AR&P

SocioEconomic Challenges, Volume 9, Issue 3, 2025
ISSN (print) – 2520-6621, ISSN (online) – 2520-6214

Intensity-Based Location Sampling Method For Investigating Socioeconomic Challenges: Ensuring External Validity In Surveys of Unknown Populations

László Letenyei, **ORCID:** https://orcid.org/0000-0002-7220-6871
Dr., Associate Professor, Institute of Social and Political Sciences, Corvinus University of Budapest, Hungary
Eliza Bodor-Eranus, **ORCID:** https://orcid.org/0000-0001-5672-3118
Dr., Researcher, Computational Social Sciences – RECENS Research Group, ELTE Centre for Social Sciences, Hungary
Gergely Horzsa, **ORCID:** https://orcid.org/0000-0003-2043-0633
Dr., Researcher, Institute for Sociology, ELTE Centre for Social Sciences, Hungary

Corresponding author: Eliza Bodor-Eranus, Bodor-Eranus.Eliza@tk.hu
Type of manuscript: research paper

Abstract: *Timely, representative evidence on users of public spaces and emergent venues is essential for diagnosing and addressing socioeconomic challenges. The Intensity-Based Location Sampling (ILS) method was developed to ensure the representativeness of populations without a prior sampling frame; its essence is the synchronous integration of real-time traffic counts with probability-based intercept surveys, allocating interviews in proportion to observed population density across zones (and, in its extension, across times). Compared with classic location or time-location sampling, external validity is strengthened, length-biased selection is reduced, and representative samples can be produced without a prior sampling frame. Scientific novelty lies in coupling counting and interviewing in real time and in an allocation rule that minimizes the discrepancy between survey and population intensities, distinguishing the approach from existing designs. The effectiveness of ILS is demonstrated on statistical data from a public urban park in Budapest, Hungary (2021): face-to-face interviews with N = 253 park visitors (embedded in a broader N = 1000 survey), with real-time traffic counts in nine zones and dynamic reallocation of fieldworkers to undersampled areas. The resulting ILS fit index (0.0179) indicated close correspondence between survey and observed distributions. These results illustrate how ILS can generate actionable, weighting-free evidence for planners and policy-makers, ensuring that interventions respond to real patterns of use and need. Beyond parks, the method is adaptable to transport hubs, markets, festivals, and crisis contexts, offering a scalable and cost-effective tool for socioeconomic research where populations are hidden, mobile, or undefined. By extending ILS with an explicit temporal component (ITLS – Intensity Based Time and Location Sampling), further prospects are opened for rigorous, low-cost study of Socioeconomic challenges in hidden, mobile, or otherwise undefined populations.*

Keywords: socioeconomic challenges, location sampling, time-location sampling, unknown population, external validity, traffic count, intensity, mixed methods, field research, data collection.
JEL Classification: C83, C81, R14, R53, O18.

Received: 18.07.2025 **Accepted:** 24.09.2025 **Published:** 03.10.2025

Funding: This work was supported by the National Research, Development and Innovation Office (Reference: OTKA FK 143024), the ELTE Centre for Social Sciences, Computational Social Science Research Group and Corvinus University of Budapest.
Publisher: Academic Research and Publishing UG (i.G.) (Germany).
Founder: Academic Research and Publishing UG (i.G.) (Germany).

Cite as: Letenyei, L., Bodor-Eranus, E., & Horzsa, G. (2025). Intensity-Based Location Sampling Method For Investigating SocioEconomic Challenges: Ensuring External Validity In Surveys Of Unknown Populations. *SocioEconomic Challenges*, *9*(3), 207-220. https://doi.org/10.61093/sec.9(3).207-220.2025

INTRODUCTION

Decisions about urban development, mobility, safety, health, culture, and public services increasingly depend on timely, representative evidence about how people use shared spaces. Public parks, marketplaces, transit hubs, festivals, and other emergent venues function as everyday infrastructures where inclusion, cohesion, and well-being are shaped. Reliable knowledge about who is present, when, and in what conditions is therefore essential for diagnosing and addressing Socioeconomic challenges.

Where evidence is most needed, measurement is most difficult. Many populations central to Socioeconomic challenges (e.g. precariously employed workers, migrants in transit, adolescents in informal settings, elderly individuals during heat events, visitors to temporary venues, or crisis response volunteers) lack conventional sampling frames. Population boundaries are fluid, denominators are unknown, and sites change across hours, days, and seasons. Reliance on small or convenience samples risks mischaracterizing inequalities and needs.

Sociological research in such environments faces recurring obstacles. Coverage and frame gaps arise when no comprehensive list of units exists and attendance fluctuates. Temporal heterogeneity alters composition across time slots. Length-biased selection favors individuals who dwell longer at a site. Spatial clustering within micro-areas (shade, play spaces, vendor clusters, transit nodes) produces local over- and under-representation. Nonresponse and approachability constraints vary with crowding, weather, events, and fieldworker visibility. Ethical and safety requirements limit intrusive designs and demand transparent, proportionate procedures. Cost and coordination pressures force on-the-spot allocation decisions, often without diagnostics to show whether the evolving sample still reflects the at-risk population.

Existing approaches only partly resolve these issues. Convenience or quota intercepts are quick but lack verifiable external validity in the case of hidden population. Classical location sampling structures where interviews occur yet seldom ties interview counts to observed population intensity, leaving bias where flows are uneven. Time-location sampling has been widely used for venue-based populations but depends on compiling and updating exhaustive lists of times and places and can still inherit length bias when sampling probabilities diverge from actual distributions across micro-locations and time slots. As venues proliferate and programming becomes more dynamic, frames become costly to maintain and quickly outdated.

Digital traces promise broader coverage but introduce distinct limitations. Camera or sensor counts, mobile-device pings, platform analytics, and administrative feeds approximate flows yet lack socio-demographic depth, depend on proprietary systems, under-represent groups with limited connectivity, and raise privacy concerns. Without linkage to direct interviewing, such traces cannot answer core sociological questions about attitudes, needs, norms, or informal practices. Overreliance on opaque technologies also risks reproducing bias.

The consequence is a persistent evidence gap especially where Socioeconomic challenges concentrate: unequal access to green space, mobility corridors marked by time poverty and safety concerns, cultural and recreational venues where participation mirrors segregation, and crisis-prone areas where resilience depends on knowing who is present and with what vulnerabilities, as well as who is present as volunteers offering assistance. When measurement lags reality, interventions are mistargeted: amenities are placed where least needed, opening hours mismatch demand, maintenance cycles ignore peak loads, and safety or health resources miss those at risk. Policy evaluations then become fragile, as shifts in flows are confounded with shifts in who is being sampled.

These conditions define practical requirements for field research in venues and populations without stable frames. Methods are needed that (i) operate *without a prior sampling frame* while retaining demonstrable external validity; (ii) *align the intensity of interviewing with the intensity of observed presence* across micro-locations and times to reduce length-biased selection; (iii) *provide real-time diagnostics* to reveal imbalances and guide the reallocation of field effort; (iv) maintain *ethical* transparency and protect privacy; and (v) *remain feasible* for municipal agencies, cultural institutions, and research teams at modest cost. Meeting these requirements would narrow the gap between what is measured and what matters for policies addressing Socioeconomic challenges.

In sum, the central difficulty is not only that relevant populations are hard to reach but also that their presence is hard to measure representatively in space and time. Strengthening the link between observed population flows and the allocation of interviewing offers a practical path to more reliable evidence—and, in turn, to better-informed responses to socioeconomic challenges.

SocioEconomic Challenges, Volume 9, Issue 3, 2025
ISSN (print) – 2520-6621, ISSN (online) – 2520-6214

LITERATURE REVIEW

Sampling in Frame-Deficient Settings: Conceptual Anchors

Empirical research in public and emergent venues frequently proceeds without a well-defined sampling frame. During fieldwork, it is rarely possible to obtain a true-random sample (Honigmann, 1982), and classic sampling methods are not always useful for answering all types of questions (Handwerker et al., 1997), especially when the population is not really known previously. The literature has therefore developed families of designs that approximate probability selection under fluid population boundaries (Kalton, 2001, 2019; Kapitány, 2010). Such sampling methods, summarized in Table 1, are known as *intensity sampling* (Patton, 2001), *extremity sampling* (Thompson, 1990), *adaptive sampling* (Thompson, 1990; Stout & Hardwick, 2005; Rogerson, 2005), *time* and/or *location sampling* (Sommen et al., 2018; Kalton, 1986), and *spatial sampling* (Haining, 2001). Each addresses recurrent threats to external validity (e.g., coverage error, length-biased selection, and time–space heterogeneity) but to different degrees and with different cost/feasibility profiles (Thompson, 1992; Leonet al., 2015; Sommen et al., 2018; Nowell & Stanley, 1991).

Table 1. Focal areas, goals, and some examples of comparable sampling methods of rare or hidden populations

Sampling type	Target	Goal	Metaphor for the technique	Examples
Intensity sampling (Patton, 2001)	Most frequented	Focus on the main element	Uranium enrichment	Eliminates most and least successful applicants
Extremity sampling (Thompson, 1990)	Rare population	Find rare cases	Gold panning	Identify most successful businesses
Adaptive sampling (Thompson, 1990; Stout & Hardwick, 2005; Rogerson, 2005)	Rare phenomenon	Find rare cases and intensify these in sample	Searching for a gold vein using gold-panning samples	Biostatistics, Cancer, HIV, Slot machines
Location sampling (Kalton, 2001)	Rare and hidden population	Find rare cases in a specific place	Lurking	Identify drug users
Time-location sampling (Sommen et al., 2018; Kalton, 1986)	Rare and hidden population	Find rare cases in a specific place and time	Hunting tower	Identify LMBTQ people who frequent gay parties
Spatial sampling (Haining, 2001; Rogerson, 2005)	Field, or a geographical area	Representative of parts of a field (regardless of users)	Urban planning	Urban studies

Source: edited by authors.

Intensity and *extremity sampling*, often used in exploratory or evaluative settings, can surface important mechanisms or rare outcomes but provide limited guarantees of representativeness if used alone. *Adaptive sampling* (originally developed for clustered or rare populations) offers a principled way to concentrate effort where signals are strongest while retaining design-based inference when selection rules are explicit and inclusion probabilities are tracked (Thompson, 1992). These strands set the stage for *TLS/VBS (time location / venue-based sampling)* and *spatial designs* that move closer to probability logic in open venues.

Time-Location/Venue-Based Sampling: Strengths and Limits

TLS/VBS formalizes selection in places and time blocks where the target population congregates and is utilized to select individuals who frequent specific locations such as libraries, museums, shopping centeres, and polling places (Kalton, 2001). Sampling can occur either upon entry to or exit from the location. While location sampling concentrates solely on the physical site, time-location sampling (TLS), also referred to as time-space sampling, incorporates temporal considerations alongside spatial factors. For instance, in the case of MSM (men who have sex with men), respondents are recruited in physical venues at times when they usually congregate, such as gay bars or clubs (Sommen et al., 2018). Typically, the sample design follows a two-stage process (Kalton, 1986), initially focusing on location selection followed by timing considerations. As it can be seen, TLS is used to sample members of a rare population and was successfully used in the case of individuals participating in illegal activities or stigmatized individuals, such as illicit drug users, men who have sex with men, and female sex workers (Sommen et al., 2018; Siegfried et al., 2023), and is also applicable in the case of populations at high risk of infectious diseases (Leon et al., 2015). A design-based foundation treats venue-day-time units as clusters and adjusts for unequal probabilities, including multiplicity due to differing attendance frequencies (Leon et al., 2015). Applications outside classic "hidden populations" demonstrate feasibility for general program participants when sites and windows are sampled probabilistically, and weights incorporate exposure (Siegfried et al., 2023). Empirical evidence further shows that failing to incorporate frequency of venue attendance (FVA) induces bias, while generalized weight-share adjustments can restore design consistency (Sommen et al., 2018). Two limitations recur in open public spaces. First, coverage can be incomplete where flows are diffuse or rapidly changing, even when venue lists are extensive. Second, length-biased selection persists within selected time frames if the chance of being approached is not explicitly tied to instantaneous presence (rather than dwell time) – a bias documented in intercept contexts (Bush & Hair, 1985) and conceptually linked to length-biased sampling in the statistical literature (Kish, 1965). These concerns motivate coupling TLS-style fieldwork with observed intensity so that interview allocation follows the population actually present.

Spatial and Spatiotemporal Allocation: Aligning Effort with Intensity

The purpose of spatial sampling is to collect samples across an area, whether one- or multidimensional (Haining, 2001). Spatial sampling encompasses random or systematic selection of samples within a defined geographic scope. This method is typically employed for estimating the overall mean of a parameter within an area, optimizing parameter estimations for unsampled locations, or predicting the movement of objects (Rogers, 2005). As a result, spatial/spatiotemporal sampling offers complementary tools for heterogeneous fields. Designs such as stratification across micro-areas, rotating panels, and generalized random tessellation stratified sampling seek broad, design-controlled coverage (Stevens & Olsen, 2012). Surveillance-oriented reviews emphasize that information yield improves when allocation tracks the underlying intensity surface (presence/flow) and adapts as conditions evolve (Jalilian et al., 2024). Targeting high-priority places and times identified by probabilistic spatiotemporal sampling improves cost efficiency and yields the most informative evidence on the study variable's spatiotemporal patterns. These ideas underpin designs that explicitly link where/when interviews occur to measured density.

Intercept Surveys, Nonresponse Dynamics, and Field-Period Effects

Intercept methods are attractive in public venues but are vulnerable to selection processes beyond location choice. Evidence shows that nonresponse bias is not a mechanical function of response rate; rather, it varies with mode and timing of fieldwork, underscoring the value of monitoring and within-field adjustments (Struminskaya et al., 2022). Responsive/adaptive designs operationalize such adjustments, reallocating effort toward strata with higher bias propensity without disproportionate cost (Peytchev et al., 2022). These findings support real-time diagnostics and dynamic reallocation as core features of venue-based field surveys when aiming for representativeness.

Counts and Sensors as Auxiliary Denominators for Design and Inference

Rapid progress in urban measurement has made pedestrian counts (from manual tallies to LiDAR and computer vision-based systems) both accessible and reliable. City-scale analyses demonstrate stable diurnal/seasonal structure in pedestrian flows and strong associations with built environment, implying that short-window counts can inform broader exposure measures (Bolin et al., 2021). Comparative sensor studies indicate that LiDAR (Light Detection and Ranging) maintains robust detection in varied light conditions, while video-based methods perform strongly in daylight; multi-sensor fusion improves reliability, supporting the use of counts as operational denominators for sampling and auxiliaries for calibration (Guan et al., 2023). In transport and safety analytics, validated pedestrian-volume models refine exposure estimation and exemplify the integration of counts and covariates for policy-relevant inference (Wolpert & Omer, 2025). Collectively, these studies justify treating observed presence as a size measure for probability-proportional-to-size allocation and as an auxiliary for estimator calibration.

Positioning Intensity-Based Location Sampling (ILS)

The reviewed methodological approaches converge on three validity threats that are specific to open, frame-deficient public spaces: (i) coverage error when venues/times are incompletely framed; (ii) length-biased selection when inclusion rises with dwell time; and (iii) time–space heterogeneity in composition within a field day or season. TLS/VBS provides a probabilistic scaffold but is sensitive to incomplete coverage and within-window bias. Spatial/spatiotemporal designs broaden coverage but do not, by themselves, ensure that samples reflect who is present at the times and micro-locations where people actually gather. Intercept-survey research and responsive-design studies argue for within-field monitoring and adaptive reallocation to reduce bias.

Building on these insights, Intensity-based Location Sampling (ILS) – as articulated in this manuscript – ties selection to observed intensity by (a) synchronizing real-time counts with probability-based intercepts; (b) allocating interviews in proportion to measured density across zones (and, by extension, across times); and (c) furnishing on-the-spot diagnostics to guide dynamic reallocation. When counts are further used as calibration totals, design-based inference benefits from auxiliary information (Deville & Särndal, 1992). This positioning shows how ILS addresses the main validity threats while remaining feasible for municipal, cultural, and public-health applications that confront socioeconomic challenges in transient populations.

The literature supports three practical commitments that structure the empirical application: (1) operate without a prior frame while preserving probabilistic logic; (2) align interviewing with observed presence to reduce length bias and improve external validity; and (3) monitor and adapt during fieldwork to correct emerging imbalances. By embedding ILS in a real-world public-space study and documenting diagnostics (fit between observed and achieved distributions), the present research advances a field-deployable blueprint consistent with contemporary survey-methods and spatial-sampling evidence.

Beyond its methodological novelty, ILS also addresses persistent blind spots in Socioeconomic research where conventional surveys struggle. For example, it can improve migrant visibility in urban spaces, it captures the distribution of environmental amenities and burdens, it's feasible for studying informal labor in public venues (Barratt et al., 2015; Manski, 1993; Bolin et al., 2021; Wolpert & Omer, 2025; Kalton, 2001; Singh & Masuku, 2014) or other transient groups (such as volunteers in crisis responses or at mega events). Barratt et al. (2015) reported notable demographic and drug use differences, arguing that online purposive samples of hidden populations should be interpreted alongside probability-based surveys and ethnographic fieldwork. Bolin et al. (2021) found that pedestrian counts depend mainly on built density and street type, with markets and transit stops having stronger effects than schools. Temporal fluctuations are driven by density rather than street type, confirming earlier findings, and the results are useful for urban design, though city-specific differences limit generalization. Wolpert and Omer (2025) reported that machine learning models best predict city-wide pedestrian flows, while agent-based models work better at the neighborhood scale; regression analyses perform poorly overall. However, as a response to the problems related by Manski (1993) – extrapolation of regressions, the selection problem, identification of endogenous social effects, and identification of subjective phenomena – some of the answers are related by Singh and Masuku (2014), who are arguing that different research problems require different sampling techniques, as no single method is suitable for all cases.

By aligning interviewing with real-time presence, ILS enables the generation of representative benchmarks in precisely those contexts where socioeconomic inequalities concentrate but evidence is scarce. This bridging function underscores the approach's relevance not only for survey methodologists but also for urban policy-makers, labor researchers, market researchers and environmental justice scholars.

METHODOLOGY

Design Overview

A design-based Intensity-Based Location Sampling (ILS) protocol was implemented to obtain externally valid estimates from an unknown, mobile population in a public venue (Sudman, 1980). ILS synchronizes short-interval traffic counts with probability-based surveys and allocates interviewing proportional to observed population intensity across micro-locations (zones) and – where applicable – time windows. The objective is to minimize the discrepancy between the achieved survey distribution and the concurrently observed population distribution, thereby reducing length-biased selection while operating without a prior sampling frame.

Setting and Frame

The demonstration study was conducted in Városmajor Park (Budapest, 2021) as part of a larger survey (total $N = 1{,}000$), of which respondents from park visitors constituted $N = 253$. To create a workable spatial frame, the park was subdivided into nine zones using rapid ethnographic reconnaissance and mental mapping (Letenyei & Dobák, 2020) with local informants (e.g., maintenance staff, vendors), followed by pilot counts to verify boundaries and visibility. Fieldwork covered daytime hours on weekdays and weekends; nighttime hours were excluded due to negligible attendance.

Traffic Counting (Auxiliary Measure)

Within each active zone, fieldworkers conducted brief, repeated person counts at regular intervals and recorded them into a shared online sheet. For each zone i, population density and density rate (zone share of total observed attendance) were computed in real time. Where feasible, non-identifying attributes (e.g., broad age bands, apparent group composition) were also noted to support diagnostics without capturing personal data.

Probability-Based Intercept Within Zones

Immediately after a count, one individual was selected at random from those currently present using a simple, pre-specified systematic walk rule adapted to local density (a fixed-interval selection along a clockwise path starting from a random bearing). If the selected person refused or was ineligible (e.g., underage threshold), the next interval unit was approached. After each completed contact (interview or refusal), the fieldworker resumed counting, producing an alternating count–approach–survey cycle. No directly identifying information was collected; participation was voluntary with oral consent.

Real-Time Allocation ("Compass")

A lightweight coordination tool ("Compass") continuously compared survey density with population density rate (Table 2).

Table 2. Compass: ILS fieldworkers' destination management tool

1	2	3	4	5	6	7	8
ZONE	Number of observations	Observed units	Population density (3/2)	Population density rate	Number of surveys	Survey intensity rate	COMPASS (5/7)
CODE	piece	person	person	%	piece	%	Undersampled, oversampled, or optimal

Source: https://docs.google.com/spreadsheets/d/1G-q22caPLnqWzoiN6kTH3fgJXeL92rd-JrNZTydqAXA/edit?gid=2045934498#gid=2045934498.

Note: Legend for the Compass: Column 1: Zone identifier code. Column 2: Number of observations in the zone - indicates how many times a fieldworker undertook traffic counting. Column 3: Observed attendees - denotes the number of observed people in the zone. Column 4: Intensity or population density - calculated as Column 3 divided by Column 2 (average traffic per zone). Column 5: Intensity rate or population density rate (%) - compares the attendees in the zone to the total attendees. Column 6: Number of surveys - variable indicating successful surveying (1 = successfully filled-in questionnaire). Column 7: Survey intensity rate or sampling rate (%) - compares surveys conducted in a zone to the total. Column 8: The Compass - an orientation tool for fieldworkers calculated using the Intensity rate (Column 5) and Survey intensity rate (Column 7). The database and the Compass are available online: https://docs.google.com/spreadsheets/d/1G-q22caPLnqWzoiN6kTH3fgJXeL92rd-JrNZTydqAXA/edit?gid=2045934498#gid=2045934498.

The overall goal of the data collection made with the ILS method was to minimize the difference between the *survey density rate* and the *population density rate* of the survey:

$$min\left|\frac{r_z}{\sum_{i=1}^{n} r_i} - \frac{p_z}{\sum_{i=1}^{n} p_i}\right| \text{ for each zone } z, \quad (1)$$

where n is the number of zones, r_i, is the aggregated number of respondents in zone i, and $p_{i,t}$ is aggregated number of observed people in zone i.

Zones were flagged as *undersampled*, *oversampled*, or *balanced*. Survey fieldworkers continuously checked during the data collection if oversampling took place in the zone (z) where they actually worked. In operational terms, if

$$\frac{r_z}{\sum_{i=1}^{n} r_i} - \frac{p_z}{\sum_{i=1}^{n} p_i} > 0, \quad (2)$$

then they look for respondents in another zone z^*, preferably in the vicinity of z, in which this difference is negative. Fieldworkers were instructed to relocate *from oversampled to undersampled* zones subject to safety and access. This procedure aligns interviewing with instantaneous presence rather than dwell time, mitigating length bias.

Diagnostic Fit and Estimation

Design performance was summarized by an ILS fit index: the square root of the sum of squares of differences. Values near 0 indicate close agreement between survey and observed distributions; large values indicate misalignment. Because selection is tied to observed intensity and monitored during fieldwork, post-survey weighting was not required in this application; descriptive estimates were reported unweighted with zone-robust standard errors.

Sample Size, Team, and Training

The park-visitor respondents' sample achieved N = 253 via rotating 2–4 interviewer teams over multiple dayparts. Non-responses were also recorded. A standardized briefing and field manual (Figure 1) covered zone definitions and handoffs, the count protocol, the intercept rule, refusal scripting, device-based data entry, and ethics (consent script, bystander privacy). A pilot day was used to verify zone visibility and to calibrate the reallocation tolerance.

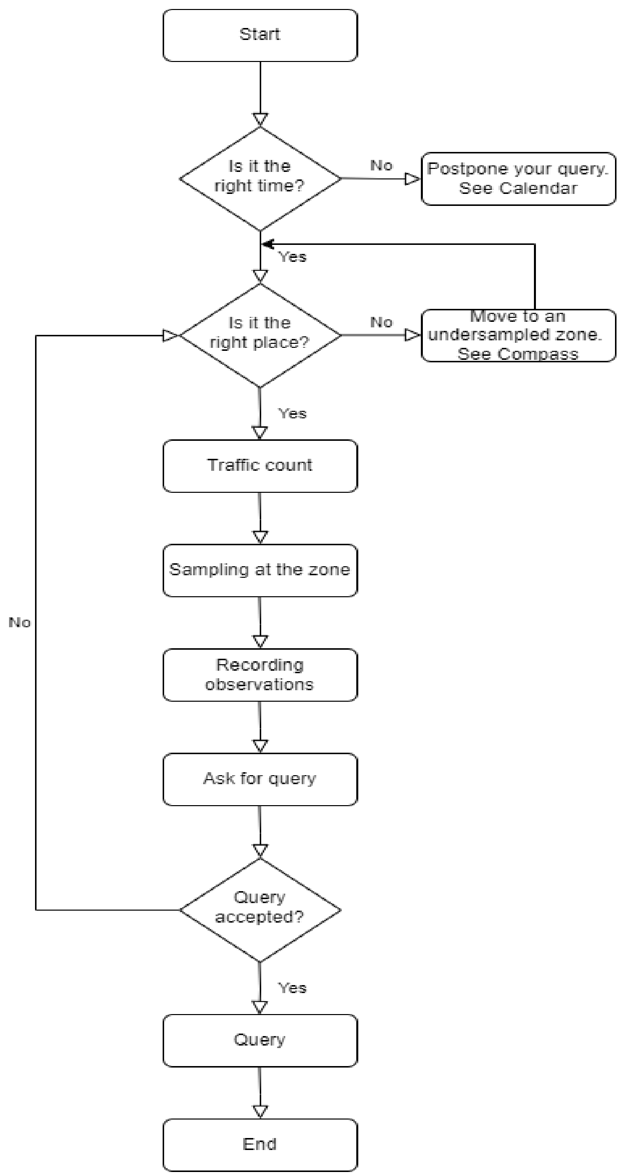

Figure 1. ILS fieldworkers' manual

Source: own processing, created with app.diagrams.net.

Data Quality and Ethics

Quality controls included spot back-checks (supervisor shadowing), automated range and logic checks in the questionnaire, and timestamp/GPS stamps for session auditing (where permitted). The protocol used aggregate counts as auxiliaries, did not record names, contact details, photos, or precise identifiers, and complied with standard principles of minimal risk research in public spaces (brief information, voluntary participation, immediate withdrawal option). The public sponsor authorized use of de-identified data for research and publication; a public link provides a limited, non-sensitive extract of the operational sheet.

A temporal extension (ITLS) applies the same principles to time slots as strata, targeting alignment over [t_0, t_c] and enabling diagnostics for both space and time; this extension is summarized in the manuscript's discussion.

RESULTS

Field Execution and Achieved Sample

As noted earlier, traffic counts were conducted repeatedly in each of the nine zones of the investigated park and immediately paired with probability-based intercepts. Counts and interviews were logged in a shared sheet that refreshed proportions in real time to guide fieldworker reallocation (Figure 2).

Figure 2. Zones of Városmajor Park, based on mental mapping

Source: own processing, created with Google Maps.

Partitioning the park into zones was crucial for analyzing the overall fit between the share of visitors and the share of survey respondents by zone. The duration of fieldwork covered nine days.

Alignment between Observed Intensity and Achieved Interviews

For measuring sample fitness, an index was applied by taking the square root of the sum of squares of differences. The result may have a value of between 0.000 and 1.000, with the former meaning total similarity between the two sets of data and the latter meaning complete difference.

Table 3 compares each zone's share of total traffic count (population intensity) to its share of interviews (survey intensity). Absolute zone-level mismatches were small (range ≈ 0.24–0.86 percentage points), and the sum of squared differences was 0.0003. Zone 8 showed the largest absolute gap, $|\Delta| = 0.0086$; Zones 5 and 3 showed the smallest, $|\Delta| = 0.0024$ and $|\Delta| = 0.0045$, respectively.

SocioEconomic Challenges, Volume 9, Issue 3, 2025
ISSN (print) – 2520-6621, ISSN (online) – 2520-6214

Table 3. Fitness of sample

Zones	Share of visitors per zone according to the traffic count (%)	Share of survey respondents per zones (%)	Differences
Zone No. **1**	**14.48**	**13.94**	**0.0054**
Zone No. **2**	**8**	**7.57**	**0.0043**
Zone No. **3**	**9.91**	**10.36**	**-0.0045**
Zone No. **4**	**21.64**	**22.31**	**-0.0067**
Zone No. **5**	**14.18**	**13.94**	**0.0024**
Zone No. **6**	**5.78**	**6.37**	**-0.0059**
Zone No. **7**	**9.85**	**9.16**	**0.0069**
Zone No. **8**	**7.51**	**8.37**	**-0.0086**
Zone No. **9**	**8.64**	**7.97**	**0.0067**
Sum	**1**	**1**	
		Sum of squares	**0.0003**
		Index of ITLS fit	**0.0179**

Source: authors' calculation.

The overall ILS fit index – the square root of that sum – was 0.0179, indicating a close correspondence between the achieved sample and observed presence; consequently, post-survey weighting was not required for descriptive statistics.

Nonresponse and Process Indicators

Non-respondents were recorded explicitly within the ILS protocol; basic observable characteristics were noted when possible, and refusals were marked as such. The refusal rate was 33.6%: the number of intercepted persons was 459, of which non-respondents were 154. In a further 52 cases, respondents reported their park visitation was occasional or extremely rare, and thus, they were not included in the survey (they wouldn't be able to answer the questions regarding the needed park development). This provides a final N = 253 database, which is 55.1% of those approached. Non-respondents follow respondents' reported estimated age group and gender patterns.

The real-time "Compass" sheet continuously flagged undersampled and oversampled zones, prompting within-day reallocation of fieldworkers and contributing to the observed alignment of survey and traffic distributions.

Benchmarks For the Venue Population

Even without a prior sampling frame, the synchronized count – interview design enabled venue – level population benchmarks. For example, regarding demographic composition, we observed an even gender distribution among attendees. Similarly, attendees of all age groups were present during the investigated period. We also observed and estimated other features, like children or dog ownership and social status. All these features were recorded when choosing a respondent but before asking the questions of the survey.

The surveys' results were in concordance with the observations. Demographically, the survey indicates that 46% of respondents identified as male, while 78% reported being married. Furthermore, 45% of respondents indicated they had children, and 41% reported ownership of a dog. In terms of educational attainment, a notable 70% of respondents claimed to have attained higher education qualifications. Moreover, 33% of respondents self-reported a social status above the statistical mean.

According to the geographical distribution of the catchment zone, the survey revealed that 7% of respondents hailed from the 1st district of Budapest, 39% from the 2nd district, and a significant proportion, amounting to 32%, originated from the 12th district, where the park under consideration is located. Additionally, 23% of respondents were sourced from other districts within Budapest.

These findings underscore the diverse demographic composition of respondents, shedding light on their geographical origins, familial statuses, pet ownership, educational backgrounds, and perceived social statuses. Likewise, we estimated attitudes towards proposed development ideas relative to general attendee parameters. A primary research goal was to gauge park attendees' attitudes towards a government-planned green corridor connecting various city parks. Although this would result in the elimination of several parking spots, it would contribute to reduced automotive traffic and more pedestrian-friendly streets. The question was whether attendees visiting parks at different frequencies would support such development under varying circumstances.

Attitudes and Park-Use Priorities

Attitudinal items covered a city green-corridor proposal (with potential parking-space reductions) and the relative importance of park functions. The design enabled relating these attitudes to general attendee parameters (e.g., visit frequency), although detailed cross-tabulated results are not reported in the current draft.

Using a Likert scale ranging from 1 to 5, we assessed the perceived importance of various park functions among respondents (Figure 4). Results indicated that the most crucial functions included children's playgrounds, restroom facilities, botanical gardens, and areas designated for dog walking, all scoring above 4 on the scale average. Additionally, respondents emphasized the significance of amenities such as free sports grounds, water surfaces, food kiosks, cycling paths, outdoor fitness areas, running tracks, outdoor stages, and spaces for outdoor events, with an average score of 3 on the scale. While the current configuration of the park encompasses many of these functionalities, their order of importance was elucidated through our study. Furthermore, we gauged respondent satisfaction with different park functions, revealing a notable disjunction between the perceived importance and actual satisfaction levels of these amenities.

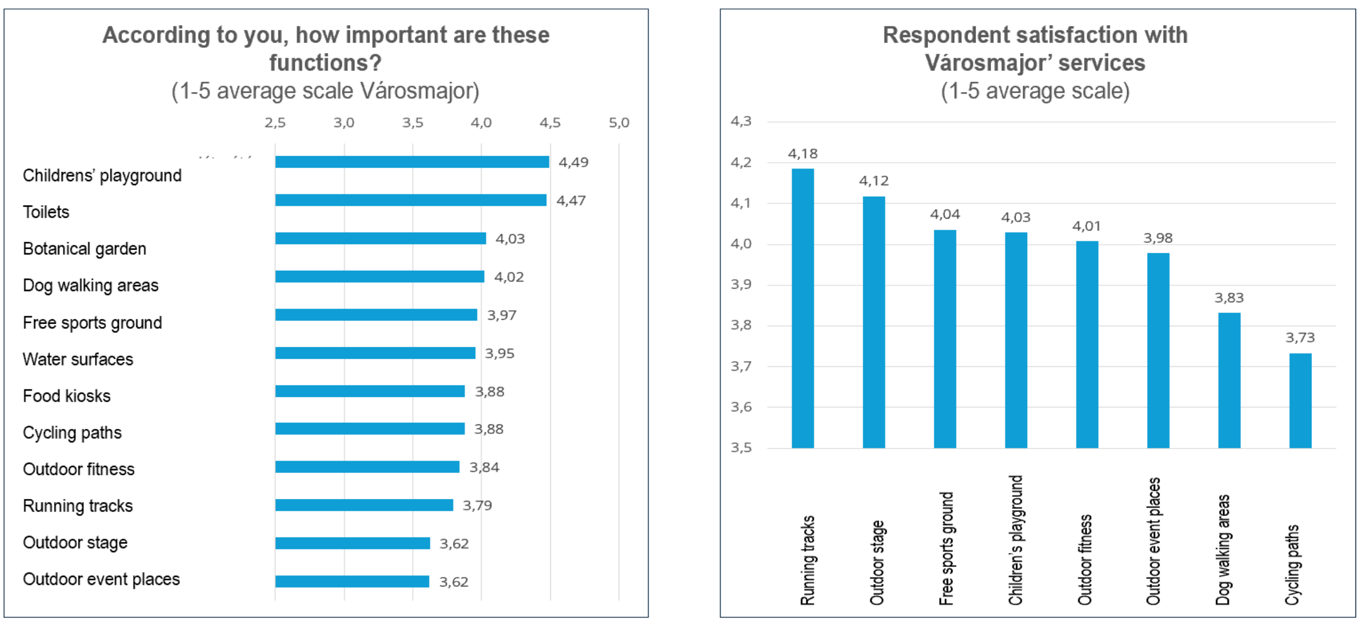

Figure 4. Percieved importance of Park Functions and Satisfaction

Source: authors' calculation.

Understanding attendance frequency differences posed a challenge due to the relatively narrow research timeframe. However, the applied methodology offered a solution to these challenges. Additionally, we distinguished between park attendees with varying frequencies of attendance. Results indicated significantly higher support for building a green corridor between parks and eliminating some parking spots among those who visited parks less frequently – contrary to initial hypotheses. Moreover, we identified generally high support for the green corridor and multifunctional reconstruction of the respective city park.

Mechanism Checks and Threats to Validity

The protocol explicitly targets length-biased selection by tying approach opportunities to instantaneous presence (zone-level counts), then aligning interview allocation accordingly. The real-time sheet operationalized this by highlighting discrepancies between survey and population intensity rates to trigger movement of interviewers. Two practical limitations are acknowledged. The limitation regarding temporal coverage means that only daytime windows

SocioEconomic Challenges, Volume 9, Issue 3, 2025
ISSN (print) – 2520-6621, ISSN (online) – 2520-6214

were included; seasonal and event effects (e.g., heat, rain, special programming) were not measured. The limitation regarding sample size and movement means that the venue module size (N = 253) constrains precision; double counting is possible when visitors traverse multiple zones or dwell long in one zone, though the method's focus on instantaneous presence reduces the inferential impact of dwell-time heterogeneity.

Substantive Implications for Venue Planning

Three readouts are policy-relevant given the city's planning context. The small ILS fit index indicates that observed flows were proportionately represented in interviewing, supporting weighting-free descriptive summaries for the park's daytime users. The importance–satisfaction disjunction suggests where marginal investments may deliver high user benefit (e.g., restrooms, playgrounds, dog areas), though numeric effect sizes would benefit from explicit reporting of mean differences and confidence intervals. The concentration from Districts II and XII implies differing exposure and equity considerations across neighborhoods; targeted outreach or access interventions could be evaluated in follow-up ILS runs.

The results demonstrate that synchronizing counts and intercepts and reallocating field effort toward undersampled zones can control within-window selection without a prior frame – an acknowledged weakness of classic TLS/VBS in open public spaces. The near-zero fit index is empirical evidence that interview allocation tracked population presence, the core mechanism posited in the review for reducing length bias and improving external validity.

The ILS implementation – counts as auxiliary denominators and real-time allocation – aligns with design-based principles (allocation according to an intensity surface and calibration to reliable totals) highlighted in the literature review. The venue results are consistent with that logic: small spatial discrepancies, interpretable venue benchmarks, and no need for post weights for the descriptive aims of this module.

In some cases, added data would strengthen the contribution of ILS. One of these is the consideration of seasonality and events: the extension of coverage to night hours and multiple seasons and the tagging of weather and programming would help to analyze if temporal ILS fits to verify stability across conditions. Another feature to be considered is the dwell time/re-approach risk: adding dwell-time tallies or "time-since-entry" proxies to directly quantify residual length bias and to estimate the probability of re-approach for movers across zones would also improve the methods' validity. A third aspect is that of precision and variance: an increased N and a report of zone-robust standard errors (or replicate-day variance) are key proportions and differences (e.g., importance–satisfaction gaps). The current *N* = 253 limits precision for subgroup contrasts. Finally, the external benchmarks and sensors (digital/video counting) would also help the real-time planning of data collection based on venue intensity.

Extension of the Sampling Method

The potential extension of this method is by applying the time dimension into the sampling procedure. The difference between the ILS and ITLS (intensity-based time-location sampling) is that the later one take into account the exact time of the sampling.

In the case of Intensity-based Time Location Sampling (ITLS), the goal is to minimize the difference between the survey density rate and the population density rate at the t_c closing time of the survey:

$$min\left|\frac{r_{z,t_c}}{\sum_{i=1}^{n} r_{i,t_c}} - \frac{p_{z,t_c}}{\sum_{i=1}^{n} p_{i,t_c}}\right| \text{ for each zone } z, \quad (3)$$

where t_0 and t_c are, respectively, the opening and closing times of the data collection, *n* is the number of zones, $r_{i,t}$ is the aggregated number of respondents in zone *i* at time *t*, and $p_{i,t}$ is the aggregated number of observed people in zone *i* at time *t*.

In conclusion, the ILS/ITLS mechanism described in the manuscript performs as intended in this pilot: achieved interviews closely mirrored observed flows, enabling credible, weighting-free venue benchmarks and priority readouts for planning. Extending coverage in time, enriching process data on dwell and nonresponse, and formalizing precision estimates would convert a strong proof of concept into a generalizable template for studies of public-space users under socioeconomic challenges.

CONCLUSIONS

This study implemented an Intensity-based Location Sampling (ILS) protocol to generate externally valid evidence from an unknown, mobile population in a public urban park. By synchronizing short-interval traffic counts with probability-based intercepts and allocating interviews in proportion to observed presence, the design directly targeted coverage gaps, length-biased selection, and time–space heterogeneity—three central threats identified in the literature for frame-deficient settings.

Main Empirical Patterns

The results show that interview allocation closely tracked the contemporaneous distribution of visitors across zones, yielding a very low ILS fit index (0.0179) and small zone-level discrepancies. In practical terms, the achieved sample mirrored who was actually present in the park at given times, strengthening external validity without requiring post-hoc weighting. Despite the absence of a prior sampling frame, the design produced interpretable benchmarks of venue composition: gender balance, age profile, household characteristics, education levels, and catchment by district. Importantly, the method also revealed socioeconomic asymmetries. Attendance concentrated disproportionately in adjacent districts, highlighting unequal access and exposure to green infrastructure across neighborhoods. Attitudinal measures further pointed to gaps between the importance and satisfaction of amenities – notably restrooms, playgrounds, and dog-walking areas – indicating where marginal investments could generate significant improvements in well-being. Finally, attitudes toward a planned green corridor suggested that environmental sustainability initiatives can attract substantial support, particularly among less frequent park visitors, counter to initial expectations.

Limitations

Three main limitations should be acknowledged. First, the temporal scope was restricted to daytime windows during a single season; thus, seasonal variations, weather shocks, special events, and night-time use remain unmeasured. Second, while the venue sample (N = 253) was sufficient for descriptive benchmarks, it limited precision for subgroup analyses and constrained the ability to examine interactions such as age and zone or attitudes and visit frequency. Third, residual risks remain concerning re-approach and dwell-time bias: although selection was tied to instantaneous presence, individuals with longer stays may still have faced higher approach probabilities. Lightweight dwell-time indicators and refusal typologies could strengthen diagnostics in future applications. Additionally, while manual counts provided a feasible denominator, sensor validation (e.g. video spot checks) would improve measurement reliability. Finally, as this was a single-site demonstration, further replication in diverse settings (such as transit nodes, markets, waterfronts, crisis responses volunteers) is required to establish generalizability.

Prospects for Studying Social-Economic Challenges

Most importantly, ILS and its temporal extension (ITLS) open new avenues for studying socioeconomic challenges in contexts where traditional sampling cannot deliver. By aligning interviewing with real-time flows, the method provides representative evidence on populations that are otherwise hidden, transient, or poorly captured in official registers. Applications include:

Urban health and climate resilience: tracking who is exposed to heat, shade, or hydration facilities under different weather conditions; evaluating targeted micro-interventions such as misting stations or shaded rest areas.

Mobility equity and time poverty: assessing crowding, waiting times, and safety conditions at transport hubs, and informing more equitable scheduling or pedestrian infrastructure investments.

Environmental justice: documenting unequal access to green spaces, amenities, and public infrastructure across neighborhoods, enabling cities to target resources where gaps are largest.

Cultural participation and inclusion: capturing attendance patterns at festivals, markets, and temporary venues, including groups excluded from standard surveys, and identifying barriers to equitable participation.

Informal labor and livelihoods: providing non-stigmatizing, probability-oriented evidence on street vendors, gig workers, and other informal workers who operate in public space.

Crisis response and displacement: deploying rapid ILS modules to identify who is present during protests, service disruptions, or emergencies, guiding the siting of assistance points and interventions.

Taken together, these applications show that ILS is not only a methodological innovation but also a substantive tool for socioeconomic inquiry. It equips policy-makers, urban planners, and social scientists dealing with socioeconomic challenges with reliable, low-cost, and scalable evidence in precisely those contexts where inequalities concentrate and existing data are weakest. By extending coverage across time, enriching process data on dwell and nonresponse, and integrating sensor-assisted validation, future studies can enhance the robustness of this approach. Yet even in its current form, ILS demonstrates that probability logic can be operationalized in open venues with modest resources, producing evidence that directly informs interventions in urban equity, resilience, and inclusion.

Author Contributions

Conceptualization: G. H., L. L., E. B. E.; data curation: L. L., G. H..; formal analysis: G. H.; funding acquisition:; investigation: L. L., G. H., E. B. E.; methodology: G. H., L. L., E. B. E..; project administration: L. L.; resources; software:; supervision: L. L., G. H., E. B. E.; validation: L. L., G.H.; visualization: L. L., G. H..; writing – original draft: E. B. E., G. H., L. L.; writing – review & editing: E. B. E., G. H., L. L.

Acknowledgements

Authors thank László Füstös, Gábor Péli, and Péter Futó for their suggestions and Zsófia Eördögh for the discussions about the fieldwork. Authors thanks for the BFK for the approval for using the data of the applied research for an academic publication.

Conflicts of Interest

Not applicable.

Data Availability Statement

Data collection was carried out on behalf of the Budapest Development Center (BFK).

The database and the Compass are available online:

https://docs.google.com/spreadsheets/d/1G-q22caPLnqWzoiN6kTH3fgJXeL92rd-JrNZTydqAXA/edit?gid=2045934498#gid=2045934498

Informed Consent Statement

The authors have obtained and maintained oral consent from all subjects involved in the study. No personal data was recorded through data collection; respondents cannot be identified through data analysis.

List of abbreviations

ILS – Inetnsity-based location sampling
ITLS – Intesity-based time and location sampling
TLS – Time location sampling
VBS – venue-based sampling

References

1. Barratt, M. J., Ferris, J. A., & Lenton, S. (2015). Hidden populations, online purposive sampling, and external validity: Taking off the blindfold. *Field Methods, 27*(1), 3–21. [CrossRef]
2. Bolin, D., Verendel, V., Berghauser Pont, M., Stavroulaki, I., Ivarsson, O., & Håkansson, E. (2021). Functional ANOVA modelling of pedestrian counts on streets in three European cities. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 184*(4), 1176–1198. [Link]
3. Bush, A. J., & Hair, J. F. (1985). An assessment of the mall intercept as a data collection method. *Journal of Marketing Research, 22*(2), 158–167. [CrossRef]
4. Deville, J.-C., & Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association, 87*(418), 376–382. [Link]
5. Guan, F., Xu, H., & Tian, Y. (2023). Evaluation of roadside LiDAR-based and vision-based multi-model all-traffic trajectory data. *Sensors, 23*(12), 5377. [Link]
6. Haining, R. P. (2001). Spatial sampling. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (pp. 14822–14827). Pergamon. [Link]
7. Handwerker, W. P., Hatcherson, J., & Herbert, J. (1997). Sampling guidelines for cultural data. *Cultural Anthropology Methods, 9*(1), 7–9. [CrossRef]
8. Holmes, J. (1967). Problems in location sampling. *Annals of the Association of American Geographers, 57*(4), 757–780. [CrossRef]
9. Honigmann, J. J. (1982). Sampling in ethnographic fieldwork. In R. G. Burgess (Ed.), Field research: A sourcebook and field manual (pp. 79–90). London: Routledge. [Link]

10. Jalilian, A., Mateu, J., & Sedda, L. (2024). A brief review and guidance on the spatiotemporal sampling designs for disease vector surveillance. *Current Research in Parasitology & Vector-Borne Diseases, 6,* 100208. [Link]
11. Kalton, G. (2001). Practical methods for sampling rare and mobile populations. In *Proceedings of the Annual Meeting of the American Statistical Association*. [Link]
12. Kalton, G. (2019). Developments in survey research over the past 60 years: A personal perspective. *International Statistical Review, 87*(S1), S10–S30. [CrossRef]
13. Kalton, G., & Anderson, D. W. (1986). Sampling rare populations. *Journal of the Royal Statistical Society. Series A (General), 149*(1), 65–82. [CrossRef]
14. Kapitány, B. (2010). Mintavételi módszerek ritka populációk esetén. *Statisztikai Szemle, 88*(7–8), 739–754. [Link]
15. Kish, L. (1965). *Survey sampling*. Wiley. [Link]
16. Leon, L., Jauffret-Roustide, M., & Le Strat, Y. (2015). Design-based inference in time-location sampling. *Biostatistics, 16*(3), 565–579. [CrossRef]]
17. Letenyei, L., & Dobák, J. (Eds.). (n.d.). *Mental mapping. The science of orientation*. Schenk Verlag. [Link]
18. Manski, C. F. (1993). Identification problems in the social sciences. *Sociological Methodology, 23*, 1–56. [CrossRef]
19. Nakanishi, N. (1978). Indefinite-metric quantum field theory of general relativity. *Progress of Theoretical Physics, 59*(3), 972–985. [CrossRef]
20. Nowell, C., & Stanley, L. R. (1991). Length-biased sampling in mall intercept surveys. *Journal of Marketing Research, 28*(4), 475–479. [CrossRef]
21. Patton, M. Q. (2015). *Qualitative research & evaluation methods* (4th ed.). SAGE. [CrossRef]
22. Peytchev, A., Pratt, D., & Duprey, M. (2022). Responsive and adaptive survey design: Use of bias propensity during data collection to reduce nonresponse bias. *Journal of Survey Statistics and Methodology, 10*(1), 131–148. [CrossRef]
23. Rogerson, P. A. (2005). Spatial sampling. In K. Kempf-Leonard (Ed.), *Encyclopedia of social measurement* (pp. 633–638). Elsevier. [CrossRef]
24. Siegfried, Y., DeMatteis, J. M., & Gollapudi, B. (2023). Implementation of a sample design for a survey of program participants using time-location sampling. *PLOS ONE, 18*(5), e0285877. [CrossRef]
25. Singh, A., & Masuku, M. (2014). Sampling techniques and determination of sample size in applied statistics research: An overview. *International Journal of Commerce and Management, 2*, 1–22. [Link]
26. Sommen, C., Saboni, L., Sauvage, C., Alexandre, A., Lot, F., Barin, F., et al. (2018). Time location sampling in men who have sex with men in the HIV context: The importance of taking into account sampling weights and frequency of venue attendance. *Epidemiology and Infection, 146*(7), 913–919. [CrossRef]
27. Sommen, C., Praud, D., Régnier, F., et al. (2019). Time-location sampling in men who have sex with men in the HIV context: The importance of taking into account sampling weights and frequency of venue attendance. *Epidemiology & Infection, 147*, e201. [CrossRef]
28. Ste Stevens, D. L., Jr., & Olsen, A. R. (2004). Spatially balanced sampling of natural resources. *Journal of the American Statistical Association, 99*(465), 262–278. [CrossRef]
29. Stigler, S. M. (1990). *The history of statistics: The measurement of uncertainty before 1900* (Reprint ed.). Belknap Press of Harvard University Press. [Link]
30. Stout, Q. F., & Hardwick, J. (2005). Adaptive designs. In *Handbook of parallel computing and statistics*. Chapman & Hall/CRC. [CrossRef]
31. Struminskaya, B., & Gummer, T. (2022). Risk of nonresponse bias and the length of the field period in a mixed-mode general population panel. *Journal of Survey Statistics and Methodology, 10*(1), 161–182. [CrossRef]
32. Sudman, S. (1980). Improving the quality of shopping center sampling. *Journal of Marketing Research, 17*(4), 423–431. [CrossRef]
33. Thompson, S. K. (1990). Adaptive cluster sampling. *Journal of the American Statistical Association, 85*(412), 1050–1059. [CrossRef]
34. Thompson, S. K. (1992). Sampling. *Journal of the American Statistical Association, 87*(418), 1054–1067. [Link]
35. Wolpert, L., & Omer, I. (2025). Comparative analysis of pedestrian volume models: Agent-based models, machine learning methods and multiple regression analysis. *Computers, Environment and Urban Systems, 117*, 102238. [CrossRef]