and the vertical line marks ten trials. c, Continuous threshold n^*_a,0.90 and 95% percentile intervals from 5,000 archive-code cluster bootstrap replicates. Superscript (0) indicates the $\varphi = 0$ working scenario; action subscripts are omitted in the figure for clarity. FA, forehand attack; FD, forehand drive; FP, forehand push; BA, backhand attack; BD, backhand drive; BP, backhand push.

Figure 4 | Key sensitivities to residual dependence, peak registration and the BD quality rule. a, Lag-1 L^2 trace-type correlation by within-cell archive rank. The grey envelope spans the smallest 2.5% lower bound to the largest 97.5% upper bound among the 12 condition-specific permutation distributions and is not a joint confidence band; archive rank is not verified acquisition order. b, Minimum integer count reaching 0.90 under AR(1) working scenarios $\varphi = 0$, 0.10, 0.20 and 0.30; R denotes racket and H aggregated body. c, Continuous thresholds before and after robust racket proxy-peak registration; registration changes the estimand and is not a truth correction. d, Continuous BD racket thresholds under the primary and Hampel-

type rules after leaving out one archive code at a time. Point area is proportional to the number of flagged values for that code; dashed lines show the complete 30-code estimates.

Tables

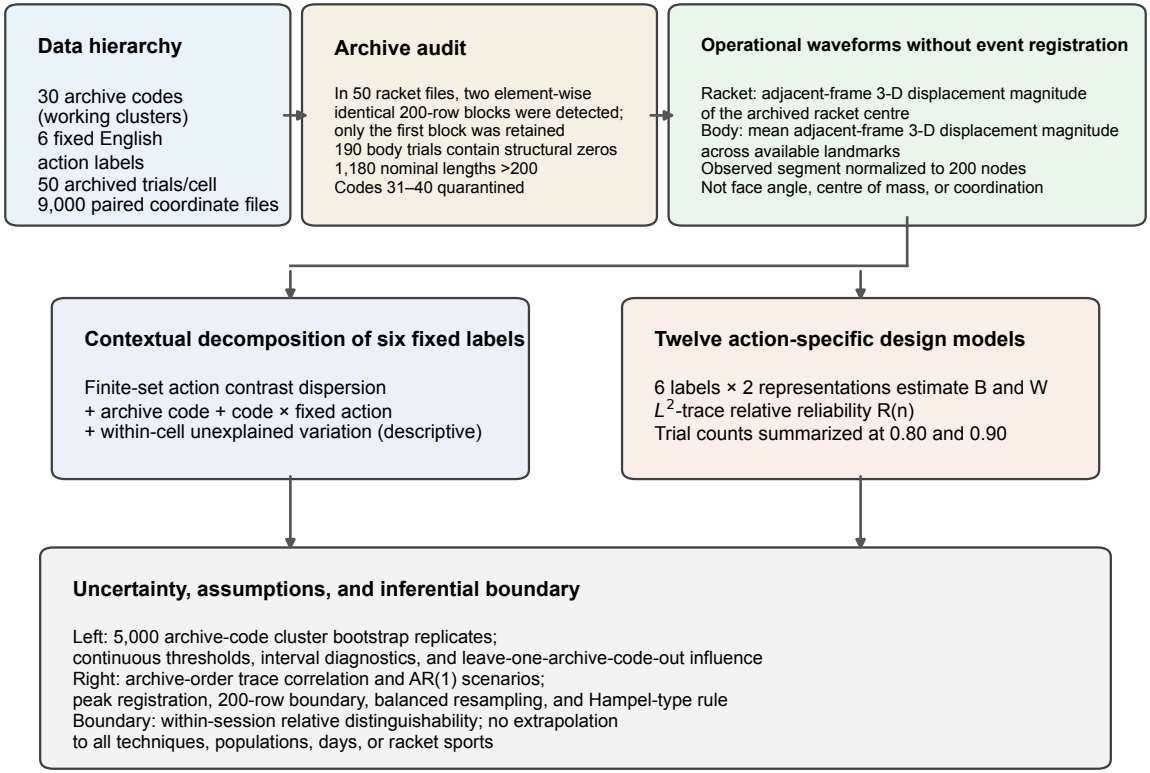
Table 1 | Data hierarchy, action labels, waveforms and primary inferential boundaries

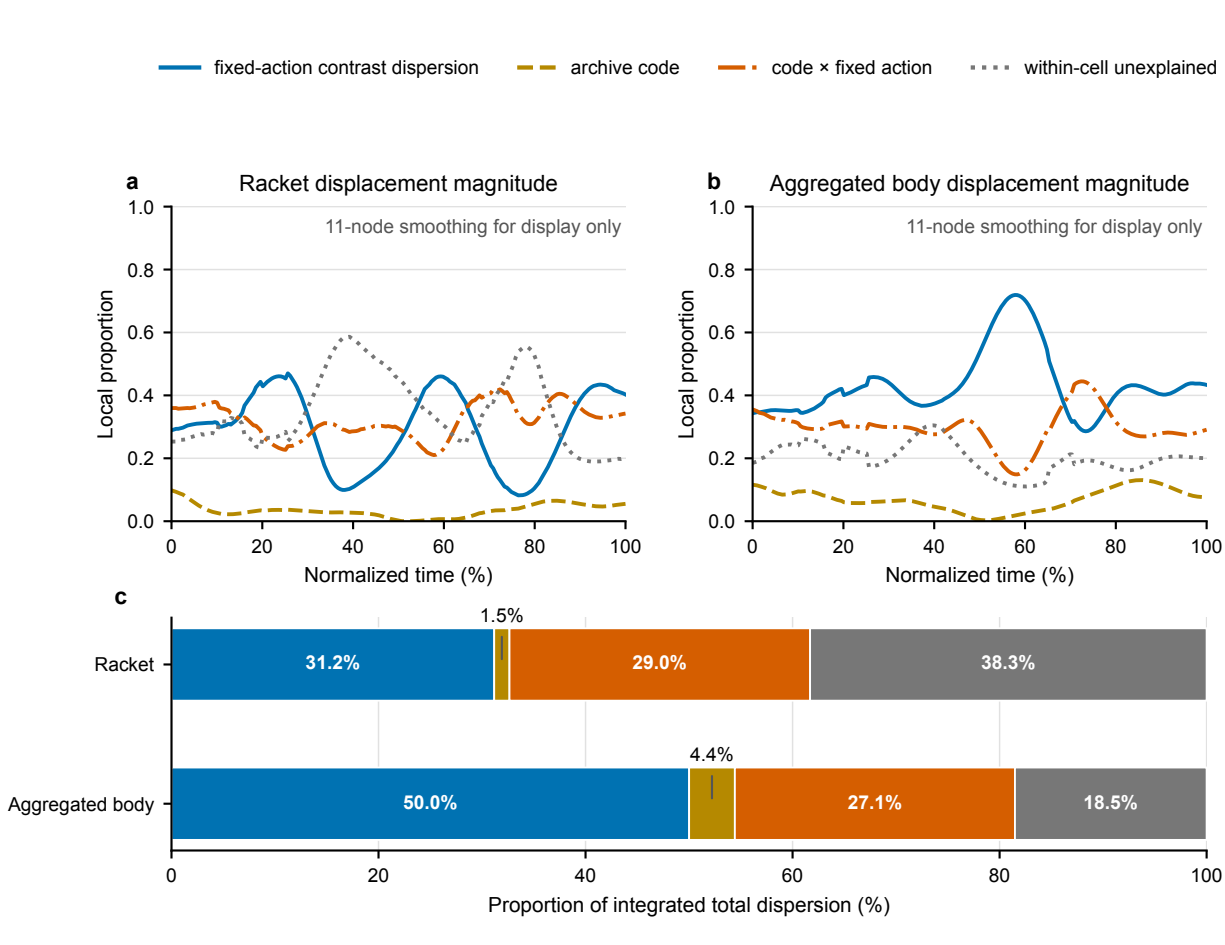
Item	Operational definition	Outside the present estimand
Independent unit	Primary analysis clustered 30 archive codes; each code contained six labels × 50 trials	A source-author-confirmed one-to-one participant identity or 9,000 independent samples
Fixed labels	FA forehand attack; FD forehand drive; FP forehand push; BA backhand attack; BD backhand drive; BP backhand push	Fully trial-validated technical constructs; the source paper validated only five racket animations per class
Racket waveform	Three-dimensional displacement magnitude of the racket centre between adjacent frames	Racket-face orientation, impact impulse or complete three-dimensional technique
Aggregated-body waveform	Equal mean displacement magnitude across the 14 published landmarks available at both adjacent frames	Centre of mass, joint angle, coordination, torque or power
Time processing	Observable archive interval linearly normalised to 200 nodes; no event registration in the primary analysis	An impact-aligned cycle or pure amplitude variation
D(A)	Noise-corrected finite-set action-contrast dispersion among the six fixed labels	Random action variance sampled from all table-tennis techniques
$R(L^2,a;n)$	Within-session L^2 trace-type relative distinguishability of a mean waveform under a fixed label	A standard universal functional G coefficient, between-day absolute agreement or classification accuracy
0.80/0.90	Reference levels for reporting and design comparison	Technical competence, training or clinical thresholds
Primary count assumption	Negligible trial-residual dependence within a cell	Independence verified from true acquisition order
Generalisability	Current TTMD6 right-handed, single-institution cohort and 50 public trials	Left-handed or novice players, competition, all table-tennis actions or other racket sports

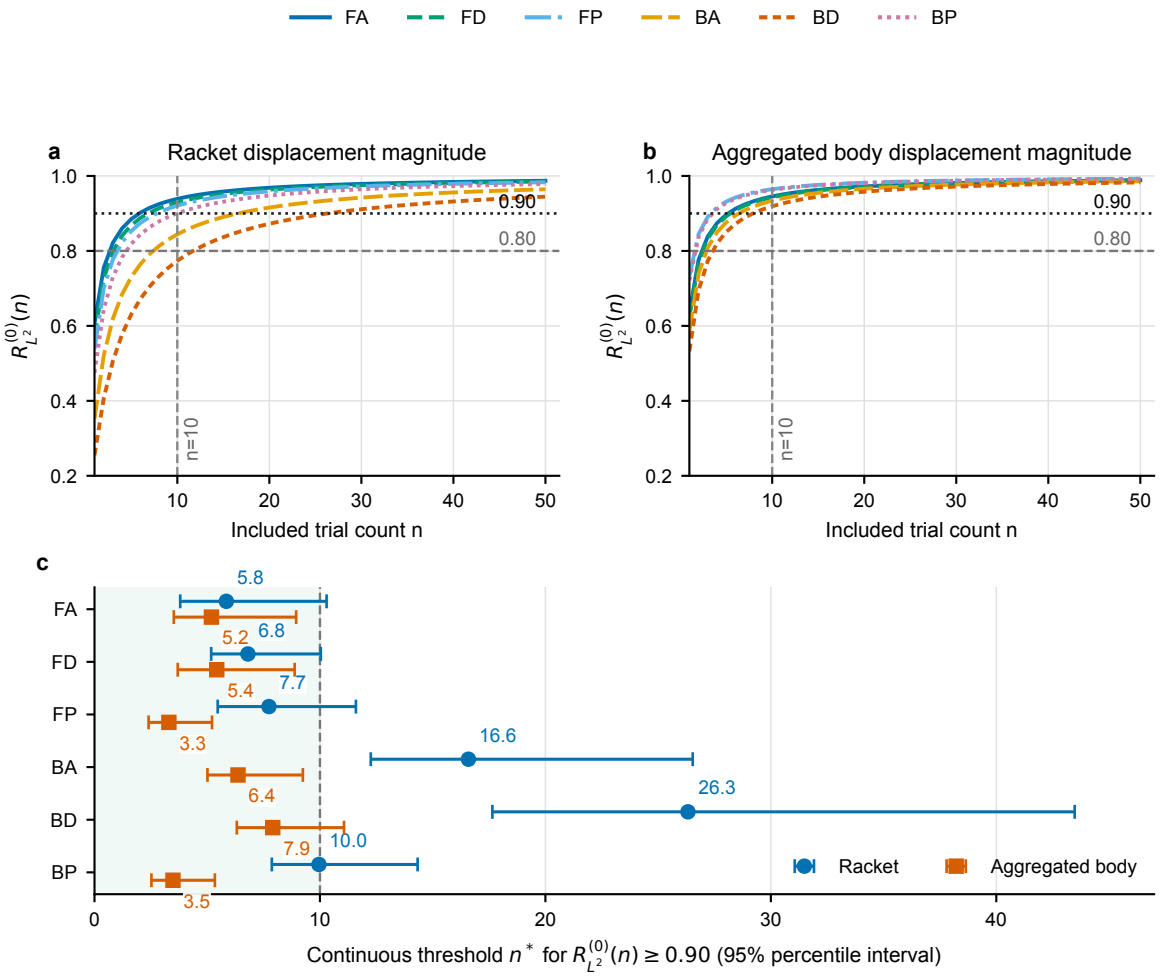
Table 2 | Action-specific integrated variances and trial-count thresholds under negligible trial-residual dependence

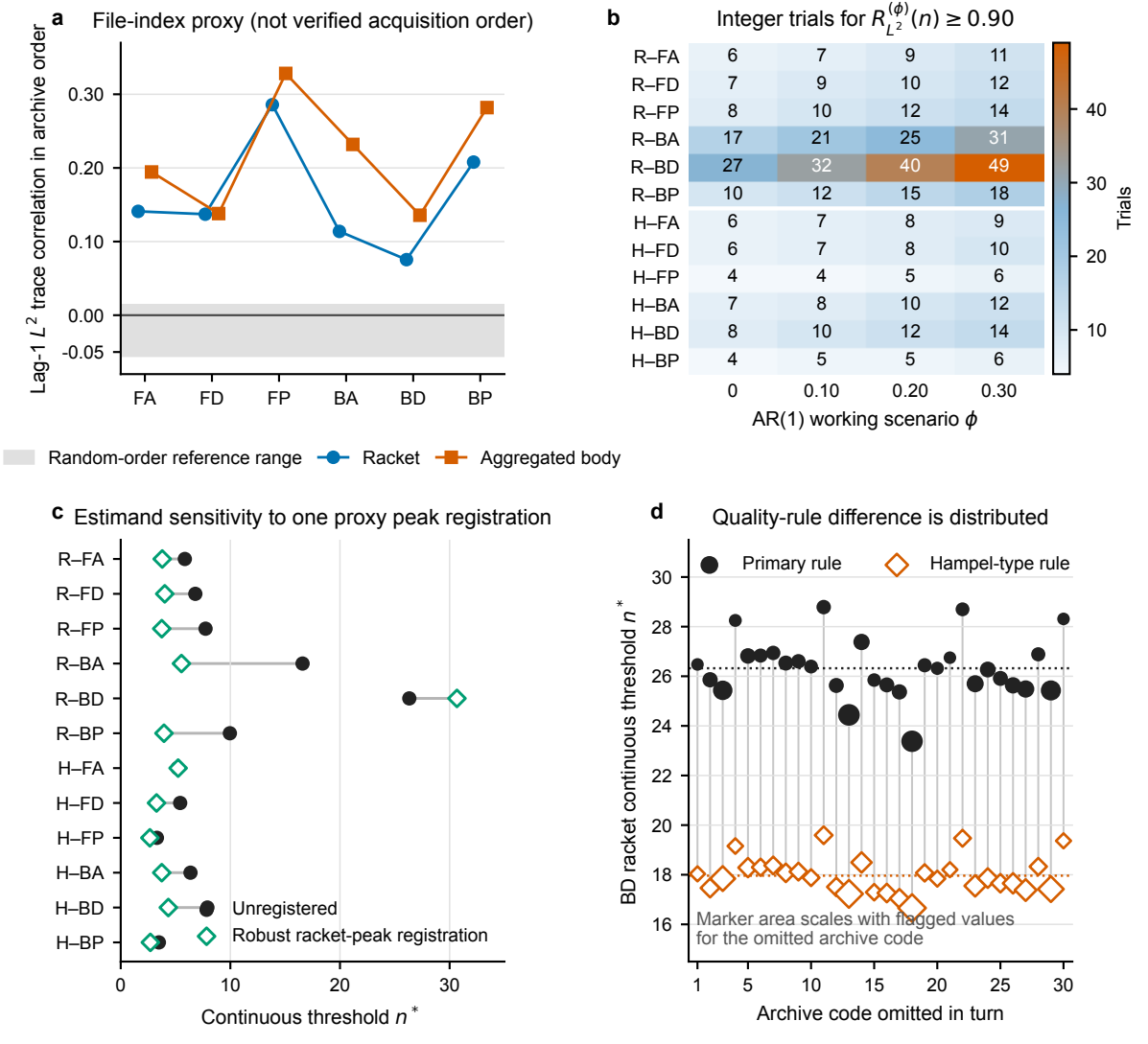
Representation	Label	B_a	W_a	W_a/B_a	Continuous trace $n^*_{0.90}$	Trace integer	Pointwise-mean integer	Trace 95% interval
Racket	FA	56.546	36.740	0.650	5.85	6	7	3.80–10.29
Racket	FD	183.268	138.580	0.756	6.81	7	8	5.18–10.04
Racket	FP	17.479	15.027	0.860	7.74	8	9	5.47–11.60
Racket	BA	26.309	48.484	1.843	16.59	17	16	12.26–26.53
Racket	BD	72.758	212.798	2.925	26.32	27	17	17.65–43.48
Racket	BP	16.410	18.154	1.106	9.96	10	11	7.87–14.33
Aggregated body	FA	1.154	0.665	0.576	5.19	6	6	3.52–8.94
Aggregated body	FD	4.687	2.824	0.602	5.42	6	6	3.69–8.88
Aggregated body	FP	1.432	0.525	0.367	3.30	4	4	2.40–5.21
Aggregated body	BA	0.648	0.458	0.707	6.36	7	7	5.01–9.24
Aggregated body	BD	1.410	1.239	0.878	7.91	8	8	6.31–11.07
Aggregated body	BP	1.101	0.426	0.387	3.48	4	5	2.52–5.34

B_a and W_a are the between-archive-code and within-code trial variances of the primary unregistered waveforms, trapezoidally integrated over 200 nodes. The continuous trace threshold is $n^*_{0.90} = 9W_a/B_a$ and the trace integer is its ceiling. The pointwise-mean integer first calculates pointwise relative reliability and then averages it over phase by trapezoidal integration. Intervals are 2.5% and 97.5% percentiles from 5,000 archive-code cluster bootstrap replicates and apply only to the continuous trace threshold. The integer columns are different functional summaries, both conditional on negligible residual dependence; neither is a general field prescription.


Data hierarchy
30 archive codes
(working clusters)
6 fixed English
action labels
50 archived trials/cell
9,000 paired coordinate files
Archive audit
In 50 racket files, two element-wise
identical 200-row blocks were detected;
only the first block was retained
190 body trials contain structural zeros
1,180 nominal lengths >200
Codes 31–40 quarantined
Operational waveforms without event registration
Racket: adjacent-frame 3-D displacement magnitude
of the archived racket centre
Body: mean adjacent-frame 3-D displacement magnitude
across available landmarks
Observed segment normalized to 200 nodes
Not face angle, centre of mass, or coordination
Contextual decomposition of six fixed labels
Finite-set action contrast dispersion
+ archive code + code × fixed action
+ within-cell unexplained variation (descriptive)
Twelve action-specific design models
6 labels × 2 representations estimate B and W
L^2-trace relative reliability R(n)
Trial counts summarized at 0.80 and 0.90
Uncertainty, assumptions, and inferential boundary
Left: 5,000 archive-code cluster bootstrap replicates;
continuous thresholds, interval diagnostics, and leave-one-archive-code-out influence
Right: archive-order trace correlation and AR(1) scenarios;
peak registration, 200-row boundary, balanced resampling, and Hampel-type rule
Boundary: within-session relative distinguishability; no extrapolation
to all techniques, populations, days, or racket sports


fixed-action contrast dispersion
archive code
code × fixed action
within-cell unexplained
a
Racket displacement magnitude
11-node smoothing for display only
Local proportion
Normalized time (%)
b
Aggregated body displacement magnitude
11-node smoothing for display only
Local proportion
Normalized time (%)
c
Racket
31.2%
1.5%
29.0%
38.3%
Aggregated body
50.0%
4.4%
27.1%
18.5%
Proportion of integrated total dispersion (%)


FA
FD
FP
BA
BD
BP
a
Racket displacement magnitude
b
Aggregated body displacement magnitude
$R^{(0)}_{L^2}(n)$
1.0
0.8
0.6
0.4
0.2
0.90
0.80
n=10
10
20
30
40
50
Included trial count n
c
FA
FD
FP
BA
BD
BP
5.8
5.2
6.8
5.4
7.7
3.3
16.6
6.4
26.3
7.9
10.0
3.5
Racket
Aggregated body
0
10
20
30
40
Continuous threshold n^* for $R^{(0)}_{L^2}(n) \geq 0.90$ (95% percentile interval)


a
File-index proxy (not verified acquisition order)
Lag-1 L^2 trace correlation in archive order
0.30
0.20
0.10
0.00
-0.05
FA
FD
FP
BA
BD
BP
Random-order reference range
Racket
Aggregated body
b
Integer trials for $R^{(\phi)}_{L^2}(n) \geq 0.90$
R–FA 6 7 9 11
R–FD 7 9 10 12
R–FP 8 10 12 14
R–BA 17 21 25 31
R–BD 27 32 40 49
R–BP 10 12 15 18
H–FA 6 7 8 9
H–FD 6 7 8 10
H–FP 4 4 5 6
H–BA 7 8 10 12
H–BD 8 10 12 14
H–BP 4 5 5 6
0 0.10 0.20 0.30
AR(1) working scenario ϕ
Trials
40
30
20
10
c
Estimand sensitivity to one proxy peak registration
R–FA
R–FD
R–FP
R–BA
R–BD
R–BP
H–FA
H–FD
H–FP
H–BA
H–BD
H–BP
Unregistered
Robust racket-peak registration
0
10
20
30
Continuous threshold n^*
d
Quality-rule difference is distributed
Primary rule
Hampel-type rule
BD racket continuous threshold n^*
30
28
26
24
22
20
18
16
Marker area scales with flagged values
for the omitted archive code
1
5
10
15
20
25
30
Archive code omitted in turn

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf