AKTI SURVEY

GTPS-CV Methodology

Geo-Trust Progressive Sampling with Cross-Validation

Version: 1.0

Status: Production-Ready Framework

Updated: January 2026

Executive Summary

GTPS-CV is a field-driven survey methodology designed to produce representative, manipulation-resistant data in environments where digital surveys are vulnerable to farming, outsourcing, or coordinated bias. Version 1.0 introduces structured sampling protocols, expanded demographic capture, situational location classification, and statistical weighting frameworks to achieve population representativeness.

Core Principles

1. Trust earned through performance — not inherited or assumed

2. Geographic diversity as a quality signal — preventing regional capture

3. Cross-validation among coordinators — detecting anomalies through comparison

4. Stratified sampling with post-stratification weighting — achieving representativeness

5. Diversity over volume — rewarding coverage breadth, not response counts

Part 1: Coordinator Structure and Trust System

1.1 Trust Hierarchy

The platform operates through a hierarchical coordinator network:

Tier	Trust Range	Appointment	Capacity
Tier 1 (Anchor)	90–100	Site-appointed	10–20 coordinators
Tier 2	70–89	Recruited by Tier 1	Up to 10 per parent
Tier 3	50–69	Recruited by Tier 2	Up to 10 per parent
Tier 4 (Field)	30–49	Recruited by Tier 3	Up to 10 per parent

1.2 Trust Initialization

New coordinators are initialized using the formula:

Initial Trust = min(Parent Trust × 0.85, Tier Maximum) - Admin Adjustment

Where Parent Trust × 0.85 creates 15% minimum decay, Tier Maximum caps trust at tier ceiling, and Admin Adjustment allows manual reduction (0–20 points).

1.3 Trust Dynamics

Trust scores update after each survey period based on four factors:

Factor	Weight	Measurement
Cross-validation consistency	35%	Deviation from area peers
Geographic cluster diversity	30%	Unique location clusters covered
Demographic cluster diversity	20%	Spread across demographic cells
Protocol compliance	15%	Geo-fence adherence, timing rules

Trust is bounded: minimum 10, maximum 100. Coordinators falling below 20 are flagged for review.

Part 2: Sampling Framework

2.1 Location Selection Protocol

Representativeness begins with systematic location selection, not convenience. Each survey defines Primary Sampling Units (PSUs) based on administrative boundaries, population density zones, and known demographic distributions.

Location Assignment Methods

Randomized Grid Assignment: Geographic area divided into grid cells; coordinators assigned random cells.

Quota-Based Distribution: Cells weighted by population; more coordinators assigned to denser areas.

Rotation Schedule: Assignments rotate weekly to prevent familiarity bias.

2.2 Respondent Selection Protocol

Within assigned locations, coordinators follow systematic time-interval sampling:

- Begin collection at assigned start time
- Approach first eligible adult after arrival
- After each completion, wait 3 minutes before next approach
- If refused, wait 1 minute, approach next eligible person
- No targeting based on appearance, dress, or perceived demographics
- Record all refusals for response rate calculation

2.3 Situational Location Classification

Every response captures situational context to enable stratification and bias detection:

Code	Location Type	Expected Demographics
RES-APT	Apartment complex	Mixed urban
RES-HSE	Residential house area	Suburban/family
EDU-SCH	School vicinity	Parents, staff
EDU-COL	College/University	Young adults 18–25
TRN-STA	Train/Metro station	Commuters, mixed
TRN-RDE	Train/Metro ride	Commuters, mixed
BUS-STP	Bus stop	Mixed, lower-middle income
BUS-RDE	Bus ride	Mixed, lower-middle income
COM-MLL	Shopping mall	Consumers, mixed
COM-MKT	Street market/bazaar	Local community
COM-CAF	Café/Restaurant	Urban, varied income
REL-MOS	Mosque vicinity	Muslim community
REL-CHR	Church vicinity	Christian community
REL-TMP	Temple vicinity	Hindu/Buddhist community
REC-PRK	Public park	Families, recreation
WRK-OFF	Office district	White-collar workers
WRK-IND	Industrial area	Blue-collar workers
GOV-OFF	Government office	Citizens, bureaucrats
HLT-HSP	Hospital vicinity	Patients, caregivers

Urban/Rural Classification

Code	Definition	Criteria
URB-1	Metro urban	City population > 1 million
URB-2	Urban	City population 100K–1M
URB-3	Semi-urban	Town population 20K–100K
RUR-1	Rural town	Population 5K–20K
RUR-2	Rural village	Population < 5K

Part 3: Data Collection Specification

3.1 Geographic Data (Geo-Cluster)

Every response captures a three-part geo-cluster:

Component	Field	Source	Required
Physical	GPS Coordinates	Browser GPS	Yes
Situational	Location Type	Coordinator selection	Yes
Settlement	Urban/Rural Code	Derived + Coordinator	Yes

GPS accuracy threshold: ≤ 50 meters. Responses without GPS are flagged as "unverified" and weighted at 50% in analysis. More than 20% unverified from a coordinator triggers review.

3.2 Demographic Data (Demo-Cluster)

Four demographic attributes form the demo-cluster:

Gender

Code	Label
M	Male
F	Female
X	Other/Prefer not to say

Age Bracket

Code	Range
A1	18–24
A2	25–34
A3	35–44
A4	45–54
A5	55–64
A6	65+

Occupation Category

Code	Category	Examples
OCC-STU	Student	School, college, university
OCC-UNE	Unemployed/Seeking	Job seekers
OCC-HOM	Homemaker	Primary household managers
OCC-RET	Retired	Pensioners
OCC-AGR	Agriculture/Fishing	Farmers, fishers, laborers
OCC-MAN	Manual/Trade	Construction, factory, drivers
OCC-SVC	Service sector	Retail, hospitality, security
OCC-CLR	Clerical/Office	Admin, data entry, reception
OCC-PRO	Professional	Engineers, doctors, lawyers, teachers
OCC-MGT	Management/Executive	Managers, directors, owners
OCC-GOV	Government/Public	Civil servants, military, police
OCC-OTH	Other	Specify in notes

Income Level (Self-Reported)

Code	Description	Anchor Question
INC-1	Struggling	Difficulty meeting basic needs
INC-2	Getting by	Cover basics with little extra
INC-3	Comfortable	Meet needs with some savings
INC-4	Well-off	Comfortable with regular savings
INC-5	Affluent	No financial concerns
INC-X	Prefer not to say	—

Part 4: Diversity Scoring and Rewards

4.1 Core Principle: Diversity Over Volume

Coordinators are never rewarded for volume. The incentive system rewards coverage diversity across geographic and demographic clusters. A coordinator with 20 responses across 15 unique clusters scores higher than one with 100 responses from 3 clusters.

4.2 Diversity Score Calculation

Geographic Diversity Index (GDI)

Measures spread across unique geo-clusters:

GDI = (Unique Geo-Clusters Covered / Total Possible Geo-Clusters in Area) × 100

Where Unique Geo-Cluster = unique combination of GPS grid cell (500m × 500m) + Situational location type + Urban/Rural code.

Demographic Diversity Index (DDI)

Measures spread across demographic cells:

DDI = (Unique Demo-Clusters Covered / Target Demo-Clusters) × 100

Maximum theoretical cells = 3 × 6 × 12 × 6 = 1,296. Practical target cells ≈ 100–200.

Combined Diversity Score (CDS)

CDS = (GDI × 0.5) + (DDI × 0.5)

Both indices weighted equally to prevent gaming either dimension.

4.3 Reward Point System

Points are earned based on diversity contribution, not response count:

Action	Points	Condition
New geo-cluster coverage	10	First response in that geo-cluster
New demo-cluster coverage	10	First response in that demo-cluster
Repeat geo-cluster	1	Additional response in covered cluster
Repeat demo-cluster	1	Additional response in covered cluster
Cross-validation bonus	5	Response aligns with nearby coordinator
Hard-to-reach bonus	15	Underserved area/demographic

Point Decay for Repetition

To actively discourage volume farming:

Points per repeat = 1 / (1 + repeat_count_in_cluster)

Response 1: 10 points (new). Response 2: 0.5 points. Response 3: 0.33 points. Response 10: 0.1 points.

4.4 Leaderboard Rankings

Leaderboards display coordinators ranked by:

1. Primary Rank: Combined Diversity Score (CDS)

2. Secondary Rank: Cross-validation consistency rate

3. Tertiary Rank: Protocol compliance rate

Volume (total responses) is displayed but never used for ranking.

4.5 Reward Tiers

Tier	CDS Threshold	Reward Type
Platinum	CDS ≥ 80	Monetary + Certificate + Badge
Gold	CDS 60–79	Monetary + Certificate
Silver	CDS 40–59	Certificate + Recognition
Bronze	CDS 20–39	Recognition
Participant	CDS < 20	Participation acknowledgment

Monetary Distribution: Coordinator Share = (Coordinator CDS / Sum of All CDS) × Sponsor Pool

Part 5: Cross-Validation System

5.1 Validation Clusters

Coordinators operating in overlapping areas are grouped into validation clusters. Requirements: Minimum 3 coordinators, minimum 20 responses per coordinator, same survey, same time window.

5.2 Statistical Comparison

For each survey question, answer distributions are compared using Jensen-Shannon Divergence (JSD): 0 = identical distributions, 1 = completely different.

JSD Score	Interpretation	Action
0.00–0.05	Excellent alignment	Trust boost (+2)
0.05–0.10	Acceptable variance	No change
0.10–0.20	Elevated variance	Flag for review
0.20+	Significant deviation	Trust penalty (−5), manual review

5.3 Handling Legitimate Outliers

To prevent penalizing accurate outliers:

Cluster Segmentation: If a coordinator's profile differs significantly, they form a separate validation sub-cluster.

Historical Comparison: Results compared to coordinator's own historical patterns.

Manual Override: Flagged cases reviewed by Tier 1 coordinators before trust penalties apply.

Part 6: Post-Stratification Weighting

6.1 Purpose

Raw survey data rarely matches population proportions. Post-stratification weighting adjusts for over/under-representation to produce population-representative estimates.

6.2 Weighting Cells

Dimension	Categories	Source
Region	Administrative units	Census data
Settlement	URB-1 through RUR-2	Census data
Gender	M, F, X	Census data
Age	A1–A6	Census data

6.3 Weight Calculation

Cell Weight = (Population Proportion in Cell / Sample Proportion in Cell)

Example: If population is 25% URB-1/Female/A2 but sample is 35%, weight = 0.25 / 0.35 = 0.71

6.4 Weight Trimming

Extreme weights distort variance. Weights are trimmed:

Trimmed Weight = max(0.2, min(5.0, Raw Weight))

6.5 Effective Sample Size

Weighting reduces statistical power. Report effective sample size:

$n\_eff = (\Sigma \text{ weights})^2 / \Sigma(\text{weights}^2)$

Part 7: Anti-Manipulation Design

7.1 Structural Defenses

Threat	Defense Mechanism
GPS spoofing	Cross-reference with IP, device fingerprint, movement patterns
Fake coordinators	Trust decay, referral accountability, minimum activity thresholds
Answer farming	Cross-validation detects anomalous uniformity
Demographic targeting	Diversity scoring discourages cluster concentration
Coordinator collusion	Random validation cluster assignment, rotation
Volume gaming	Zero reward for volume; diversity-only incentives

7.2 Detection Signals

• Velocity anomalies: Too many responses too quickly

• Pattern uniformity: Identical or near-identical answer sequences

• Geo-impossibility: Responses from locations too far apart in time

• Demographic skew: Extreme concentration in single demo-cluster

• Cross-validation failure: Persistent deviation from area peers

7.3 Response Actions

Signal Severity	Automatic Action	Human Review
Low	Flag for monitoring	No
Medium	Reduce trust score	Optional
High	Suspend data acceptance	Required
Critical	Suspend coordinator	Required

Part 8: Reporting Standards

8.1 Required Disclosures

All published results must include:

1. Methodology version used (e.g., GTPS-CV v1.0)

2. Raw sample size and effective sample size

3. Collection period and geographic scope

4. Weighting variables and trimming applied

5. Coverage gaps: Which geo/demo clusters are under-represented

6. Coordinator network size and average trust score

7. Cross-validation pass rate

8.2 Confidence Reporting

Results presented with: Point estimate, 95% confidence interval (accounting for design effect), and margin of error.

8.3 Limitations Statement

GTPS-CV produces probability-approximating samples through systematic field protocols, but is not a true probability sample. Results should be interpreted as indicative of population sentiment with the disclosed margins of uncertainty. Post-stratification weights adjust for known demographic imbalances but cannot correct for unmeasured biases.

Part 9: System Parameters

Configurable parameters with defaults:

Parameter	Default	Range	Description
trust_decay_rate	0.85	0.70–0.95	Trust multiplier for new recruits
learning_rate	0.10	0.05–0.20	Speed of trust adjustment
cv_margin_acceptable	0.10	0.05–0.15	JSD threshold for acceptable variance
weight_trim_lower	0.20	0.10–0.50	Minimum weight
weight_trim_upper	5.00	3.00–10.00	Maximum weight
gps_accuracy_threshold	50m	20–100m	Max acceptable GPS uncertainty
min_cluster_size_cv	3	2–5	Min coordinators for cross-validation
new_cluster_points	10	5–20	Points for new cluster coverage
gdi_weight	0.50	0.30–0.70	Weight of GDI in CDS
ddi_weight	0.50	0.30–0.70	Weight of DDI in CDS

Part 10: Governance and Sponsor Neutrality

10.1 Sponsor Restrictions

Sponsors cannot influence: Survey question wording or answer options, trust algorithm parameters, geographic targeting or exclusion, weighting methodology, data acceptance logic, coordinator selection or rewards, or diversity scoring formulas.

10.2 Sponsor Permissions

Sponsors may: Fund the reward pool, request specific geographic coverage (without exclusions), receive anonymized weighted aggregate results, and display brand acknowledgment in survey interface.

10.3 Editorial Independence

Survey design and analysis remain under platform editorial control. Sponsor-requested surveys undergo review to ensure questions are non-leading, answer options are balanced, and topic is appropriate for field collection.

Appendix A: Glossary

Term	Definition
Geo-cluster	Unique combination of GPS grid cell + situational location + settlement type
Demo-cluster	Unique combination of gender + age + occupation + income
CDS	Combined Diversity Score: weighted average of GDI and DDI
GDI	Geographic Diversity Index
DDI	Demographic Diversity Index
JSD	Jensen-Shannon Divergence: statistical measure of distribution difference
PSU	Primary Sampling Unit
Post-stratification	Statistical adjustment to align sample with population proportions
Trust decay	Reduction in initial trust score across coordinator tiers
Cross-validation	Comparison of results between coordinators in same area
n_eff	Effective sample size after weighting adjustment

Appendix B: Comparison with Original Methodology

Aspect	Original	Version 1.0
Location data	GPS only	GPS + Situational + Settlement
Demographics	Gender, Age	Gender, Age, Occupation, Income
Sampling protocol	Undefined	Systematic time-interval
Reward basis	Volume-influenced	Diversity-only (CDS-based)

Cross-validation	Fixed % threshold	Statistical (JSD)
Weighting	Geographic normalization	Full post-stratification
Trust formula	Unspecified	Published formula
Reporting standards	None	Required disclosures

Appendix C: Implementation Checklist

☐ Configure PSUs for target geography

☐ Load census benchmarks for weighting cells

☐ Set system parameters (or accept defaults)

☐ Onboard Tier 1 coordinators (10–20)

☐ Train coordinators on time-interval selection protocol

☐ Deploy mobile app with 3-part geo-cluster capture

☐ Deploy demographic collection UI with 4-part demo-cluster

☐ Establish validation cluster boundaries

☐ Configure leaderboard (diversity-ranked)

☐ Implement point decay formula for repeat clusters

☐ Draft limitations disclosure template

☐ Schedule first survey pilot (target: 500–1,000 responses)

☐ Post-pilot: Review cluster coverage gaps

☐ Post-pilot: Calibrate weighting parameters