

Gap-Entropy Testing (GET): A Framework for Structural and Distributional Differences

Vangelis D. Karalis^{1,2}

¹ Department of Pharmacy, School of Health Sciences, National and Kapodistrian University of Athens, 15784 Athens, Greece; vkarakis@pharm.uoa.gr

² Institute of Applied and Computational Mathematics, Foundation for Research and Technology Hellas (FORTH), 70013 Heraklion, Greece

How To Cite: Karalis, V.D. Gap-Entropy Testing (GET): A Framework for Structural and Distributional Differences. *Applied Mathematics and Statistics* 2026, 3(2), 19. <https://doi.org/10.53941/ams.2026.100019>

Received: 12 May 2026
 Revised: 5 September 2026
 Accepted: 14 September 2026
 Published: 18 September 2026

Abstract: The classical two-sample problem is framed as the detection of differences in marginal distributions, in terms of location, scale, or overall shape. In this paper, Gap-Entropy Testing (GET) is proposed, a novel nonparametric framework that extends the two-sample paradigm by adding a spacing-sensitive component to conventional distributional analysis. The approach is based on order statistics, $\log 1p$ -transformed adjacent spacings normalized by their median gap, and entropy estimation using a Vasicek-type estimator. Inference is performed by permutation testing under a common continuous i.i.d. null. Three complementary statistics were developed: GET-V compares spacing-entropy estimates; GET-Plus combines spacing and median-location components; and GET- Ω additionally incorporates energy distance. Under the common i.i.d. null, GET-V has a type-I error close to the nominal 0.05 level and is insensitive to a pure location shift. For the lattice and micro-cluster example, GET-V has large rejection rates, which is related to the spacing regularity and the organization of the local gap. GET-V is also sensitive to marginal shape, and individual-label permutation is anti-conservative under matched lattice and shifted-lattice nulls. Application to public benchmark datasets allows us to use GET-V for continuous variables without ties. GET is a reproducible and interpretable framework that adds spacing information to standard two-sample comparisons.

Keywords: statistical comparison; spacing-based method; spacing-sensitive analysis; permutation testing; entropy estimation

1. Introduction

The comparison of two samples has been a key issue in statistics and has evolved with the advent of statistical thinking. Early approaches focused on parametric inference with Student's t -test [1] to find differences in means when sampling is used. While these methods are good when the assumptions are correct, they perform poorly with non-normality, heteroscedasticity, or outliers [2,3]. To address these problems, nonparametric approaches were developed [4]. Rank-based approaches such as the Mann-Whitney test [5] consider the order of the observations but do not require a parametric distribution. Empirical-distribution function methods such as the Kolmogorov-Smirnov test allow the comparison beyond the location to determine the differences in shape, spread, and tails [6–9]. In recent years, the focus has shifted to more general measures of between-distribution discrepancy. Distance and energy-based methods measure the difference across metric spaces and usually pairwise distances between observations [10,11]. These methods can be applied to multivariate and high-dimensional cases and can measure the relationship between the dependence structure and higher-order features [12]. Kernel-based methods have also been used to compare distributional differences further and more generally [13]. The adjacent spacings are a



complementary representation of a sample and emphasize local regularity, clustering, and gap heterogeneity that location summaries do not. In the usual i.i.d. model, their distribution is determined by the parent distribution, so the spacing effect should be taken into account as well as marginal shape. This leads to a separate spacing statistic that can be investigated in detail.

The Gap-Entropy Testing (GET) is proposed as a unified framework (Figure 1) based on entropy summaries of transformed adjacent gaps [14,15]. GET-V is the spacing-sensitive component, GET-Plus combines spacing and median-location components, and GET- Ω additionally incorporates energy distance [10]. The permutation of labels is useful to calibrate the framework under a common i.i.d. exchangeability null [16,17]. GET-V is invariant to translation and positive rescaling, so it has a distinct place to be tested in addition to location and overall distributional difference.

Three GET statistics are defined and their behavior is explored in a common i.i.d. null, constructed spacing alternatives, i.i.d. shape alternatives, and matched dependent processes. It is then demonstrated their use with public benchmark data. The analysis presents a reproducible approach and the conditions to take into account in the application of two-sample comparisons.

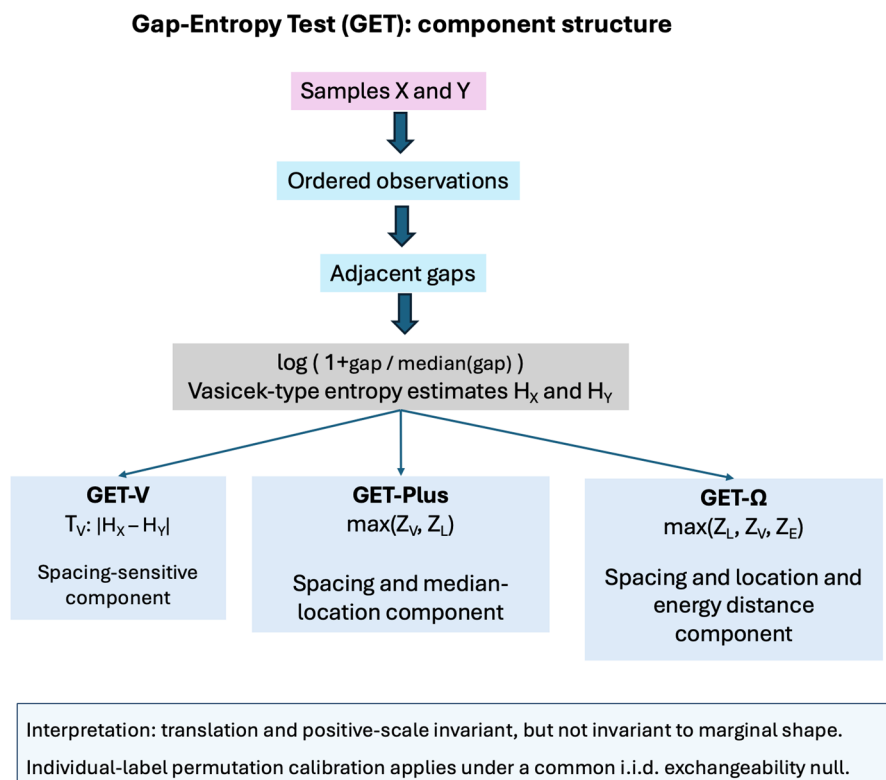


Figure 1. Conceptual overview of the GET framework. GET-V compares spacing-entropy estimates, GET-Plus adds a median-location component, and GET- Ω also adds energy distance. The components are calibrated by permutation under the common i.i.d. exchangeability null.

2. Methods

2.1. The GET Concept

2.1.1. Problem Setting

Let $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$ denote two samples of real-valued observations. The formal permutation null assumes that all observations are exchangeable under a common continuous i.i.d. distribution F . Under this null, relabeling pooled observations leaves their joint distribution unchanged. The constructed lattice and clustered scenarios are stress-test alternatives to this null; they are not used to claim general validity for comparing arbitrary dependent point processes.

$$H_0: F = G \quad (1)$$

Formal permutation inference is based on the common i.i.d. null, while M2–M4 evaluate sensitivity to specified gap patterns.

Assumptions

GET-V and GET-Plus do not require a parametric family. The energy-distance component used in GET- Ω is interpreted under its usual finite-first-moment condition. Individual-label permutation requires exchangeability under the null, and GET-V additionally requires continuous measurements without ties. In the implementation used here, gaps are bounded below by 10^{-12} for numerical stability; because this convention can dominate results for discrete or rounded variables, such variables are classified as ineligible for GET-V inference rather than analyzed through arbitrary jittering.

2.1.2. Order Statistics and Spacing Representation

For each sample, the observations are ordered before adjacent gaps are computed [14]:

$$D_i^x = X_{(i+1)} - X_{(i)}, \quad i = 1, \dots, n_X - 1$$

$$D_j^y = Y_{(j+1)} - Y_{(j)}, \quad j = 1, \dots, n_Y - 1$$

Adjacent gaps are normalized by the sample median gap. This definition, rather than the median of the raw observations, ensures positive-scale invariance.

$$\tilde{D}_i^x = D_i^x / \text{median}(D^x) \quad (2)$$

$$\tilde{D}_j^y = D_j^y / \text{median}(D^y) \quad (3)$$

The implemented transformation is $\log 1p(\text{gap}/\text{median gap})$, which is finite for positive gaps and reduces the influence of exceptionally large normalized gaps.

$$Z_i^x = \log(1 + \tilde{D}_i^x) \quad (4)$$

$$Z_j^y = \log(1 + \tilde{D}_j^y) \quad (5)$$

The gap construction is translation invariant and median-gap normalization is invariant to positive rescaling. Neither operation removes the influence of skewness, tail weight, multimodality, or other marginal-shape features.

2.1.3. Entropy Estimation

For each sample, the transformed gaps are sorted and summarized by a Vasicek-type entropy estimator [15,18]. The implementation uses window $m = \text{floor}(\sqrt{n_g})$, where n_g is the number of transformed gaps, clipped when necessary, so that $2m < n_g$. Boundary indices are truncated to the observed range, and numerical spacings are bounded below by 10^{-12} .

For the ordered transformed gaps $Z_{(1)}, \dots, Z_{(n_g)}$, the implemented estimator is:

$$\hat{H}(Z) = (1/n_g) \sum_i \log[(n_g/(2m))(Z_{(i+m)} - Z_{(i-m)})] \quad (6)$$

$m = \text{floor}(\sqrt{n_g})$, clipped when necessary, so that $2m < n_g$.

2.1.4. GET-V

Let H_X and H_Y denote the two spacing-entropy estimates. GET-V is based on their absolute difference:

$$T_V = |\hat{H}_X - \hat{H}_Y| \quad (7)$$

Here, $T_V = |H_X - H_Y|$. GET-V is spacing-sensitive and translation/positive-scale invariant, but it is not marginal-shape invariant. Its p -value is obtained from the permutation distribution of T_V under the common i.i.d. exchangeability null.

2.1.5. GET-Plus

The location component is $T_L = |\text{median}(X) - \text{median}(Y)|$.

$$T_L = |\text{median}(X) - \text{median}(Y)|$$

For GET-Plus and GET- Ω , each observed component is standardized relative to its own permutation distribution using the permutation median and the normal-consistent median absolute deviation. A standard deviation fallback is used if the MAD is numerically zero.

$$Z_C = [T_{C(\text{obs})} - \text{median}\{T_{C(\text{perm})}\}]/\text{MAD}_{\text{normal}\{T_{C(\text{perm})}\}} \quad (8)$$

For a component T , the standardized score compares the observed statistic with the center and spread of its permutation values. The same center and spread are applied to the permuted component values used to calibrate the composite maximum.

Then, the GET-Plus statistic is defined as:

$$T_+ = \max(Z_V, Z_L) \quad (9)$$

This formulation combines the component scores on a common permutation-standardized scale. GET-Plus is interpreted through both its spacing and median-location components rather than as evidence for a unique type of difference.

2.1.6. GET- Ω

GET- Ω adds the energy-distance component T_E [10]. The same component-wise standardization in Equation (8) is used to obtain Z_E .

The GET- Ω statistic is then defined as:

$$T_\Omega = \max(Z_V, Z_L, Z_E) \quad (10)$$

GET- Ω is the maximum of the standardized spacing, median-location, and energy components. It is an omnibus statistic: significance indicates that at least one component departs from its permutation reference distribution, but the component statistics must be inspected to interpret the source of the signal.

2.1.7. Inference via Permutation Testing

Statistical significance is assessed by individual-label permutation [16,17]. Under the common i.i.d. null, pooled observations are exchangeable and are randomly reassigned to groups of the original sizes. This calibration is not generally valid for arbitrary dependent null processes; that limitation is evaluated separately in Section 2.2.8.

For GET-V, the transformed gaps and T_V are recomputed after every relabeling. For GET-Plus and GET- Ω , all components and their standardized maxima are recomputed using the same permutation reference distributions.

The permutation p -value can be computed as:

$$p = [1 + \sum_{\beta} 1\{T_b \geq T_{\text{obs}}\}]/(B + 1) \quad (11)$$

This finite-sample correction (i.e., $B + 1$ in the denominator) prevents zero p -values and maintains valid permutation calibration.

2.2. Simulation Framework

The simulation study has three purposes: evaluation under a classical i.i.d. location alternative, stress testing under constructed spacing alternatives, and calibration assessment under relevant null mechanisms. Figure 2 summarizes the workflow.

2.2.1. Scenario Design

The core evaluation includes one classical i.i.d. alternative and three constructed spacing stress tests:

M1 (location shift): $X \sim N(0,1)$ and $Y \sim N(0.5,1)$.

M2 (lattice stress test): X is i.i.d. $U(0,1)$;

Y is based on $(i + 0.5)/n$ with independent $N(0, (0.18/n)^2)$ jitter and clipping to $[0,1]$. M3 (micro-cluster stress test): both samples allocate equal counts to ten bins; X is uniform within bins, whereas Y is concentrated around two within-bin centers (0.03 and 0.07 above the left boundary) with SD 0.006 and is clipped to the corresponding bin boundaries. The two groups are generated independently in M2–M4, but observations within each constructed Y sample are not i.i.d.

M4 (shifted-lattice stress test): X is i.i.d. $U(0,1)$; Y is a randomly shifted grid modulo one with $N(0, 0.0008^2)$ jitter and clipping. The one-point distributions are approximately similar, but Y is dependent by construction.

Table 1 summarizes the exact generators and parameters for the M2–M4 spacing stress tests. Figure A1 shows the simulated value and spacing histograms for the constructed stress-test alternatives.

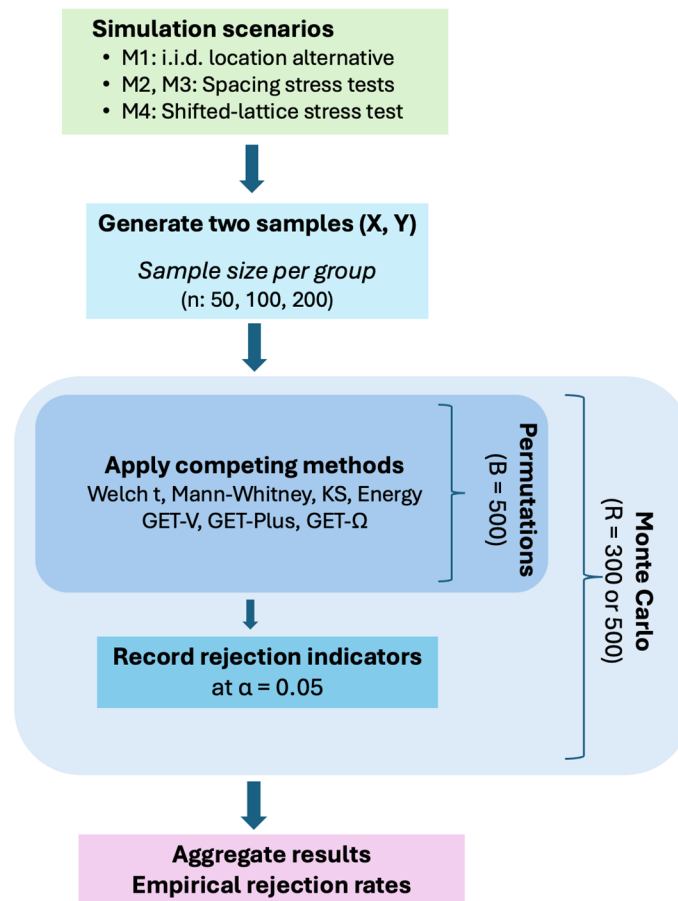


Figure 2. Simulation and validation workflow. The archived implementation is used for the core $R = 500$, $B = 500$ evaluation. Separate analyses assess population-standardized i.i.d. shape sensitivity ($R = 500$, $B = 500$) and matched-process null calibration ($R = 300$, $B = 500$). ‘Morphology’ and ‘Same-marginals’ denote the M2–M3 and M4 stress tests, respectively. For M2–M4, ‘power’ denotes descriptive rejection rates, not conventional power or evidence of equal marginal distributions.

Table 1. Core simulation scenarios. M2–M4 are constructed spacing stress tests outside the common i.i.d. null.

Scenario	Implemented Generator	Interpretation
M1	$X \sim N(0,1)$; $Y \sim N(0.5,1)$	i.i.d. location alternative
M2	$X \sim U(0,1)$; Y grid $(i + 0.5)/n + N(0, (0.18/n)^2)$ jitter	lattice stress test
M3	10 equal bins; X uniform; Y two centers/bin, SD 0.006; clipped within bins	micro-cluster stress test
M4	$X \sim U(0,1)$; Y random-shift grid $+ N(0, 0.0008^2)$ jitter	shifted-lattice stress test

2.2.2. Sampling Design

For each scenario, two groups of equal size are generated with $n = 50$, 100, or 200 observations per group. The groups are generated independently, although observations within the constructed non-i.i.d. mechanisms are dependent. Balanced group sizes $n_X = n_Y = n$ were used throughout to remove sample-size imbalance as a confounder.

2.2.3. Monte Carlo Procedure

For each core scenario and sample size, $R = 500$ Monte Carlo replications were performed with seed 42. Each permutation-based procedure used $B = 500$ label permutations (Table 2).

At each replication, two groups are generated according to the specified scenario, all competing methods are applied, and a binary rejection indicator is recorded for each method. The reported values are rejection proportions across replications; only M1 and the common i.i.d. null receive conventional power/type-I interpretations.

2.2.4. Competing Methods

The proposed GET statistics were compared with four standard two-sample procedures: Welch’s t -test, the Mann-Whitney test, the Kolmogorov-Smirnov test, and the energy-distance test. Welch’s t -test was used as a

location-sensitive parametric benchmark allowing unequal variances. The Mann-Whitney test was included as a rank-based nonparametric benchmark, and the Kolmogorov-Smirnov test as an empirical-distribution-function benchmark. The energy-distance test was included as a distance-based distributional benchmark.

For one-dimensional samples $X = (X_1, \dots, X_{n_X})$ and $Y = (Y_1, \dots, Y_{n_Y})$, the empirical energy-distance statistic $T_E(X, Y)$ was computed using the standard two-sample energy formulation [10]:

$$T_E(X, Y) = 2/(n_X n_Y) \sum_i \sum_j |X_i - Y_j| - 1/n_X^2 \sum_i \sum_j |X_i - X_j| - 1/n_Y^2 \sum_i \sum_j |Y_i - Y_j|$$

The same energy-distance statistic was used both as a standalone competing method and as the distribution-sensitive component T_E in GET- Ω . Classical tests were computed using standard implementations, whereas Energy and the GET-family statistics were calibrated by permutation testing under the stated exchangeability null.

Table 2. Technical specifications of the implemented analysis.

Component	Implemented Specification
Core Monte Carlo replications	R = 500
Core permutations	B = 500
Sample sizes	n = 50, 100, 200 per group
Significance level	alpha = 0.05
Core power seed	42
Core type-I seed	10042
GET-V transform	log1p (gap/median gap)
Gap floor	10^{-12} (numerical only; tied variables ineligible in applications)
Entropy window	m = floor (sqrt(number of transformed gaps))
GET-Plus location component	absolute difference in sample medians
Composite standardization	permutation median and normal-consistent MAD
Marginal-shape sensitivity analysis	R = 500, B = 500, seed = 20260831
Matched-null audit	R = 300, B = 500, seed = 20260901

2.2.5. Permutation Calibration

Permutation-based methods were calibrated by repeatedly relabeling the pooled observations while holding the original group sizes fixed [17]. Under the stated exchangeability null, this procedure provides finite-sample calibration.

2.2.6. Performance Metrics

For M1 and the common i.i.d. null, empirical rejection rates have their conventional power and type-I-error interpretations. For M2–M4, they are described as empirical rejection rates under constructed stress-test alternatives. Monte Carlo uncertainty is considered when interpreting the additional sensitivity and calibration analyses.

2.2.7. Sensitivity to i.i.d. Marginal Shape

Four population-standardized i.i.d. alternatives were performed: logistic with scale $\sqrt{3}/\pi$; $t(5)$ scaled by $\sqrt{3/5}$; a centered and variance-standardized lognormal distribution with log-scale SD 0.8; and an equal-weight symmetric mixture of $N(-1.5, 0.55^2)$ and $N(1.5, 0.55^2)$, divided by $\sqrt{1.5^2 + 0.55^2}$. Each was compared with $N(0,1)$ using R = 500 replications and B = 500 permutations.

2.2.8. Calibration Audit under Matched Process Nulls

To define the scope of individual-label permutation beyond the i.i.d. null, there are generated two samples from the same mechanism under four matched nulls: i.i.d. $U(0,1)$, the lattice generator used in M2, the micro-cluster generator used in M3, and the shifted-lattice generator used in M4. For each sample size, the audit used R = 300 replications, B = 500 permutations, and seed 20260901. This analysis identifies the dependence structures for which alternative resampling methods are required.

2.3. Application to Benchmark Datasets

It is presented a workflow that utilizes three publicly available benchmark datasets: an explicitly synthetic Alzheimer dataset, a dementia dataset with limited provenance, and a public stroke prediction dataset [19–21]. These examples illustrate the approach to selecting method selection, tie screening, and component-wise interpretation instead of clinical inference. The datasets contain binary group labels and a mix of continuous, rounded, ordinal, and binary predictors. Variables were not considered clinical cohorts unless stated in the source.

Numeric predictors were separately screened for each dataset. Identifier-like variables were excluded, any observations that were missing were removed for each comparison, and all observations in both groups were to be at least ten. GET-V eligibility required all observed values within each group to be unique. Any variable with a repeated value was considered as tied and was not selected to be GET-V, GET-Plus, or spacing-based in interpretation. Where appropriate, traditional tests and energy distance were still reported.

The main text set contains four tie-free comparisons and two high-tie stroke variables that illustrate the eligibility rule. Table A2 shows the complete set of tie-free comparisons.

2.3.1. Data Preprocessing and Selection

Real data analyses were performed variable by variable. Identifier-like fields were discarded, missing predictor or outcome values were separately removed for each comparison, and each binary outcome group was required to contain at least ten observations. Before any spacing-based analysis, uniqueness was checked in each group. Variables with one or more repeated observations were not considered valid for GET-V and its composites, and repeated subsampling was not used as a substitute for tie handling.

2.3.2. Test Comparison and Interpretation Framework

For tie-free variables, groups were compared using Welch's t -test, Mann-Whitney, Kolmogorov-Smirnov, energy distance, GET-V, GET-Plus, and GET- Ω . For tied variables, we only consider conventional and energy distance results; the GET-family results are not considered. The p -values are exploratory and not adjusted for multiple comparisons. We do not make any multiplicity controlled claim in this paper. GET-V assesses differences in empirical spacing entropy and is interpreted together with the companion marginal distribution analyses. GET-Plus and GET- Ω add location and energy components, so we can look at the source of each result directly.

The subsampling analysis was removed from the main inferential claims because repeated subsampling does not resolve ties or validate individual-label permutation under dependent nulls.

The sensitivity to nonlinear transformation is reported as a change of scale analysis, not an invariance property. Positive rescaling leaves GET-V unchanged while nonlinear transformations can change the result.

2.4. Implementation

For reproducibility, the software stack is pinned to the latest stable releases available on 31 March 2026: Python 3.14.3, NumPy 2.4.4, pandas 3.0.2, SciPy 1.17.1, and Matplotlib 3.10.8. These version pins specify the March 2026 reproduction environment and do not imply that the original analyses were first executed with those later releases. The core power analysis used seed 42, whereas the archived type-I simulation initialized its generator with seed 10042. The core output files correspond to $R = 500$ replications and $B = 500$ permutations. The marginal-shape sensitivity analysis used the archived GET-V implementation with seed 20260831, $R = 500$, and $B = 500$; the matched-process calibration audit used seed 20260901, $R = 300$, and $B = 500$.

3. Results

3.1. Simulation Results

The simulation results are reported in Tables 3–7 and Figures 3–5. Tables 3–5 summarize the core scenarios, while Tables 6 and 7 present the additional analyses of marginal-shape sensitivity and matched-process calibration.

Table 3. Empirical rejection rates across the four core simulation scenarios and three sample sizes. For M1, the rates have the conventional power interpretation; M2–M4 are constructed spacing stress tests.

Scenario	n	Welch t	MW	KS	Energy	GET-V	GET-Plus	GET- Ω
M1_shift	50	0.712	0.690	0.514	0.674	0.044	0.436	0.672
M1_shift	100	0.918	0.904	0.782	0.896	0.066	0.724	0.890
M1_shift	200	1.000	1.000	0.988	1.000	0.064	0.974	1.000
M2_lattice	50	0.002	0.002	0.002	0.002	1.000	1.000	1.000
M2_lattice	100	0.012	0.010	0.002	0.008	1.000	1.000	1.000
M2_lattice	200	0.006	0.006	0.000	0.004	1.000	1.000	1.000
M3_micro	50	0.000	0.000	0.000	0.000	0.460	0.280	0.074
M3_micro	100	0.000	0.000	0.000	0.000	0.882	0.770	0.434
M3_micro	200	0.000	0.000	0.000	0.000	0.980	0.950	0.802
M4_same_marg	50	0.008	0.008	0.004	0.004	1.000	1.000	1.000
M4_same_marg	100	0.004	0.004	0.000	0.004	1.000	1.000	1.000
M4 same marg	200	0.004	0.004	0.000	0.002	1.000	1.000	1.000

3.1.1. Empirical Rejection Rates across Scenarios and Sample Sizes

Empirical rejection rates for all competing methods are presented in Table 3 for the four core simulation scenarios and three sample sizes. Several distinct patterns emerge.

Under M1, rejection rates increase with sample size for Welch's t -test, Mann–Whitney, Kolmogorov–Smirnov, energy distance, GET-Plus, and GET- Ω . GET-V remains near the nominal level (0.044–0.066), consistent with invariance to a pure location shift.

Under the M2 lattice stress test, GET-family rejection rates are 1.000, whereas the conventional and energy procedures have very low rejection rates. This strong contrast demonstrates the ability of the GET statistics to detect the constructed regular gap pattern.

Under M3, GET-V rejection increases from 0.460 at $n = 50$ to 0.980 at $n = 200$. GET-Plus and GET- Ω also increase with sample size. These results describe behavior under the specified micro-cluster generator.

Under M4, GET-family rejection rates are 1.000 at all sample sizes, while the marginal-distribution procedures have low rejection rates. M4 therefore demonstrates strong sensitivity to shifted-lattice gap regularity.

Figure 3 displays the distinct response profiles of the procedures across the four core scenarios.

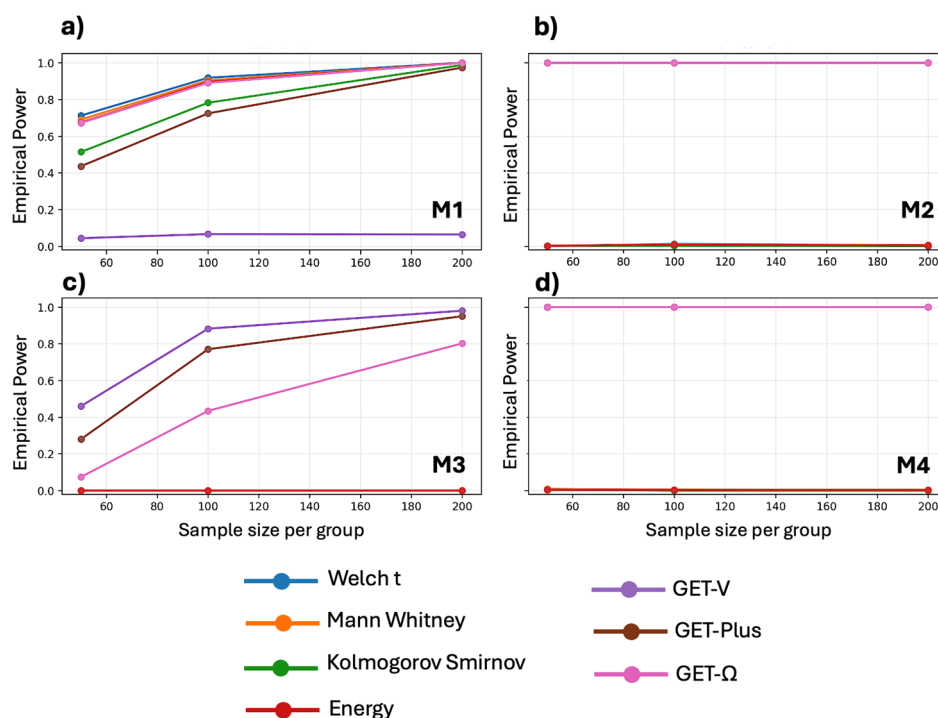


Figure 3. Empirical rejection rates across the four core scenarios. (a) M1: i.i.d. location-shift alternative. (b) M2: perturbed-lattice spacing stress test. (c) M3: micro-cluster spacing stress test. (d) M4: shifted-lattice spacing stress test. M1 has the conventional power interpretation; for M2–M4, the axis label ‘Empirical Power’ denotes descriptive rejection rates.

3.1.2. Average Rejection Rates by Scenario Class

Table 4 summarizes average rejection rates by scenario class. GET-V is low for M1 and high for the constructed spacing scenarios, whereas GET- Ω provides broader sensitivity across the evaluated generators. These averages are descriptive and should not be generalized beyond the specified scenarios.

Figure 4 provides a compact descriptive summary of the same scenario-specific rejection rates. Because the original figure is preserved, its legacy class labels are defined explicitly in the caption.

Table 4. Average empirical rejection rates by scenario class. M1 is the i.i.d. location alternative; M2–M3 are spacing stress tests; and M4 is the shifted-lattice stress test.

Class	Welch t	MW	KS	Energy	GET-V	GET-Plus	GET- Ω
M1 location shift	0.877	0.865	0.761	0.857	0.058	0.711	0.854
M2–M3 spacing stress tests	0.003	0.003	0.001	0.002	0.887	0.833	0.718
M4 shifted-lattice stress test	0.005	0.005	0.001	0.003	1.000	1.000	1.000
Overall	0.222	0.219	0.191	0.216	0.708	0.845	0.823

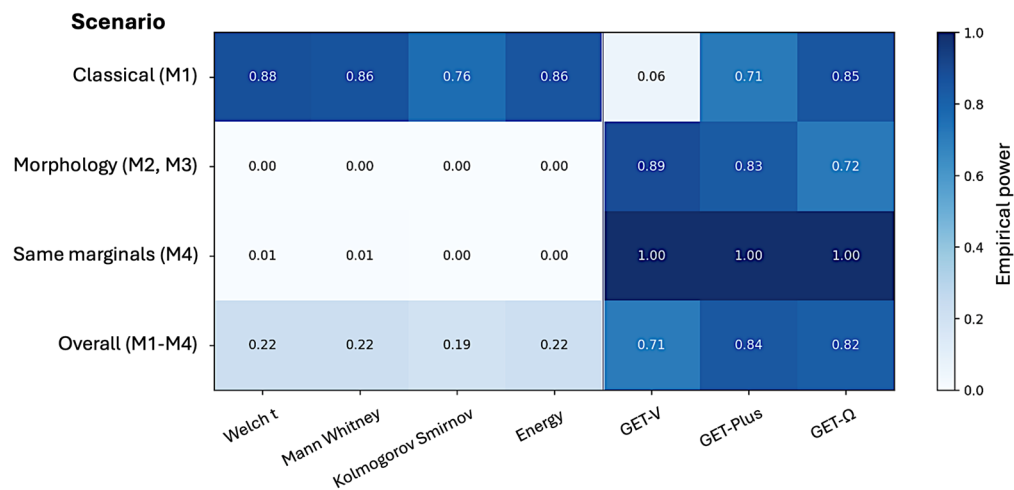


Figure 4. Average empirical rejection rates by constructed scenario class. The preserved internal labels 'Morphology' and 'Same-marginals' are a shorthand for the M2-M3 spacing stress tests and the M4 shifted-lattice stress test, respectively. These labels imply neither equality of marginal distributions nor independence from marginal shape. 'Empirical power' denotes descriptive rejection rates for the stress tests.

3.1.3. Type-I Error Control

Under the common i.i.d. $N(0,1)$ null, all procedures are close to the nominal 0.05 level. GET-V averages 0.053 across $n = 50, 100$, and 200.

Table 5. Empirical type-I error under the common i.i.d. $N(0,1)$ null ($R = 500$; $B = 500$ for permutation procedures).

n	Welch t	MW	KS	Energy	GET-V	GET-Plus	GET-Ω
50	0.044	0.044	0.044	0.040	0.058	0.058	0.034
100	0.044	0.044	0.022	0.042	0.050	0.042	0.046
200	0.044	0.042	0.026	0.036	0.052	0.030	0.036
Overall	0.044	0.043	0.031	0.039	0.053	0.043	0.039

3.1.4. Blind-Spot Illustration

Figure 5 illustrates the pronounced spacing contrast between an i.i.d. uniform sample and a shifted near-regular sample, highlighting the gap feature targeted by GET-V.

3.1.5. Sensitivity to i.i.d. Marginal Shape

Population-standardized shape alternatives show that GET-V is not specific to a marginal-independent structural feature. Logistic rejection increases from 0.066 to 0.126; $t(5)$ from 0.072 to 0.172; and skewed lognormal from 0.198 to 0.542 as n increases from 50 to 200. The bimodal mixture remains near nominal (0.036–0.058). Table 6 reports Monte Carlo standard errors.

Table 6. GET-V rejection rates under population-standardized i.i.d. shape alternatives ($R = 500$, $B = 500$). MCSE denotes Monte Carlo standard error.

Alternative vs. $N(0,1)$	n	GET-V Rejection	MCSE
Logistic	50	0.066	0.011
Logistic	100	0.068	0.011
Logistic	200	0.126	0.015
$t(5)$	50	0.072	0.012
$t(5)$	100	0.130	0.015
$t(5)$	200	0.172	0.017
Skewed lognormal	50	0.198	0.018
Skewed lognormal	100	0.364	0.022
Skewed lognormal	200	0.542	0.022
Bimodal mixture	50	0.036	0.008
Bimodal mixture	100	0.036	0.008
Bimodal mixture	200	0.058	0.010

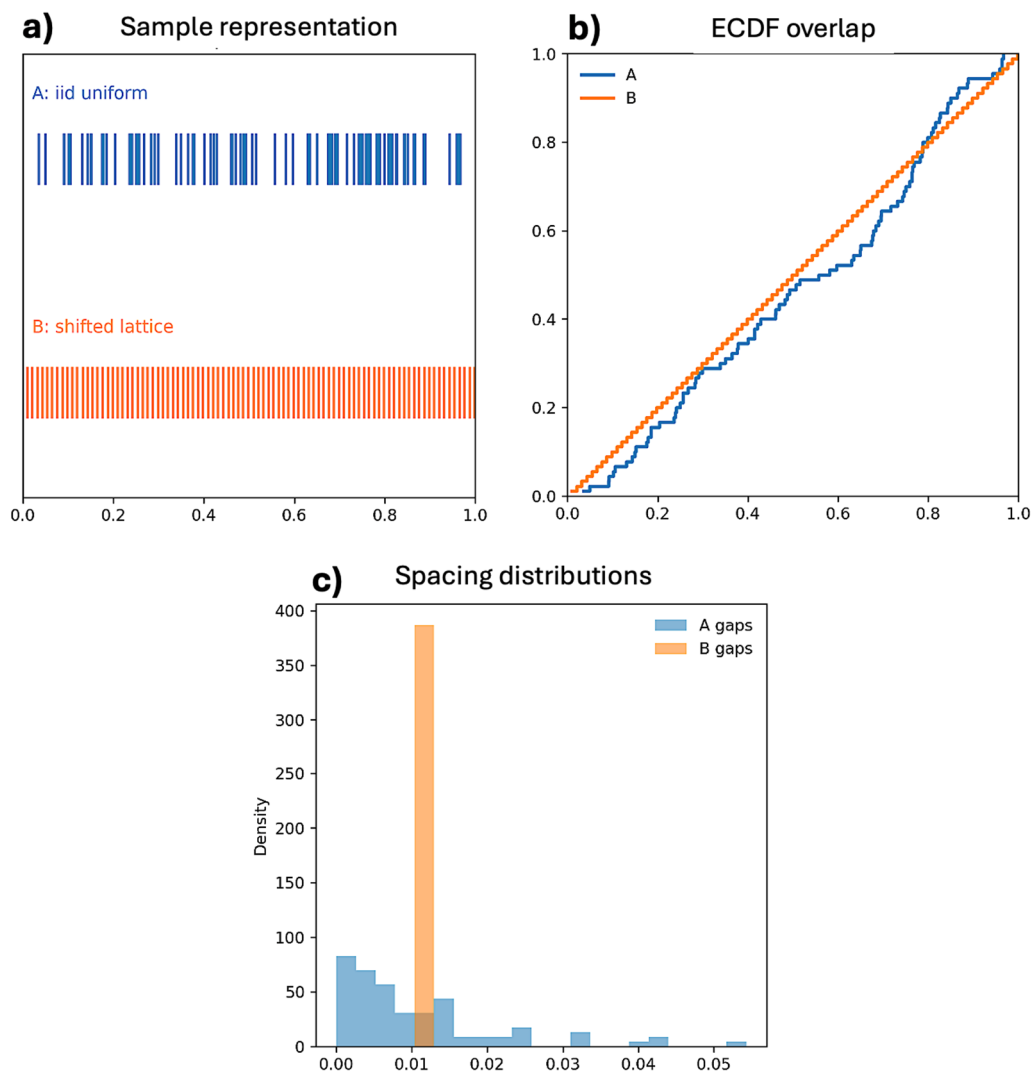


Figure 5. Illustrative spacing contrast between an i.i.d. uniform sample and a shifted near-regular sample. Any preserved ‘Same-marginals’ label is legacy shorthand for the M4 stress-test generator and not a claim of exact equality of marginal distributions.

3.1.6. Calibration under Matched Process Nulls

The matched-process audit delineates the calibration boundary of individual-label permutation. Type-I error remains near nominal for matched i.i.d. uniform and micro-cluster generators, whereas the lattice and shifted-lattice generators require a process-aware resampling strategy. Table 7 quantifies these differences.

Table 7. Diagnostic GET-V type-I-error audit under matched null generators ($R = 300$, $B = 500$).

Matched Null Generator	n	GET-V Type-I Rate	MCSE
iid uniform	50	0.050	0.013
iid uniform	100	0.043	0.012
iid uniform	200	0.043	0.012
lattice	50	0.383	0.028
lattice	100	0.323	0.027
lattice	200	0.327	0.027
micro-cluster	50	0.047	0.012
micro-cluster	100	0.053	0.013
micro-cluster	200	0.047	0.012
shifted lattice	50	0.370	0.028
shifted lattice	100	0.323	0.027
shifted lattice	200	0.343	0.027

3.2. Benchmark-Dataset Results

Illustrative benchmark results are presented in Tables 8 and 9. GET-V is evaluated only for variables with no repeated observations within either group. The reported p -values are exploratory and unadjusted for multiple comparisons.

3.2.1. Group Summaries

Table 8 presents selected group summaries and explicitly identifies variables that are ineligible for GET-V because of ties.

Table 8. Selected benchmark group summaries and GET-V tie eligibility. Group 0 and group 1 follow the source binary coding for Diagnosis (Alzheimer), Dementia (dementia), and stroke (stroke).

Dataset	Variable	n (Group 0)	Mean (Group 0)	n (Group 1)	Mean (Group 1)	GET-V Eligible
Alzheimer	MMSE	1389	16.27	760	11.99	Yes
Alzheimer	FunctionalAssessment	1389	5.86	760	3.65	Yes
dementia	Weight	515	75.02	485	73.58	Yes
dementia	MRI_Delay	515	29.59	485	30.65	Yes
stroke	age	4700	41.76	209	67.71	No
stroke	bmi	4700	28.82	209	30.47	No

Selection note: the main-text set illustrates four tie-free comparisons and two high-tie exclusions; Table A2 reports all tie-free comparisons.

Among tie-free variables, Alzheimer MMSE and FunctionalAssessment show strong marginal distributional differences but non-significant GET-V results. Dementia Weight and MRI_Delay show no convincing GET-V evidence. Stroke age and BMI contain extensive ties and are therefore retained only for conventional distributional comparison; GET-V is not evaluated.

3.2.2. Full-Sample Test Comparisons

Table 9 shows the test results. The tie-free variables show how GET-V can deliver information different from a standard and omnibus comparison, while tied variables illustrate the value of eligibility screening.

Table 9. Selected exploratory benchmark comparisons. GET-family results are not evaluated (NE) for variables with ties. p -values are unadjusted for multiplicity.

Dataset	Variable	Welch t	MW	KS	Energy	GET-V	GET-Plus	GET- Ω
Alzheimer	MMSE	<0.00001	<0.00001	<0.00001	0.00200	0.07984	0.00200	0.00200
Alzheimer	FunctionalAssessment	<0.00001	<0.00001	<0.00001	0.00200	0.09581	0.00200	0.00200
dementia	Weight	0.11814	0.12052	0.02135	0.04391	0.64870	0.65269	0.04391
dementia	MRI_Delay	0.32352	0.32550	0.42330	0.40319	0.18762	0.32335	0.42315
stroke	age	<0.00001	<0.00001	<0.00001	0.00200	NE	NE	NE
stroke	bmi	0.00033	0.00010	0.00002	0.00200	NE	NE	NE

Multiplicity note: these p -values are exploratory and unadjusted; no confirmatory family-wise inference is claimed.

Alzheimer MMSE and Functional Assessment show strong conventional, energy, and omnibus differences, and GET-V is non-significant. Since the Alzheimer dataset is synthetic, these results are more illustrative and not clinical evidence. Dementia Weight and MRI_Delay are tie-free and do not get significant GET-V results. Stroke age and BMI are conventional distributions but have very large repeated values and are not eligible for GET-V.

3.2.3. Tie Eligibility and Interpretation

The eligibility audit in Table A1 shows why rounded, ordinal, and binary variables require separate treatment. For excluded variables, the numerical gap floor in the archived implementation would replace many zero gaps with the same arbitrary constant. Such output would describe the flooring convention as much as the data. Accordingly, these variables are not assigned GET-family p -values.

3.3. Overall Empirical Synthesis

GET-V is well-calibrated under the common i.i.d. null and is not sensitive to a pure location shift. It is very responsive to the gap regularity alternatives but is also sensitive to the i.i.d. marginal shape. The matched-process audit showed that the individual label permutation is not appropriate for some dependent null processes and that process-aware resampling is required. GET therefore has a specific complementary role in two-sample analysis rather than a general interpretation as a point process test.

4. Discussion

This study introduces a unified spacing-sensitive extension to two-sample analysis. GET makes gap information explicit, combines it with location and energy components, and provides a reproducible permutation framework whose operating conditions are directly evaluated by the marginal-shape sensitivity and matched-process calibration analyses.

4.1. Main Findings

GET-V remains near nominal under an i.i.d. location shift and responds strongly to the constructed lattice and clustering alternatives. However, its response to skewness and heavy tails demonstrates that a significant result is not specific to dependent organization.

GET-Plus and GET- Ω extend the spacing statistic into transparent composite procedures. Their component-wise construction preserves interpretability while broadening sensitivity to location and overall distributional differences.

4.2. Structural Versus Distributional Differences

The phrase “structural difference” is used only for a scientifically specified dependent mechanism. In general i.i.d. applications, GET-V measures a difference in empirical spacing entropy, which can arise from marginal shape.

The calibration audit under matched-process nulls shows that the individual-label permutation procedure is appropriate for the common i.i.d. null but can be severely anti-conservative for matched lattice processes. Type-I error ranges from 0.323 to 0.383 for matched lattice nulls and from 0.323 to 0.370 for matched shifted-lattice nulls, while remaining near nominal for the matched micro-cluster generator. These results preclude a general point-process-testing interpretation.

GET-Plus includes both spacing and median-location components, and GET- Ω also includes energy distance. While they are useful in a single-source summary, neither composite can provide a unique causal or structural mechanism without inspection of its components.

Reporting the spacing, location, and broader distributional components separately helps identify which component drives the composite result.

4.3. Insights from Benchmark-Dataset Applications

This analysis shows how eligibility screening guides application of the method. The Alzheimer data are synthetic, the dementia source has limited provenance information, and several clinically familiar variables are rounded or discrete; these features make the datasets useful methodological examples for component interpretation and tie handling.

4.4. Methodological Implications

The main methodological contribution is a reproducible spacing-entropy statistic with two interpretable composite extensions. It adds gap information to conventional two-sample analysis. GET permutation inference requires continuous, tie-free measurements under the common i.i.d. exchangeability null.

GET-V is invariant to translation and positive rescaling only. It is not invariant to general monotone transformations or marginal shape. The shape analysis quantifies this limitation.

The Vasicek-type estimator provides a convenient local summary of transformed gaps. Its window rule and boundary implementation are now reported, but other estimators and tuning strategies need to be tested further. The current estimator was chosen for simplicity and computational feasibility in permutation loops. This practical choice should not be viewed as evidence that it is optimal for all distributions, sample sizes, or dependence structures.

Practical selection of a GET statistic. Use GET-V for a spacing-sensitive question with continuous, tie-free measurements under the common i.i.d. exchangeability null. Use GET-Plus where both spacing and median location are important and GET- Ω for an omnibus comparison incorporating energy distance. Report the component results together with usual distributional summaries.

4.5. Limitations and Future Directions

Several limitations qualify interpretation. GET-V is sensitive to the shape of the population and it cannot find a marginal-independent population feature in the standard i.i.d. model. Individual-label permutation is not generally valid under dependent point-process nulls. Tied data require a validated extension rather than numerical

flooring. The benchmark datasets are not clinical validation and one is explicitly synthetic. Finally, the Vasicek window rule, the finite-sample behavior and k-sample extensions need further study.

In future work, this framework could be extended to resampling for replicated point patterns or exchangeability blocks, to principled treatment of ties and discrete measurements, and to compare other entropy estimators.

4.6. Implications

These analyses place GET within a defined scope as a complementary spacing-sensitive method for continuous, tie-free data. Valid extensions to tied data and dependent processes will require dedicated methods.

By making gap-based information explicit, GET broadens two-sample analysis beyond conventional location and distributional summaries. Its transparent component structure and direct calibration audits support informed use across a range of continuous-data applications.

5. Conclusions

GET adds a reproducible spacing-entropy component to conventional two-sample analysis. Under a common i.i.d. exchangeability null, GET-V is approximately calibrated and insensitive to a pure location shift, while responding strongly to the constructed regularity and clustering alternatives. GET-Plus and GET- Ω combine spacing information with location and energy components. The shape-sensitivity and matched-process analyses also show where interpretation must be restricted. Within these conditions, GET offers a novel and interpretable way to examine gap information alongside conventional distributional comparisons. Extensions to tied data and dependent processes require dedicated methods.

Author Contributions

V.D.K.: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, supervision, project administration. The author has read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The datasets that were used for the implementation of this analysis were accessed and obtained from the original publications [19–21]. Derived data supporting the findings of this study are available from the corresponding author [V.D.K.] on request.

Conflicts of Interest

The author declares no conflict of interest.

Use of AI and AI-Assisted Technologies

During the preparation of this work, the author used Grammarly to assist with language editing and clarity improvements. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article.

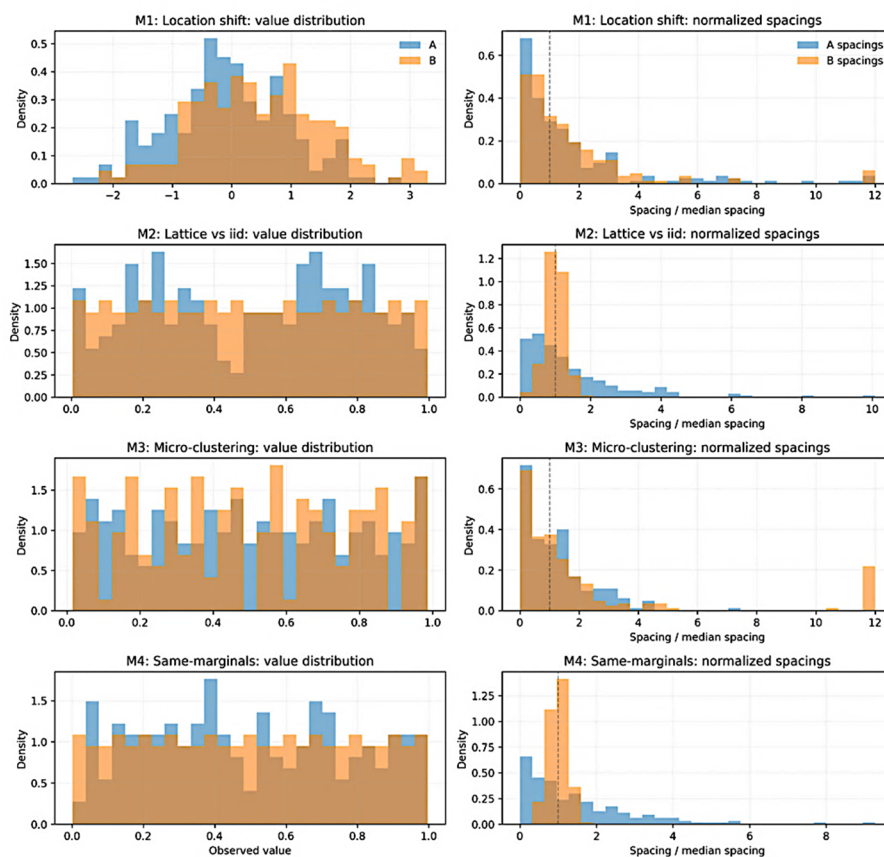
Appendix A

Table A1. Eligibility audit for selected benchmark variables. GET-V was evaluated only when both groups contained no repeated values.

Dataset	Variable	Within-Group Repeated Observations	GET-V Status
Alzheimer	MMSE	0%	Eligible
Alzheimer	FunctionalAssessment	0%	Eligible
dementia	Weight	0%	Eligible
dementia	MRI_Delay	0%	Eligible
dementia	Cognitive_Test_Scores	98–99%	Excluded
stroke	age	82–98%	Excluded
stroke	bmi	35–91%	Excluded

Table A2. Full tie-free benchmark comparisons for GET-V and companion methods.

Dataset	Variable	Welch t	KS	Energy	GET-V
Alzheimer	BMI	0.22465	0.11567	0.19960	0.97006
Alzheimer	AlcoholConsumption	0.72435	0.71483	0.72854	0.10579
Alzheimer	PhysicalActivity	0.78226	0.85059	0.95609	0.27745
Alzheimer	DietQuality	0.69360	0.85331	0.80240	0.46507
Alzheimer	SleepQuality	0.00867	0.03946	0.00998	0.66467
Alzheimer	CholesterolTotal	0.76858	0.59190	0.74251	0.28543
Alzheimer	CholesterolLDL	0.13793	0.47938	0.26547	0.89222
Alzheimer	CholesterolHDL	0.04895	0.03749	0.04790	0.53693
Alzheimer	CholesterolTriglycerides	0.29380	0.32604	0.40120	0.16567
Alzheimer	MMSE	<0.00001	<0.00001	0.00200	0.07984
Alzheimer	FunctionalAssessment	<0.00001	<0.00001	0.00200	0.09581
dementia	Weight	0.11814	0.02135	0.04391	0.64870
dementia	MRI_Delay	0.32352	0.42330	0.40319	0.18762

**Figure A1.** Simulated value and spacing histograms for the constructed stress-test alternatives. The preserved panel label ‘Same-marginals’ is legacy shorthand for the M4 shifted-lattice generator and does not denote exact equality of the marginal distributions. The panels highlight the empirical gap patterns targeted by GET-V.

References

1. Student. The Probable Error of a Mean. *Biometrika* **1908**, *6*, 1–25.
2. Box, G.E.P. Non-Normality and Tests on Variances. *Biometrika* **1953**, *40*, 318–335. <https://doi.org/10.1093/biomet/40.3-4.318>.
3. Welch, B.L. The Generalization of “Student’s” Problem when Several Different Population Variances Are Involved. *Biometrika* **1947**, *34*, 28–35. <https://doi.org/10.1093/biomet/34.1-2.28>.
4. Lehmann, E.L. *Nonparametrics: Statistical Methods Based on Ranks*; Holden-Day: San Francisco, CA, USA, 1975.
5. Mann, H.B.; Whitney, D.R. On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other. *Ann. Math. Stat.* **1947**, *18*, 50–60. <https://doi.org/10.1214/aoms/1177730491>.
6. Hájek, J.; Šidák, Z. *Theory of Rank Tests*; Academic Press: New York, NY, USA, 1967.
7. Kolmogorov, A.N. Sulla Determinazione Empirica Di Una Legge Di Distribuzione. *G. Ist. Ital. Attuari* **1933**, *4*, 83–91.
8. Smirnov, N. Table for Estimating the Goodness of Fit of Empirical Distributions. *Ann. Math. Stat.* **1948**, *19*, 279–281. <https://doi.org/10.1214/aoms/1177730256>.
9. Cramér, H. *Mathematical Methods of Statistics*; Princeton University Press: Princeton, NJ, USA, 1946.
10. Székely, G.J.; Rizzo, M.L. Energy Statistics: A Class of Statistics Based on Distances. *J. Stat. Plan. Inference* **2013**, *143*, 1249–1272. <https://doi.org/10.1016/j.jspi.2013.03.018>.
11. Székely, G.J.; Rizzo, M.L.; Bakirov, N.K. Measuring and Testing Dependence by Correlation of Distances. *Ann. Statist.* **2007**, *35*, 2769–2794. <https://doi.org/10.1214/009053607000000505>.
12. Gretton, A.; Borgwardt, K.; Rasch, M.J.; et al. A Kernel Two-Sample Test. *J. Mach. Learn. Res.* **2012**, *13*, 723–773.
13. Berlinet, A.; Thomas-Agnan, C. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*; Springer US: New York, NY, USA, 2004. <https://doi.org/10.1007/978-1-4419-9096-9>.
14. Pyke, R. Spacings. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1965**, *27*, 395–436. <https://doi.org/10.1111/j.2517-6161.1965.tb00602.x>.
15. Vasicek, O. A Test for Normality Based on Sample Entropy. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1976**, *38*, 54–59. <https://doi.org/10.1111/j.2517-6161.1976.tb01566.x>.
16. Good, P. *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*, 2nd ed.; Springer: New York, NY, USA, 2000. <https://doi.org/10.1007/978-1-4757-3235-1>.
17. Edgington, E.S.; Onghena, P. *Randomization Tests*, 4th ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2007. <https://doi.org/10.1201/9781420011814>.
18. Ebrahimi, N.; Pflughoeft, K.; Soofi, E.S. Two Measures of Sample Entropy. *Stat. Probab. Lett.* **1994**, *20*, 225–234. [https://doi.org/10.1016/0167-7152\(94\)90046-9](https://doi.org/10.1016/0167-7152(94)90046-9).
19. Fedesoriano. Stroke Prediction Dataset. Available online: <https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset> (accessed on 1 May 2026).
20. El Kharoua, R. Alzheimer’s Disease Dataset. Available online: <https://www.kaggle.com/datasets/rabieelkharoua/alzheimers-disease-dataset> (accessed on 1 May 2026).
21. Adeyemi, T. Dementia Patient Health Dataset. Available online: <https://www.kaggle.com/datasets/timothyadeyemi/dementia-patient-health-dataset> (accessed on 1 May 2026).