CME LIVE OPEN ACCESS

Advances
Brain Research and Neuroscience

Research Article

Efficient Detection of Rare Disease Hotspots Using Fuzzy Adaptive Cluster Sampling under Inhomogeneous Spatial Point Processes

Dr. Janardan Behera*

Department of Statistics, Ravenshaw University, India

Corresponding Author: Janardan Behera, Department of Statistics, Ravenshaw University, India.

Received: 2026 Jan 02 **Accepted:** 2026 Jan 29 **Published:** 2026 Feb 04

Abstract

Reliable detection and estimation of rare disease hotspots are fundamentally hindered by low prevalence, weak spatial signals, and severe operational constraints in surveillance systems. Classical adaptive cluster sampling, although theoretically suitable for rare events, often suffers from sharp instability due to rigid threshold-based triggering and uncontrolled cluster expansion. This study develops a novel fuzzy adaptive cluster sampling framework for hotspot detection and inference under inhomogeneous spatial point processes by replacing binary triggering with probabilistic, membership- driven adaptive expansion. A complete design-based inferential structure is established through de- fuzzified Horvitz–Thompson type estimators, Monte Carlo approximation of inclusion probabilities, and network-level variance estimation, with formal proofs of design unbiasedness, consistency, and variance stabilization. Extensive simulation experiments based on inhomogeneous Poisson spatial point process models with embedded rare disease hotspots demonstrate that the proposed fuzzy adaptive design consistently achieves higher detection sensitivity under weak and diffuse clustering while maintaining controlled false alarm rates and stable sampling effort. In-depth sensitivity tests looking at fuzziness levels, contrast between hotspots and backgrounds, spatial neighborhood patterns, and how rare the disease is further back up how strong and dependable this system is in practice. The results establish fuzzy adaptive cluster sampling as a statistically principled, op- erationally stable, and practically implementable methodology for modern rare disease surveillance under spatial uncertainty.

Keywords: Fuzzy Adaptive Cluster Sampling, Rare Disease Surveillance, Spatial Point Processes, Hotspot Detection, Inhomogeneous Poisson Process, Design-Based Estimation, Fuzzy Logic, Spatial Epidemiology, Monte Carlo Inclusion Probabilities, Sensitivity Analysis

1. Introduction

The detection and reliable estimation of rare disease hotspots remain among the most difficult and practically important problems in spatial epidemiology and public health surveillance. Rare diseases are typically associated with very low baseline prevalence, strong spatial sparsity, and highly heterogeneous underlying risk surfaces. Because of these features, conventional probability sampling designs often perform poorly, as a large proportion of sampled units yield zero observations while a small number of spatially localized regions account for a major part of the total disease burden. Early identification of such concentrated risk regions is crucial for timely intervention, targeted deployment of limited public health resources, and effective outbreak control. However, early stage hotspot signals are usually weak and diffuse, and they are frequently obscured by random variation in disease occurrence, imperfect diagnosis, and delays or errors in reporting. These intrinsic characteristics of rare disease processes create a fundamental challenge for surveillance systems, which must balance the need for early and sensitive detection with the requirements of statistical reliability, operational stability, and cost efficiency.

Classical probability sampling schemes such as simple random sampling, stratified sampling, and conventional cluster sampling have long been used in disease surveillance because of their strong theoretical foundations and design based inferential guarantees. However, in rare disease settings these designs are inherently inefficient, as very large sample sizes are required to achieve even moderate precision. This inefficiency has motivated the development of adaptive sampling strategies in which the sampling effort is increased in areas where cases are detected. Among these strategies, adaptive cluster sampling has received considerable attention because of its ability to concentrate sampling locally around observed cases and thereby improve estimation efficiency for rare and clustered populations. In essence, a unit is sampled once it meets a certain condition (typically the presence of at least one case), then not only this unit but also all of its neighboring units are sampled, and this procedure goes on recursively until it halts. This mechanism allows the sampling design to respond directly to the spatial structure of the disease process rather than relying solely on a fixed initial design.

Adaptive cluster sampling was designed initially to increase efficiency when sampling rare and clustered populations and has been examined extensively in both applied and theoretical work. The work of Thompson was the starting point for the development of the adaptive cluster sampling design and shows how large efficiency gains can be achieved when there is strong local clustering of the units of the population under study. Additional research resulted in further modifications to the design and its properties under various network and estimation procedures were investigated. In principle, adaptive cluster sampling operates by initiating a conventional probability sample and then expanding the sample whenever a predefined condition is satisfied, typically the presence of at least one case in a sampled unit. All neighboring units are then added to the sample and are themselves subject to the same triggering rule. This recursive expansion mechanism enables the sampling design to concentrate effort around regions of elevated disease occurrence and to exploit spatial dependence in the underlying process [1,2].

Despite its theoretical appeal, classical adaptive cluster sampling is known to suffer from serious operational and inferential limitations when applied to rare disease surveillance, particularly under weak signal conditions. Several studies have documented that the rigid threshold based triggering mecha- nism can lead to highly unstable final sample sizes and extreme design weights, especially when disease prevalence is low and spatial clustering is only moderate. In such situations, adaptive expansion often fails to activate at all, causing the design to collapse into inefficient simple random sampling, while in other cases a single triggering event can generate uncontrolled cluster growth and excessive oversampling of background regions. These instabilities have direct consequences for estimation reliability, as highly variable inclusion probabilities translate into inflated variance and poor finite sample behavior. In the context of early outbreak detection, where signals are weak and uncertainty is high, such instability severely undermines the practical usefulness of crisp adaptive cluster sampling designs [3-5].

The statistical modeling of spatial disease incidence is most naturally formulated within the frame- work of spatial point processes, where individual disease cases are represented as random points over a continuous spatial domain. Among the various point process families, the inhomogeneous Poisson point process has become established as one of the most popular models in spatial epidemiology due to its analytical tractability and the natural way in which it allows for varying levels of risk over space through a nonstationary intensity function. Within this context, localized disease clusters are at odds with the underlying null model and correspond to areas where the intensity surface maintains significantly higher values as compared to those of background risk. Spatial point process models have been used exten- sively for the analysis of infectious disease data, environmental exposure mapping, and cluster detection, and they provide a principled stochastic foundation for linking spatial risk surfaces with sampling and inference procedures [6,7].

Within the spatial epidemiology literature, a wide range of statistical procedures has been developed for disease hotspot detection, including scan statistics, kernel intensity smoothing, and measures of local spatial autocorrelation. The spatial scan statistic introduced by Kulldorff has become one of the most widely used tools for identifying localized clusters in surveillance data and has been adopted in many real time outbreak monitoring systems [8,9]. Despite their popularity, such cluster detection methods rely heavily on adequate sampling intensity and sufficiently strong contrast between hotspot and background regions. When disease prevalence is low and spatial signals are weak or diffuse, the power of these detection procedures deteriorates rapidly, leading to delayed detection and a high risk of missed clusters[10]. These limitations are further exacerbated in practice by sparse and uneven sampling, which introduces additional uncertainty into the estimated risk surface. As a consequence, cluster detection and sampling cannot be treated as separate stages in rare disease surveillance, and effective hotspot identification requires sampling designs that are intrinsically coupled with the stochastic spatial structure of disease occurrence.

Uncertainty and gradual transitions between low and high-risk regions are intrinsic features of real epidemiological processes, particularly during the early stages of outbreak formation. Classical statistical frameworks typically impose hard classification boundaries that force spatial units to be labeled either as hotspot or non-hotspot, even when the underlying risk structure is only weakly differentiated. Fuzzy set theory provides a mathematically consistent framework for representing such gradual transitions through graded membership functions rather than rigid binary labels. Since Zadeh brought it in fuzzy logic has seen wide use to model unclear, noisy, and word-based info across many science fields [11].

Despite substantial progress in adaptive sampling theory, spatial point process modeling, hotspot de- tection, and fuzzy spatial analysis, their unified integration for rare disease surveillance remains largely unexplored. Existing adaptive sampling methods continue to rely primarily on rigid threshold based triggering, while fuzzy approaches have been confined mostly to post hoc risk mapping rather than to the design of probabilistic sampling mechanisms. Moreover, very few studies have developed adaptive sampling methodologies directly under fully stochastic spatial point process environments with formal design based inferential guarantees. This paper addresses this methodological gap by developing a novel fuzzy adaptive cluster sampling framework for hotspot detection and inference under inhomogeneous spatial point processes. The proposed design replaces binary triggering with probabilistic, membership driven adaptive expansion and establishes a complete design based inferential structure through defuzzified Horvitz Thomson type estimation, Monte Carlo inclusion probability approximation, and network level variance estimation. Through rigorous theoretical development, extensive simulation experiments, and comprehensive sensitivity analysis, the study demonstrates that the proposed framework achieves superior detection sensitivity, improved estimation efficiency, and stable sampling effort under weak and diffuse hotspot conditions that typically undermine classical adaptive designs. By integrating fuzzy logic with probabilistic spatial sampling, this work provides a statistically principled and operationally credible foundation for modern rare disease surveillance under profound uncertainty.

2. Literature Review

The literature on rare disease surveillance, adaptive sampling, spatial point process modeling, and fuzzy based spatial analysis has expanded rapidly over the past two decades. This section reviews the most relevant methodological and applied contributions that form the foundation of the present study. For clarity and consistency, the reviewed studies are summarized in tabular form focusing on their primary objectives and key findings.

Adaptive and spatial methods have become an important tool in the analysis of rare events, clus- tered events and spatial heterogeneity. Adaptive cluster and adaptive sampling designs have been the subject of fundamental work that developed powerful inferential mechanisms to study rare and spatially clustered populations efficiently, providing design based theory and also, for restricted and complex situ-ations, practical solutions [4,13–16]. Complementary developments in spatial statistics and point process theory have provided the mathematical and computational foundation for modeling spatial dependence, clustering and spatio-temporal patterns in epidemiological data. Detection of disease clusters and surveillance are also enhanced with scan statistics and prospective monitoring schemes that facilitate a systematic search for spatial or temporal hotspots while practical issues in operational early warning systems have been critically reviewed in public health ap- plications BennettEtAl2015].Implementation of Fuzzy logic based methods Fuzzy set theory also allowed for a formal way to handle uncertain and partial membership in turn leading to fuzzy rule based method- ologies for spatial interpolation and modeling of disease risk with vague or missing data of further note are more recent developments that combine adaptive sampling, Bayesian updating and state of the art computational methodologies to promote more rigorous approaches to the analysis of spatial disease surveillance, hotspot detection and visualization indicative of a trend towards flexible, data adaptive and uncertainty aware geospatial modeling frameworks in public health [4-9,11-19].

Reference	Key Focus / Contribution	Findings / Conclusions
Thompson (1990)	Introduced the classical adaptive cluster sampling design for rare and clustered populations.	Demonstrated major efficiency gains for clustered rare populations but highlighted sensitivity to network structure.
Salehi and Seber (1997)	Developed inferential theory for adaptive sampling in rare populations.	Established variance inflation issues under weak clustering and unstable triggering.
Brown and Manly (1998)	Proposed restricted adaptive cluster sampling to limit oversampling.	Showed improved cost control but at the expense of detection power.
Diggle (2013)	Developed full theory of spatial and spatio temporal point processes for disease modeling.	Demonstrated the suitability of inhomogeneous Poisson processes for spatial disease risk modeling.
Baddeley et al. (2015)	Comprehensive methodological treatment of spatial point pattern analysis.	Provided practical tools for intensity estimation, clustering, and model diagnostics.

Kulldorff (1997, 2001)	Developed spatial and space time scan statistics for disease cluster detection.	Established scan statistics as a standard hotspot detection tool but requiring strong signal contrast.
Wang et al. (2012)	Applied fuzzy logic for uncertainty aware spatial interpolation.	Demonstrated improved performance over crisp methods under noisy spatial conditions.
Saha et al. (2019)	Applied fuzzy GIS for disease risk surface modeling.	Showed that fuzzy membership improves representation of diffuse risk boundaries.
Andrade Pacheco et al. (2020)	Developed adaptive spatial sampling for infectious disease hotspot detection.	Demonstrated that adaptive sampling outperforms fixed designs under strong clustering.
Bennett et al. (2015)	Reviewed early warning systems for disease surveillance.	Identified weak signal detection and sparse data as major unresolved barriers.
Zhang et al. (2021)	Developed adaptive spatial sampling for infectious disease surveillance using Bayesian intensity updating.	Showed improved hotspot detection under moderate clustering but relied on deterministic triggering.
Moraga (2022)	Developed spatial point process models for disease mapping under uncertainty.	Demonstrated improved risk estimation under sparse disease conditions using inhomogeneous Poisson models.
Lee and Lawson (2022)	Reviewed modern methods for spatial and spatio-temporal disease cluster detection.	Highlighted poor performance of classical scan statistics under weak signal regimes.
Rahman et al. (2023)	Proposed fuzzy based spatial risk assessment for epidemiological exposure mapping.	Demonstrated superior representation of diffuse disease risk boundaries under uncertainty.
Gomez-Rubio and Rue (2024)	Developed scalable Bayesian spatial models for disease surveillance using point processes.	Established robust inference under extremely sparse disease data with computational efficiency.

Table 1: Key Studies Related to Adaptive Sampling, Spatial Point Processes, and Fuzzy Based Disease Surveillance

From the literature summarized in Table 1, it is evident that substantial progress has been achieved in adaptive sampling methodology, spatial point process and Bayesian disease mapping models, disease cluster detection techniques, and fuzzy based spatial risk analysis. However, these developments have largely evolved in isolation, and their unified integration for rare disease surveillance remains limited. Classical adaptive cluster sampling continues to rely on rigid threshold based triggering mechanisms and is known to suffer from severe instability under weak and diffuse clustering. Spatial scan statistics and kernel-based detection methods exhibit rapid deterioration in performance under sparse sampling and low disease prevalence. Modern spatial point process and Bayesian disease mapping models provide pow- erful inferential tools but are typically applied after data collection rather than being embedded directly within the sampling design. Likewise, fuzzy logic has been used successfully for post hoc risk surface representation and uncertainty aware classification, but it has not been formally incorporated into the construction of probabilistic adaptive sampling schemes under stochastic spatial disease models. More- over, very few existing studies simultaneously establish fuzzy driven adaptive sampling, inhomogeneous point process modeling, and formal design based inferential guarantees within a unified framework. The present study directly addresses these unresolved gaps by developing a fuzzy adaptive cluster sampling methodology with rigorous estimation theory and comprehensive simulation-based validation for rare disease hotspot detection under realistic epidemiological conditions.

Table 2 presents a systematic comparison between dominant methodological characteristics of existing research on adaptive sampling, spatial disease modeling, and hotspot detection, and the corresponding novel features introduced by the proposed fuzzy adaptive cluster sampling framework. The comparison highlights how most existing approaches remain restricted to rigid threshold-based triggering, post hoc spatial modeling, and binary hotspot classification, whereas the proposed methodology integrates fuzzy driven adaptive expansion directly with inhomogeneous spatial point process modeling and design-based inference. This side by side assessment allows the methodological advancements of the present study to be interpreted in direct relation to well-established limitations reported in the current surveillance and sampling literature.

Methodological Aspect	Existing Research Approaches	Proposed Methodology in This Article
Adaptive triggering mechanism	Uses rigid threshold based binary triggering in classical adaptive cluster sampling.	Uses probabilistic fuzzy membership based triggering allowing smooth adaptive expansion.
Handling of weak and diffuse hotspots	Performance deteriorates rapidly under weak signal and low prevalence conditions.	Maintains stable detection sensitivity even under weak and diffuse hotspot regimes.
Integration with spatial point processes	Spatial point process models are generally applied only at the analysis stage.	Explicitly embeds inhomogeneous spatial point process modeling within the adaptive sampling design.
Treatment of uncertainty in hotspot boundaries	Hotspot boundaries are defined through sharp binary classification.	Hotspot boundaries are represented through graded fuzzy membership functions.
Design based inference	Design based inference exists but suffers from extreme weight instability.	Establishes defuzzified Horvitz Thomson type estimators with stabilized inclusion probabilities.
Variance behavior	Exhibits variance inflation and unstable final sample sizes.	Produces smooth variance behavior and predictable sampling effort.
Use of Bayesian disease mapping	Applied post hoc for smoothing and risk estimation only.	Sampling design itself is driven by stochastic risk structure prior to estimation.
Computational implementation	Often relies on fixed neighborhood expansion with heuristic restrictions.	Uses controlled Monte Carlo inclusion probability approximation with network level variance estimation.
Operational suitability for early surveillance	Limited practical reliability under early outbreak uncertainty.	Specifically designed for reliable early stage rare disease surveillance.
Sensitivity analysis	Rarely examined in a structured multi parameter framework.	Conducts comprehensive sensitivity analysis across fuzziness, intensity contrast, neighborhood, and rarity.
Unified methodological framework	Methods are fragmented across sampling, detection, and modeling.	Provides a unified fuzzy adaptive sampling, detection, and inference framework.

Table 2: Comparison of Existing Research Frameworks with the Proposed Fuzzy Adaptive Cluster Sampling Methodology

The comparative synthesis in Table 2 clearly demonstrates that the proposed framework does not represent a marginal refinement of existing adaptive sampling strategies, but instead constitutes a substantive methodological integration across fuzzy logic, spatial point process theory, and probabilistic design-based inference. By embedding uncertainty directly within the triggering mechanism, stabilizing inclusion probabilities, and coupling sampling with stochastic spatial risk structure, the proposed methodology addresses precisely those weaknesses that have consistently constrained the practical re- liability of classical approaches. This unified advancement establishes a clear conceptual and technical departure from existing research and provides a rigorous foundation for reliable rare disease hotspot detection under weak signal and high uncertainty conditions.

2.1. Preliminaries and theoretical background

This section develops the fundamental theoretical structures that are required to rigorously formulate the proposed fuzzy adaptive cluster sampling framework for rare disease hotspot detection. Since the core of the methodology is built upon a spatial stochastic representation of disease incidence, followed by an adaptive sampling mechanism driven by fuzzy decision rules, it is essential to clearly establish the probabilistic foundations of spatial point processes, the design-based logic of adaptive cluster sampling, and the mathematical structure of fuzzy sets and membership functions before proceeding to the proposed model. The primary objective of this section is therefore to ensure conceptual clarity and mathematical consistency by introducing the basic objects, assumptions, and notational conventions that will be systematically used in the subsequent methodological development. By formally unifying these three elements within a single theoretical framework, this section provides the necessary groundwork for constructing efficient estimators and for carrying out valid design-based inference under spatially heterogeneous and fuzzily defined disease risk environments.

2.2. Inhomogeneous spatial point processes for disease incidence

The spatial occurrence of rare disease cases is inherently stochastic in nature and is strongly influenced by underlying environmental, demographic, socioeconomic, and behavioral risk factors that vary continuously over geographic space. A rigorous statistical representation of such spatially heterogeneous disease incidence therefore requires a probabilistic framework capable of describing both the random location of cases and the non-uniform structure of disease risk across the study region. Spatial point process theory provides a natural and mathematically coherent foundation for this purpose, as it treats each observed disease case as a random point located within a continuous spatial domain and characterizes the overall disease pattern through its associated intensity function. In the context of rare disease surveillance, the assumption of spatial inhomogeneity is particularly important because the underlying risk is seldom constant over space and is instead shaped by localized clusters of exposure, environmental contaminants, or socio-economic vulnerabilities.

Let $D \subset \mathbb{R}^2$ denote a bounded spatial region representing the geographical area under surveillance, such as a district, a state, or any other administratively defined domain. A spatial point process X on D is defined as a random countable subset of D, where each realization of X corresponds to the observed set of disease case locations during a specified time period. The fundamental quantity that governs the probabilistic structure of a spatial point process is the intensity function $\lambda(s)$, where $s = (s_1, s_2) \in D$ denotes a generic spatial location. The intensity function is defined such that for an infinitesimally small region ds around s, the expected number of points falling in that region satisfies

$$E\{N(ds)\} = \lambda(s)\, ds,$$

where $N(\cdot)$ denotes the counting measure associated with the point process. In the present context, $\lambda(s)$ represents the local disease risk surface, with larger values indicating regions of higher expected disease incidence.

A spatial point process is said to be inhomogeneous when the intensity function varies with location, that is, when $\lambda(s)$ is not constant over D. Among inhomogeneous processes, the inhomogeneous Poisson point process plays a central role due to its mathematical tractability and its ability to represent spatial randomness driven solely by a spatially varying risk surface. Under an inhomogeneous Poisson process, the numbers of points occurring in disjoint spatial regions are independent random variables, and for any measurable subset $A \subset D$, the number of disease cases $N(A)$ follows a Poisson distribution with mean

$$E\{N(A)\} = \int_A \lambda(\mathbf{s})\, d\mathbf{s}.$$

This model is particularly appropriate for rare diseases, where individual cases occur infrequently and are often driven by localized risk factors rather than by strong interaction effects between cases.

The adaptability of the inhomogeneous Poisson process is due to the fact that the intensity function $\lambda(s)$ may be related to a set of spatial covariates, $z(s)$, via a log-linear model of the form

$$\lambda(\mathbf{s}) = \exp\left\{\beta_0 + \sum_{j=1}^{p} \beta_j z_j(\mathbf{s})\right\},$$

where $z_1(s), \ldots, z_p(s)$ denote spatially indexed covariates such as population density, environmental pollution, sanitation access, or socioeconomic indicators, and $\beta = (\beta_0, \beta_1, \ldots, \beta_p)$ is a vector of regression parameters. This formulation allows the disease risk surface to vary smoothly across space in response to observed risk factors and provides a direct probabilistic mechanism for defining spatial heterogeneity and potential hotspot regions.

From a surveillance perspective, a hotspot may be interpreted as a subregion $H \subset D$ where the local intensity $\lambda(s)$ exceeds the background level by a substantial margin. Unlike classical binary hotspot definitions based on sharp thresholds, the continuous nature of $\lambda(s)$ naturally supports a gradual representation of disease risk, which becomes especially important when fuzzy concepts are introduced in later sections. The inhomogeneous intensity function therefore serves as the mathematical bridge between the underlying continuous disease risk surface and the discrete spatial sampling units that are observed through the survey design.

When the continuous spatial domain D is partitioned into a finite collection of non overlapping areal units $U = \{1, 2, . . . ,N\}$, the integrated intensity over unit i given by

$$\mu_i = \int_{A_i} \lambda(\mathbf{s})\, d\mathbf{s},$$

represents the expected number of disease cases in that unit, where A_i denotes the spatial area of unit i. Conditional on the intensity surface, the observed count y_i within unit i may be modeled as

$$y_i \mid \lambda(\cdot) \sim \text{Poisson}(\mu_i),$$

which establishes the direct probabilistic link between the continuous point process representation and the discrete count data recorded at the level of villages, wards, or grid cells. This formulation is crucial for integrating spatial point process theory with adaptive cluster sampling, which operates on discrete spatial units rather than on continuous point locations.

Within the framework of rare disease surveillance, the inhomogeneous spatial point process thus provides a mathematically coherent description of spatially varying disease risk, enables the identification of latent hotspot structures through the intensity surface, and creates the foundational probabilistic environment within which adaptive sampling strategies can be formally studied. In the subsequent subsection, this continuous spatial stochastic representation will be connected to the design based logic of adaptive cluster sampling, which governs how sampling units are selected and expanded in response to observed disease occurrence.

2.3. Basics of adaptive cluster sampling design

Adaptive cluster sampling is a class of probability sampling designs that is specifically tailored for populations in which the characteristic of interest is rare, spatially clustered, and unevenly distributed over the study region. Classical sampling designs such as simple random sampling or systematic sampling often perform very poorly in such situations because a large proportion of sampled units contain no information about the rare population, leading to inefficient estimators with high variance. Adaptive cluster sampling overcomes this limitation by allowing the sampling design itself to respond to the ob- served data during the survey process, so that regions exhibiting the rare characteristic are sampled more intensively than regions where the characteristic is absent. This adaptive feature makes the design particularly suitable for rare disease surveillance, where cases tend to occur in localized spatial clusters rather than being uniformly distributed across space.

Formally, let the study region be partitioned into a finite collection of non overlapping primary sampling units, denoted by $U = \{1, 2, . . . ,N\}$, where each unit represents a spatial cell, village, ward, or any other appropriate areal unit. Let y_i denote the value of the variable of interest associated with unit i, which in the present context represents the number of disease cases observed in that unit during a specified time period. The defining feature of adaptive cluster sampling is the use of a pre specified condition $C(y_i)$, known as the adaptive condition, which determines whether the neighborhood of unit i should be added to the sample. A typical condition in rare disease studies is $C(y_i) = 1$ if $y_i \geq 1$ and $C(y_i) = 0$ otherwise, meaning that the presence of at least one disease case triggers adaptive expansion.

The sampling process begins with an initial sample s0 selected from U according to some conventional probability sampling design such as simple random sampling without replacement. For each unit i in s_0, the value y_i is observed and the adaptive condition is evaluated. If $C(y_i) = 1$, then all neighboring units of i, according to a pre defined neighborhood structure, are added to the sample. The same procedure is then applied recursively to each newly added unit, so that the sample grows dynamically until no further units satisfy the adaptive condition. The final sample obtained through this recursive expansion is denoted by s, and it consists of one or more disjoint clusters of spatially connected units containing at least one initial unit from s_0.

A key theoretical concept in adaptive cluster sampling is that of a network. A network is defined as the maximal set of units that are connected to each other through the neighborhood relation and that satisfy the adaptive condition either directly or indirectly. Every unit in the population belongs to exactly one network, and each network is associated with a well defined probability of being included in the final sample, which depends on the probability that at least one of its units appears in the initial sample s_0. Let Ak denote the k th network and let m_k be the number of units in that network. If the initial sample is selected by simple random sampling of size n_0, the inclusion probability of the network A_k is given by

$$\pi_k = 1 - \frac{\binom{N-m_k}{n_0}}{\binom{N}{n_0}},$$

which represents the probability that at least one of the m_k units of the network is included in the initial sample.

Once the network structure is established, estimation under adaptive cluster sampling is typically carried out using unbiased or approximately unbiased estimators that properly account for the unequal inclusion probabilities induced by the adaptive design. A commonly used estimator for the population total $T = \sum_{i=1}^{N} y_i$ is the Horvitz Thomson type estimator defined at the network level as

$$\hat{T}_{ACS} = \sum_{k \in S_N} \frac{Y_k}{\pi_k},$$

where S_N denotes the set of networks that are intersected by the final sample and $Y_k = \sum_{i \in A_k} y_i$ is the total of the variable of interest within network Ak. The performance of this estimator depends crucially on the degree of clustering of the rare population and on the choice of the adaptive condition and neighborhood structure.

In the context of rare disease surveillance, adaptive cluster sampling offers a natural mechanism for concentrating sampling effort in high risk regions without requiring a complete prior map of disease incidence. However, its classical formulation relies on crisp triggering rules and sharply defined neighborhood structures, which are often unrealistic in real public health settings where disease risk is gradual rather than binary and spatial boundaries are inherently vague. These drawbacks warrant the incorporation of fuzzy concepts in the framework of adaptive cluster sampling, which will be presented in the next sections to enable gradual hotspot member ship as well as soft adaptive expansion, in terms of levels of disease risks, not by hard threshold exceedance criteria.

2.4. Fundamentals of fuzzy sets and membership functions

The practical process of rare disease surveillance is inevitably affected by multiple layers of uncertainty that cannot be adequately represented through sharp binary classifications alone. In many real-world situations, areas are not strictly diseased or non-diseased, populations are not purely high risk or low risk, and observed case counts are often affected by misclassification, underreporting, and diagnostic ambiguity. Classical set theory, which forces every element to belong either completely to a set or not to belong at all, is therefore often too rigid to capture the gradual nature of disease risk and hotspot formation. Fuzzy set theory provides a mathematically rigorous framework for modeling such gradual transitions by allowing partial membership of elements in a set, thereby offering a flexible tool for representing spatial vagueness and uncertainty in public health surveillance.

Let X be a non empty universe of discourse representing a collection of spatial units, locations, or risk states under consideration. A fuzzy set Ã in X is defined as a collection of ordered pairs

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) : x \in X\},$$

where $\mu_{\tilde{A}} : X \to$ [0, 1] is referred to as the membership function of fuzzy set $\tilde{A}$ The value $\mu_{\tilde{A}}(x)$ can be taken to mean the degree of membership of the element x in the fuzzy set $\tilde{A}$. Zero means total non-membership, one means full membership and in-between values correspond to partial or graded membership. In the context of disease surveillance, $\underline{X}$ may represent spatial locations or areal units, while $\tilde{A}$ may represent the fuzzy set of hotspot regions, with the membership function expressing the degree of hotspot intensity at each spatial unit.

A fundamental advantage of fuzzy representation is that it allows the hotspot concept to be defined gradually rather than through a rigid threshold. For example, if $\lambda(s)$ denotes the underlying disease intensity at a spatial location s, a fuzzy hotspot membership function may be constructed as a smooth increasing transformation of λ(s), so that locations with low intensity have small membership values, locations with moderate intensity have intermediate membership values, and locations with very high intensity approach full membership in the fuzzy hotspot set. This representation faithfully reflects the epidemiological reality that disease risk increases continuously rather than abruptly across space.

Operations on fuzzy sets are generalizations of classical set operations and are defined through corresponding operations on membership functions. If $\tilde{A}$ and $\tilde{B}$ are two fuzzy sets on X with membership functions $\mu_{\tilde{A}}(x)$ and $\mu_{\tilde{B}}(x)$, respectively, then the fuzzy intersection, union, and complement are defined by

$$\mu_{\tilde{A}\cap\tilde{B}}(x) = \min\{\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)\}, \quad \mu_{\tilde{A}\cup\tilde{B}}(x) = \max\{\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)\}, \quad \mu_{\tilde{A}^c}(x) = 1 - \mu_{\tilde{A}}(x).$$

These operations make it possible to confer multiple sources of fuzzy information, such as environmental exposure, demographic susceptibility, and reported disease incidence, in a coherent manner under a single mathematical umbrella. A particularly important concept in fuzzy modeling is that of the α cut. For any $\alpha \in [0, 1]$, the α cut of a fuzzy set $\tilde{A}$ is the classical set defined by

$$A_\alpha = \{x \in X : \mu_{\tilde{A}}(x) \geq \alpha\}.$$

The α cut converts a fuzzy representation into a family of nested crisp sets indexed by the confidence level α. In the present context, α cuts will play a central role in linking fuzzy hotspot representations to adaptive sampling decisions, as they provide a mechanism for translating gradual risk information into operational sampling thresholds while retaining control over the degree of conservatism in hotspot detection.

In addition to fuzzy sets defined on spatial locations, fuzzy numbers are also useful for representing imprecise disease counts and uncertain diagnostic outcomes. A fuzzy number $\tilde{Y}$ is characterized by a membership function defined on the real line that satisfies normality, convexity, and upper semi continuity. In surveillance applications, fuzzy counts can arise naturally when reported cases include suspected and probable diagnoses with varying degrees of diagnostic certainty. Such fuzzy valued observations can be incorporated into the estimation process through appropriate defuzzification procedures, allowing imprecise observations to contribute to inference without being forced into artificially sharp categories.

The concept of defuzzification provides a systematic method for converting fuzzy quantities into crisp numerical values that can be used for estimation and decision making. Among the commonly used defuzzification rules, the centroid or center of gravity method is particularly attractive due to its intuitive interpretation and favorable mathematical properties. If $\tilde{Y}$ is a fuzzy number with membership function $\mu_{\tilde{Y}}(y)$, its centroid defuzzified value is defined as

$$Y^* = \frac{\int y\, \mu_{\tilde{Y}}(y)\, dy}{\int \mu_{\tilde{Y}}(y)\, dy},$$

provided the integrals exist. This transformation will later be employed to construct crisp estimators from fuzzy weighted sampling schemes under the proposed fuzzy adaptive cluster sampling design.

Within the present methodological framework, fuzzy sets and membership functions serve as the mathematical bridge between continuous spatial disease risk and adaptive sampling decisions that must be implemented in discrete steps. By allowing spatial units to belong to hotspot regions with varying degrees of strength, fuzzy modeling enables a soft form of adaptive expansion that is better aligned with the gradual nature of disease risk surfaces generated by inhomogeneous spatial point processes. In the next subsection, this fuzzy representation of disease risk will be integrated with spatial information to define a formal fuzzy model for disease hotspots, which will serve as the core driver of the proposed fuzzy adaptive cluster sampling design.

2.5. Fuzzy modelling of disease risk and hotspots

The transition from a continuous spatial intensity surface to an operational definition of disease hotspots requires a formal mechanism for translating heterogeneous risk into interpretable and actionable spatial structures. While the intensity function of an inhomogeneous spatial point process provides a mathemat- ically precise representation of local disease risk, it does not by itself prescribe how regions of elevated risk should be classified, compared, or targeted under a sampling design. In classical hotspot analysis, this gap is typically bridged through the use of fixed thresholds, such as declaring all locations with intensity exceeding a pre specified constant as hotspots. Such a construction, although analytically con- venient, imposes an artificial sharp boundary on a phenomenon that is inherently gradual and uncertain. Fuzzy modelling resolves this conceptual incompatibility by allowing the notion of a disease hotspot to be expressed in terms of degrees of membership rather than rigid binary inclusion, thereby offering a more realistic and epidemiologically meaningful representation of spatial risk heterogeneity.

Let $\lambda(s)$ denote the inhomogeneous intensity function of the spatial point process defined over the study region D. A fuzzy hotspot can be defined as a fuzzy set $\tilde{H}$ on D with a membership function $\mu_{\tilde{H}}$ (s) that maps each spatial location $s \in D$ to a value in the unit interval [0, 1]. This membership function is constructed as a monotone increasing transformation of the underlying intensity surface, so that higher disease risk corresponds to stronger membership in the fuzzy hotspot set. A general and flexible representation is given by

$$\mu_{\tilde{H}}(\mathbf{s}) = g(\lambda(\mathbf{s})),$$

where $g : \mathbb{R}^+ \to [0, 1]$ is a smooth, non decreasing scaling function satisfying $\lim_{\lambda\to 0} g(\lambda) = 0$ and $\lim_{\lambda\to\infty} g(\lambda) = 1$. Typical choices of $g(\cdot)$ include logistic, Gompertz, or piecewise linear functions, which allow the analyst to control the rate at which hotspot membership increases with disease intensity. Through this construction, locations with very low intensity exhibit negligible membership in the hotspot set, locations with moderate intensity exhibit partial membership, and locations with very high intensity approach full membership.

When the continuous study region is discretized into areal units $A_1, A_2, \ldots, A_N$, the fuzzy hotspot membership at the unit level may be defined through spatial aggregation of the point level membership function. A natural definition is

$$\mu_{\tilde{H}}(i) = \frac{1}{|A_i|} \int_{A_i} \mu_{\tilde{H}}(\mathbf{s})\, d\mathbf{s},$$

where $|A_i|$ denotes the area of unit i. This aggregated membership value measures the average degree of hotspot intensity within unit i and serves as the fundamental fuzzy risk indicator driving the adaptive sampling design. Units with larger values of $\mu_{\tilde{H}}$ (i) are interpreted as exhibiting stronger hotspot behavior, while units with smaller values are regarded as weakly affected or background regions.

Fuzzy modelling also provides a principled framework for integrating multiple sources of risk information into a single composite hotspot indicator. Let $z_1(s)$, $z_2(s)$, . . . , $z_p(s)$ denote spatial covariates related to environmental exposure, population vulnerability, health infrastructure, and other determinants of disease risk. For each covariate, a corresponding fuzzy risk set $\tilde{R}_j$ with membership function $\mu_{\tilde{R}_j}$ (s) can be constructed. The overall fuzzy hotspot membership may then be defined through fuzzy aggregation operators such as weighted intersections or unions, for example

$$\mu_{\tilde{H}}(\mathbf{s}) = \max_{1\le j\le p} \mu_{\tilde{R}_j}(\mathbf{s}), \qquad \text{or} \qquad \mu_{\tilde{H}}(\mathbf{s}) = \sum_{j=1}^{p} w_j \mu_{\tilde{R}_j}(\mathbf{s}),$$

where $w_j \ge 0$ and $\sum_{j=1}^{p} w_j = 1$. These constructions allow heterogeneous risk determinants to be fused into a single spatially varying fuzzy hotspot surface without forcing premature dichotomization of any individual risk factor.

The operational role of fuzzy hotspot modelling becomes explicit through the use of α cuts. For any fixed $\alpha \in [0, 1]$, the α cut of the fuzzy hotspot set is defined as

$$H_\alpha = \{i \in U : \mu_{\tilde{H}}(i) \geq \alpha\},$$

which yields a crisp subset of spatial units that are classified as hotspots at confidence level α. As α varies from low to high values, the corresponding α cuts generate a nested sequence of hotspot regions, ranging from diffuse high risk belts to compact core hotspots. This graded classification is especially valuable in public health surveillance, where decision makers often need to balance the risk of missing true outbreak areas against the cost of surveying excessively large regions.

Within the proposed methodological framework, the fuzzy hotspot membership $\mu_{\tilde{H}}(i)$ will serve as the primary triggering variable for the adaptive expansion rule in the sampling design. Rather than relying on a crisp condition such as the presence or absence of at least one observed case, the fuzzy adaptive mechanism will allow the probability and extent of cluster expansion to vary smoothly with the degree of hotspot membership. This integration of fuzzy risk modelling with spatial point process intensity thus provides the conceptual and mathematical foundation for a soft, data responsive, and epidemiologically realistic adaptive cluster sampling design, which will be formally introduced in the next section.

2.6. Fuzzy adaptive cluster sampling design

The purpose of this section is to formally establish the proposed fuzzy adaptive cluster sampling frame work for efficient detection of rare disease hotspots under spatially heterogeneous risk environments. Building on the theoretical foundations of inhomogeneous spatial point processes, classical adaptive cluster sampling, and fuzzy set theory developed in the previous section, we now integrate these three components into a unified sampling design that is capable of responding gradually to varying degrees of disease risk across space. Unlike conventional adaptive cluster sampling, which relies on sharp triggering rules and binary expansion decisions, the fuzzy adaptive design introduced here allows the intensity of adaptive expansion to depend smoothly on the degree of hotspot membership associated with each spa- tial unit. This formulation not only reflects the epidemiological reality of gradual risk variation but also provides greater flexibility and stability in rare disease surveillance, especially in settings characterized by weak signals, diagnostic uncertainty, and diffuse hotspot structures.

2.7. Study region and spatial sampling framework

Let the study region be a bounded two dimensional spatial domain $D \subset \mathbb{R}^2$ representing the geographical area under surveillance. For operational implementation of the sampling design, the continuous region D is partitioned into a finite collection of N non overlapping and collectively exhaustive spatial areal units denoted by $A_1, A_2, \ldots, A_N$. These units may correspond to villages, wards, census tracts, health reporting zones, or regular grid cells, depending on the scale and administrative structure of the surveillance system. Let $U = \{1, 2, \ldots, N\}$ denote the index set of all spatial units in the population. Associated with each unit $i \in U$ is an observed disease count y_i, representing the number of reported rare disease cases during a fixed surveillance period, and a fuzzy hotspot membership value $\mu_{\tilde{H}}(i)$ derived from the underlying spatial intensity surface and auxiliary risk information.

The neighborhood structure plays a central role in defining the geometry of cluster formation under adaptive sampling. For each unit i, let $\mathcal{N}(i)$ denote the set of neighboring units according to a pre specified adjacency rule. Common neighborhood definitions include first order contiguity based on shared boundaries, distance based neighborhoods defined through a fixed radius, or graph based connections derived from transportation or social contact networks. The choice of neighborhood structure determines the spatial connectivity of adaptive clusters and directly influences the extent to which detected disease signals propagate through the sampling process.

The sampling process begins with the selection of an initial sample s_0 of size n_0 from the finite population U using a conventional probability sampling design such as simple random sampling without replacement. The initial sample provides the entry points through which the adaptive expansion is triggered. For each unit $i \in s_0$, the observed disease count y_i and the corresponding fuzzy hotspot membership $\mu_{\tilde{H}}(i)$ are evaluated. These quantities jointly determine whether and how the sampling design expands into neighboring units under the fuzzy adaptive rule, which will be formally introduced in the next subsection.

At any stage of the sampling process, let s denote the current sample consisting of all units that have been selected either through the initial sampling stage or through subsequent adaptive expansion. The final sample is obtained when no further units qualify for inclusion under the fuzzy adaptive triggering mechanism. The resulting design induces a collection of spatially connected clusters whose sizes, shapes, and inclusion probabilities depend not only on the initial sample s_0 but also on the continuous distribution of fuzzy hotspot membership values across space. This structure forms the fundamental geometric and probabilistic backbone of the proposed fuzzy adaptive cluster sampling design.

Through this spatial sampling framework, the design establishes a formal bridge between the continuous disease risk surface modeled by the inhomogeneous spatial point process and the discrete sampling units through which surveillance data are collected. In the next subsection, this framework will be augmented with a fuzzy triggering rule that governs the adaptive expansion of the sample in response to graded hotspot membership rather than binary disease indicators.

2.8. Fuzzy triggering rule for adaptive expansion

The defining feature that distinguishes the proposed fuzzy adaptive cluster sampling design from its classical counterpart is the manner in which the decision to expand the sample is governed by graded hotspot information rather than by a binary indicator of disease presence. In conventional adaptive cluster sampling, expansion is triggered whenever an observed unit satisfies a crisp condition such as the occurrence of at least one case. While such a rule is mathematically simple, it is often too rigid for rare disease surveillance where early stage outbreaks, borderline clusters, and diagnostic uncertainty are common. The fuzzy adaptive framework replaces this rigid triggering mechanism with a continuous expansion rule driven by fuzzy hotspot membership values, thereby allowing the strength of adaptive expansion to reflect the intensity of local disease risk in a gradual and stable manner.

Let $\mu_{\tilde{H}}(i)$ denote the fuzzy hotspot membership associated with unit i, as defined in the previous section. Instead of imposing a single fixed threshold for expansion, the fuzzy design introduces a triggering function

$$\phi(i) = \psi\big(\mu_{\tilde{H}}(i)\big),$$

where $\psi : [0, 1] \to [0, 1]$ is a non decreasing expansion response function. The value $\phi(i)$ represents the strength or probability with which unit i triggers adaptive expansion into its neighboring units. A simple and interpretable choice is $\phi(i) = \mu_{\tilde{H}}(i)$, under which units with weak hotspot membership trigger expansion weakly, while units with strong membership trigger expansion with high intensity. This formulation ensures that adaptive growth of the sample responds smoothly to the spatial risk gradient rather than reacting abruptly to marginal fluctuations in observed case counts.

Operationally, the fuzzy adaptive expansion proceeds as follows. For each sampled unit i, a random Bernoulli variable Bi is generated with success probability $\phi(i)$. If $B_i = 1$, then all neighboring units in $\mathcal{N}(i)$ are added to the current sample, subject to the condition that they have not already been visited. If $B_i = 0$, no adaptive expansion is initiated from unit i. This probabilistic expansion mechanism ensures that units with higher fuzzy hotspot membership contribute more strongly to sample growth, while still preserving randomization and design based inferential validity.

An alternative deterministic formulation of the fuzzy triggering rule is obtained through the use of an α level expansion criterion. In this case, adaptive expansion is initiated from unit i if and only if

$$\mu_{\tilde{H}}(i) \geq \alpha,$$

where $\alpha \in (0, 1]$ denotes a design controlled sensitivity parameter. Smaller values of α produce aggressive expansion with large adaptive clusters, while larger values yield conservative expansion restricted to only the strongest hotspot cores. By varying α, the sampling authority can explicitly control the tradeoff between early detection sensitivity and field survey cost within a unified fuzzy design framework.

The recursive structure of the fuzzy adaptive expansion is defined identically to that of classical adaptive sampling, with the crucial difference that the triggering condition is now fuzzy rather than binary. If unit i triggers expansion, then every unit $j \in \mathcal{N}(i)$ is added to the sample and its own fuzzy membership $\mu_{\tilde{H}}(j)$ is subsequently evaluated to determine whether further propagation occurs. The recursion terminates when no newly added units satisfy the fuzzy expansion condition. The final sample therefore consists of one or more spatially connected fuzzy adaptive clusters whose sizes and shapes depend jointly on the initial sample and the continuous spatial distribution of fuzzy hotspot membership.

From a design theoretic perspective, this fuzzy triggering mechanism induces unequal and risk adaptive inclusion probabilities across spatial units. Units located in high membership hotspot cores exhibit substantially higher probabilities of being included in the final sample than background units, while intermediate risk regions exhibit smoothly varying inclusion behavior. This graded inclusion structure forms the key mechanism through which the proposed fuzzy adaptive design achieves greater efficiency in detecting rare disease hotspots without wasting excessive sampling effort in low risk regions.

This fuzzy expansion rule thus serves as the operational heart of the proposed sampling framework, linking the theoretical fuzzy risk representation directly to the real time evolution of the sample. In the next subsection, this adaptive mechanism will be embedded within the formal neighborhood structure to define the precise geometry of fuzzy adaptive cluster formation.

2.8. Neighbourhood structure and cluster formation

The geometric configuration of adaptive clusters under the proposed fuzzy sampling framework is fundamentally governed by the specification of the neighborhood structure among spatial units. While the fuzzy triggering rule determines whether adaptive expansion should occur from a given unit, the neighborhood structure determines the precise spatial pathways along which this expansion propagates. In practical surveillance settings, disease transmission, environmental exposure, and social interaction are all mediated through spatial proximity and connectivity, and it is therefore essential that the sampling design incorporates a neighborhood system that faithfully represents the underlying spatial interaction mechanisms. The choice of neighborhood structure thus plays a central role in shaping the size, geometry, and epidemiological relevance of the resulting fuzzy adaptive clusters.

Let $U = \{1, 2, \ldots, N\}$ denote the finite population of spatial units and let $\mathcal{N}(i) \subset U$ denote the neighborhood set associated with unit i. The mapping $i \mapsto \mathcal{N}(i)$ defines an undirected spatial adjacency graph $\mathcal{G} = (U, E)$, where an edge $(i, j) \in E$ exists whenever $j \in \mathcal{N}(i)$ and $i \in \mathcal{N}(j)$. In areal disease surveillance, a common choice is first order contiguity, where two units are regarded as neighbors if they share a common boundary. Alternatively, neighborhoods may be defined through distance based criteria, such as

$$\mathcal{N}(i) = \{j \in U : d(i,j) \leq r\},$$

for some fixed radius r, where $d(i, j)$ denotes the Euclidean distance between the centroids of units i and j. More general neighborhood systems may also be constructed from transportation networks, river systems, or human mobility flows, depending on the dominant transmission pathways of the disease under study.

Once the neighborhood graph is specified, cluster formation under fuzzy adaptive expansion proceeds through a recursive graph traversal mechanism. Let s_0 denote the initial sample selected through the base probability design and let s denote the evolving sample. For each unit $i \in s$, the fuzzy triggering rule defined in the previous subsection is evaluated, and if unit i activates expansion, then every unit $j \in \mathcal{N}(i)$ is added to the sample provided it has not already been included. Each newly added unit then inherits the same evaluation protocol, so that cluster growth proceeds outward through the neighborhood graph in a recursive fashion until no further units satisfy the fuzzy expansion condition. The final sample therefore decomposes into a collection of disjoint fuzzy adaptive clusters, each consisting of spatial units connected through neighborhood adjacency and linked through a chain of fuzzy activation events.

Formally, let $C(i)$ denote the final adaptive cluster generated from an initiating unit $i \in s_0$. Then

$$C(i) = \left\{ j \in U : \begin{array}{l} \text{there exists a sequence } i = j_0, j_1, \ldots, j_m = j \\ \text{such that } j_{k+1} \in \mathcal{N}(j_k) \\ \text{and } j_k \text{ activates fuzzy expansion} \end{array} \right\},$$

which defines $C(i)$ as the connected component of the neighborhood graph reachable from i through fuzzy activation. The collection of all such clusters over s_0 forms a partition of the final sample s, and no spatial unit belongs to more than one fuzzy adaptive cluster.

An important distinction between classical and fuzzy adaptive clustering arises in the variability of cluster boundaries. Under a crisp triggering rule, cluster formation produces rigid spatial components whose boundaries are determined by hard threshold exceedance. Under the fuzzy design, cluster boundaries become inherently probabilistic and graded, since the likelihood of propagation decays smoothly with decreasing fuzzy hotspot membership. As a consequence, clusters formed under strong hotspot cores tend to be compact and fully connected, while clusters formed in diffuse risk regions exhibit irregular shapes and thinner connectivity. This behavior is particularly advantageous in rare disease surveillance, where true outbreak areas may not conform to regular geometric shapes and where weak signals must be allowed to propagate cautiously without inducing uncontrolled sample growth.

From an inferential perspective, the neighborhood induced fuzzy clusters define the natural resolution at which design based estimation must be conducted. As in classical adaptive cluster sampling, estimation will be performed at the cluster level rather than at the individual unit level, since all units within a fuzzy adaptive cluster share a common inclusion mechanism. However, unlike the classical case where cluster membership is deterministic conditional on the initial sample, fuzzy clusters introduce controlled randomness into the cluster boundaries through the probabilistic triggering rule. This feature leads to smoother inclusion probability gradients across space and plays a central role in stabilizing variance under weak hotspot conditions.

Thus, the neighborhood structure and the resulting fuzzy cluster formation mechanism together define the spatial skeleton of the proposed sampling design. They determine how local hotspot information is translated into extended spatial exploration and how sampling effort propagates through the disease risk surface. In the next subsection, this cluster structure will be complemented by a formal interpretation of the fuzzy adaptive design from an operational and field implementation perspective.

2.9. Estimation under fuzzy adaptive cluster sampling

The primary objective of any adaptive sampling design is not only to detect rare and spatially clustered events efficiently, but also to support statistically valid estimation of key population level characteristics under the induced unequal and data dependent inclusion structure. In the present fuzzy adaptive cluster sampling framework, the estimation problem becomes substantially more intricate than in conventional fixed designs because unit inclusion probabilities depend jointly on the initial randomization, the stochastic fuzzy triggering mechanism, and the spatial propagation of clusters through the neighborhood graph. Moreover, the presence of fuzzy membership values introduces an additional layer of uncertainty that must be carefully reconciled with design based inference principles. This section therefore develops a coherent estimation framework for rare disease surveillance under the proposed fuzzy adaptive design by formally defining the target parameters of interest, deriving the corresponding inclusion probabilities, constructing appropriate fuzzy adjusted design weights, and finally establishing operational estimators through suitable defuzzification procedures. The objective is to ensure that the resulting estimators remain interpretable, approximately unbiased, and practically implementable while fully respecting the stochastic and spatial structure induced by fuzzy adaptation.

2.10. Target Parameters for rare disease surveillance

A precise formulation of the target parameters is a necessary prerequisite for the development of statistically meaningful estimators under any complex sampling design. In the context of rare disease surveillance, the inferential goals typically extend beyond the simple estimation of aggregate disease burden and include the characterization of spatial risk intensity, the identification of hotspot magnitude, and the quantification of population level exposure heterogeneity. Under the proposed fuzzy adaptive cluster sampling framework, these targets must be defined in a manner that is compatible with both the discrete nature of the sampled spatial units and the continuous representation of disease risk induced by the underlying inhomogeneous spatial point process.

Let the study region be partitioned into N spatial units A_1, $A_2, \ldots, A_N$, and let y_i denote the observed number of disease cases in unit i during the surveillance period. The most fundamental target parameter is the population total

$$T = \sum_{i=1}^{N} y_i,$$

which accounts for the sum of rare disease cases in the whole study area. This parameter is of first order importance for public health, as it provides a direct measure of the absolute burden of disease, and is used as a basis for distribution of resources, planning for intervention and analysis of temporal trends.

Closely related to the total is the population mean disease intensity per unit, defined as

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} y_i,$$

which provides a normalized measure of disease burden that facilitates comparison across regions of different sizes or across time periods with varying numbers of spatial units. In surveillance systems where units differ substantially in population size or area, weighted versions of the mean based on unit population or spatial support may also serve as relevant target parameters.

Beyond global summaries, hotspot oriented surveillance requires the estimation of disease burden within fuzzy hotspot regions. Let $\mu_{\tilde{H}}$ (i) denote the fuzzy hotspot membership of unit i, and let H_α be the corresponding α cut hotspot region. The total disease burden within this fuzzy defined hotspot core is given by

$$T_{H_\alpha} = \sum_{i \in H_\alpha} y_i,$$

which quantifies the magnitude of disease concentration within regions of elevated risk. This quantity plays a crucial role in outbreak assessment, targeted intervention design, and the evaluation of spatial inequality in disease distribution.

Within the continuous spatial domain framework, an additional target of interest is the integrated intensity over a subregion $B \subset D$, defined as

$$\Lambda(B) = \int_B \lambda(\mathbf{s})\, d\mathbf{s},$$

which represents the expected disease burden over any epidemiologically meaningful spatial subregion. When B coincides with a union of spatial units, $\Lambda(B)$ may be approximated through discrete aggregation of the corresponding μ_i values, thereby linking the point process representation directly to the unit level surveillance data.

In many applications, the relative concentration of disease within hotspot regions is of equal or greater importance than absolute counts. This motivates the definition of prevalence ratios or concentration indices such as

$$R_\alpha = \frac{T_{H_\alpha}}{T},$$

which measures the proportion of total disease burden captured within the α level fuzzy hotspot region. Large values of R_α indicate strong spatial concentration and provide quantitative evidence of localized outbreaks.

These target parameters together define a multi resolution inferential objective for rare disease surveillance under the proposed fuzzy adaptive design. They encompass global disease burden, normalized intensity, hotspot focused burden, and relative concentration measures, thereby reflecting the diverse decision making needs of public health authorities. In the subsequent subsection, the manner in which the fuzzy adaptive sampling design induces unequal inclusion probabilities for these targets will be formally derived and analyzed.

2.11. Inclusion probabilities under fuzzy adaptive design

In any probability sampling framework, the concept of inclusion probability forms the cornerstone of valid design-based inference, since it quantifies the mechanism through which each population unit becomes observable under the survey design. Under classical fixed probability designs, inclusion probabilities are fully determined at the outset of sampling and remain unaffected by the observed data. In contrast, adaptive sampling designs introduce a data driven component into the selection mechanism, so that the probability of including a unit depends not only on the initial randomization but also on the pattern of observations revealed during the survey process. The fuzzy adaptive cluster sampling design proposed in this study extends this principle further by allowing inclusion probabilities to depend on graded hotspot membership values rather than on a binary triggering condition. As a result, the induced inclusion structure is inherently stochastic, spatially dependent, and smoothly varying across the study region.

Let I_i denote the inclusion indicator for unit i, where $I_i = 1$ if unit i is included in the final fuzzy adaptive sample and $I_i = 0$ otherwise. The first order inclusion probability of unit i is defined as

$$\pi_i = P(I_i = 1),$$

where the probability is taken over all sources of randomness in the sampling process, including the selection of the initial sample and the random outcomes of the fuzzy triggering mechanism. Under the present design, π_i admits a natural decomposition into two components corresponding to direct selection in the initial sample and indirect selection through adaptive expansion. Specifically,

$$\pi_i = \pi_i^{(0)} + \pi_i^{(A)} - \pi_i^{(0)}\pi_i^{(A)},$$

where $\pi_i^{(0)}$ denotes the probability that unit i is included in the initial sample s_0, and $\pi_i^{(A)}$ denotes the conditional probability that unit i is reached through fuzzy adaptive expansion given that it was not selected initially. This decomposition emphasizes that inclusion may occur either directly at the first stage or indirectly through spatial propagation from neighboring activating units.

When the initial sample is selected by simple random sampling of size n_0 from the population of size N, the initial inclusion probability is given by

$$\pi_i^{(0)} = \frac{n_0}{N}.$$

The adaptive component $\pi_i^{(A)}$ is substantially more complex because it depends on the spatial configuration of fuzzy hotspot membership values and the probabilistic activation of neighboring units. Let $\mathcal{P}(i)$ denote the set of all possible activation paths through the neighborhood graph that could lead to the inclusion of unit i starting from some initially selected unit. Each such path consists of a finite sequence of neighboring units whose fuzzy triggering events must occur successively. The adaptive inclusion probability may then be expressed formally as

$$\pi_i^{(A)} = 1 - \prod_{p \in \mathcal{P}(i)} \left(1 - \prod_{j \in p} \phi(j) \right),$$

where $\phi(j)$ denotes the fuzzy expansion probability associated with unit j. This expression reflects the fact that unit i is included adaptively if and only if at least one activation path reaching i is fully realized through successive fuzzy triggering events. Although this representation is generally not available in closed analytical form for large populations, it provides a clear conceptual description of how graded hotspot membership influences inclusion behavior under the fuzzy adaptive design.

From an intuitive perspective, units located in strong hotspot cores exhibit large values of $\mu_{\tilde{H}}(i)$ and hence large values of the expansion response $\phi(i)$, which substantially increases their probability of being included either directly or through multiple neighboring activation paths. Units located in diffuse or background regions, by contrast, exhibit small expansion probabilities and are therefore included primarily through the initial randomization mechanism. This smooth spatial gradient in inclusion probabilities stands in sharp contrast to the abrupt inclusion structure induced by classical adaptive sampling with a crisp triggering rule and is one of the key reasons why the fuzzy design achieves improved stability under weak hotspot conditions.

For design based estimation, it is also necessary to consider second order inclusion probabilities. Let $\pi_{ij} = P(I_i = 1, I_j = 1)$ denote the joint inclusion probability for units i and j. Under the fuzzy adaptive design, π_{ij} depends not only on their respective marginal inclusion mechanisms but also on the extent to which they share common activation paths through the neighborhood graph. Units belonging to the same fuzzy adaptive cluster exhibit strongly dependent inclusion behavior, while units separated by large spatial distances are approximately independent conditional on the initial sample. Although exact analytical expressions for π_{ij} are generally intractable, their structure can be accurately approximated through repeated Monte Carlo simulation under the known fuzzy expansion rule and neighborhood configuration.

The explicit recognition of spatially varying and data dependent inclusion probabilities is essential for the construction of valid estimators under fuzzy adaptive cluster sampling. These probabilities form the fundamental building blocks from which design weights are derived and through which unbiasedness and variance properties of estimators are assessed.

2.12. Construction of fuzzy design weights

The unequal and spatially graded inclusion probabilities induced by the fuzzy adaptive cluster sampling design must be explicitly accounted for in the estimation process through the construction of appropriate design weights. In classical survey sampling, design weights are typically defined as the reciprocals of first order inclusion probabilities and serve as the fundamental mechanism through which unequal probability sampling is corrected to recover population level quantities. Under the present fuzzy adaptive framework, the same guiding principle applies, but the stochastic, path dependent, and spatially correlated nature of inclusion necessitates a careful reinterpretation of the weight construction process so that it remains coherent, stable, and practically computable.

Let π_i denote the first order inclusion probability of unit i under the fuzzy adaptive design, as defined in the previous subsection. The basic fuzzy design weight associated with unit i is defined as

$$w_i = \frac{1}{\pi_i},$$

which represents the number of population units that unit i is expected to represent in the overall population. Units with high fuzzy hotspot membership tend to exhibit large inclusion probabilities due to repeated opportunities for adaptive propagation through neighboring activation paths and thus receive relatively smaller weights. On the other hand, background units that are added primarily via the initial randomization mechanism are assigned larger design weights. This inverse relation guarantees that the adaptive oversampling of high risk regions will not bias the estimation on the population level.

In practice, the exact analytical values of π_i are rarely available in closed form under fuzzy adaptive designs, since they depend on the joint behavior of the initial sampling stage and the recursive fuzzy expansion process across the entire neighborhood graph. Consequently, inclusion probabilities are typically approximated through design based Monte Carlo procedures. Specifically, the fuzzy adaptive sampling design is repeatedly simulated under the known expansion rule and neighborhood configuration, and the empirical inclusion proportion of each unit across a large number of repetitions is used as a consistent estimator of π_i. Let $\hat{\pi}_i$ denote such a Monte Carlo estimate based on B independent replications of the design. Then the operational fuzzy design weight is given by

$$\hat{w}_i = \frac{1}{\hat{\pi}_i}.$$

As B increases, $\hat{w}_i$ converges in probability to the true design weight w_i, thereby ensuring asymptotic design consistency of the resulting estimators.

An important feature of the fuzzy adaptive framework is that the design weights are generally smoother and less extreme than those encountered under crisp adaptive cluster sampling. Under a crisp rule, units lying just inside the triggering threshold can exhibit very small inclusion probabilities while units inside fully activated clusters exhibit near unit inclusion probabilities, leading to highly variable weights and unstable estimators. Under the fuzzy design, the continuous variation in expansion probabilities produces a gradual gradient in π_i values across space and therefore a more balanced distribution of weights. This smoothing effect is one of the primary mechanisms through which the fuzzy adaptive design stabilizes estimator variance under weak or diffuse hotspot conditions.

When estimation is performed at the cluster level rather than at the unit level, the weight construction must be expressed in terms of fuzzy adaptive clusters. Let C_k denote the kth fuzzy adaptive cluster and let π_k denote the probability that cluster C_k is intersected by the initial sample and subsequently activated. The corresponding cluster level design weight is defined as $W_k = 1/\pi_k$, and all units within C_k inherit this common weight. This cluster level weighting preserves the coherence of the adaptive inclusion mechanism and ensures that network-based estimators remain design unbiased or approximately unbiased under the fuzzy expansion rule.

From a practical standpoint, the computation of fuzzy design weights through Monte Carlo approximation is naturally integrated with the simulation-based evaluation of the sampling design that will be undertaken in later sections of the study. The same simulation engine that is used to evaluate detection efficiency and bias properties of estimators can be exploited to generate reliable numerical approximations to the inclusion probabilities and the associated weights. This dual use of simulation not only enhances computational efficiency but also embeds the estimation procedure firmly within the empirical performance assessment of the fuzzy adaptive framework.

The construction of fuzzy design weights thus provides the essential quantitative bridge between the complex inclusion structure induced by the sampling design and the classical form of weighted estimators used in survey sampling. In the next subsection, these fuzzy adjusted design weights will be combined with observed sample data through suitable defuzzification procedures to yield explicit estimators for the target parameters of rare disease surveillance.

2.13. Defuzzification of weights and final estimators

The construction of fuzzy design weights establishes a probabilistic correction for the unequal and spa- tially adaptive inclusion mechanism induced by the sampling design. However, because both the trigger- ing process and, in certain applications, even the observed disease counts may involve fuzzy quantities, the estimation procedure requires a principled mechanism to convert this fuzzy weighted structure into operational numerical estimators. This transformation is achieved through defuzzification, which serves as the formal bridge between fuzzy representations of uncertainty and crisp population level inference. The objective of this subsection is therefore to establish how fuzzy design weights and, when applicable, fuzzy valued observations are coherently transformed into final estimators for the target surveillance parameters.

Let $\tilde{w}_i$ denote the fuzzy design weight associated with unit i, which may arise either directly from a fuzzy approximation of the inclusion probability or indirectly through uncertainty in the estimation of π_i. Similarly, let $\tilde{y}_i$ denote a fuzzy representation of the observed disease count in unit i, which may occur when reported cases include suspected, probable, and confirmed diagnoses with varying degrees of diagnostic certainty. The fundamental estimating equation for the population total in fuzzy form may then be expressed as

$$\tilde{T} = \sum_{i \in s} \tilde{w}_i \tilde{y}_i,$$

where s denotes the final fuzzy adaptive sample. Since direct inference cannot be conducted on fuzzy numbers alone, this aggregated fuzzy quantity must be transformed into a crisp estimator through defuzzification.

Among the various defuzzification methods available in the literature, the centroid or center of gravity method is particularly well suited to the present context because it preserves linearity and admits a natural interpretation as an expected value under the membership distribution. If $\tilde{Z}$ denotes a generic fuzzy number with membership function μ $\mu_{\tilde{Z}}(z)$, its centroid defuzzified value is defined as

$$Z^* = \frac{\int z\, \mu_{\tilde{Z}}(z)\, dz}{\int \mu_{\tilde{Z}}(z)\, dz},$$

provided the integrals exist. Applying this transformation to the fuzzy weighted total $\tilde{T}$ yields the final operational estimator

$$\hat{T} = \left(\sum_{i \in s} \tilde{w}_i \tilde{y}_i \right)^*,$$

which represents the crisp estimate of the total disease burden obtained under the fuzzy adaptive sampling design.

When only the design weights are fuzzy while the observed disease counts are crisp, the defuzzified estimator simplifies to

$$\hat{T} = \sum_{i \in s} w_i^*\, y_i,$$

where w_i^* denotes the centroid defuzzified value of the fuzzy design weight $\tilde{w}_i$. This form preserves the classical structure of inverse probability weighted estimators while incorporating the uncertainty embedded in the fuzzy weighting mechanism. Similarly, the estimator of the population mean disease intensity is given by

$$\hat{\bar{Y}} = \frac{1}{N} \sum_{i \in s} w_i^*\, y_i.$$

For fuzzy hotspot specific targets such as the total disease burden within the α cut hotspot region, the corresponding estimator takes the form

$$\hat{T}_{H_\alpha} = \sum_{i \in s \cap H_\alpha} w_i^*\, y_i,$$

which aggregates weighted contributions only from those sampled units that belong to the crisp hotspot region defined at level α. In this manner, the fuzzy representation of risk influences estimation indirectly through its effect on inclusion probabilities and weights, while the final estimators themselves remain interpretable as conventional real valued quantities.

An important theoretical feature of the defuzzified estimators is that they inherit approximate design unbiasedness from the underlying fuzzy weighted estimators under mild regularity conditions on the form of the membership functions and the defuzzification rule. Specifically, if the fuzzy weights are centered around the true inverse inclusion probabilities and the centroid transformation is applied, then the expectation of $\hat{T}$ under the joint sampling and fuzzy randomness is approximately equal to the true population total T. Deviations from exact unbiasedness arise primarily from higher order nonlinearities in the fuzzy transformations and diminish rapidly as the fuzziness in the weight structure decreases.

From a practical point of view, the defuzzification step is important in making sure that fuzzy adaptive sampling can be used in the same interfaces of statistical reports and public health decisions as those employed in other survey methods. This results enables the analyst to take advantage of the modeling flexibility and robustness of fuzzy logic in the design and extension of sampling, while providing crisp numerical estimates that lead directly to interpretation, can be compared across regions, and finally used for policy making. In the next subsection, the variability properties of these defuzzified estimators will be formally examined through the development of variance estimation procedures adapted to the fuzzy adaptive design.

2.14. Variance estimation under fuzzy adaptive sampling

The assessment of estimator variability is an essential component of any statistically sound surveillance methodology, since point estimates alone cannot convey the degree of uncertainty associated with inferred disease burden or hotspot intensity. Under fuzzy adaptive cluster sampling, variance estimation becomes particularly delicate because the inclusion mechanism is not only unequal and spatially dependent, but also partially driven by fuzzy and probabilistic triggering rules. As a consequence, conventional variance formulas derived for fixed probability designs are no longer directly applicable, and appropriate modi- fications are required to account for both the adaptive cluster structure and the additional uncertainty induced by fuzziness.

Let $\hat{T} = \sum_{i \in s} w_i^* y_i$ denote the defuzzified estimator of the population total, where w_i^* represents the centroid defuzzified design weight for unit *i*. Under a general unequal probability sampling framework, the exact design based variance of $\hat{T}$ may be expressed through the well known Horvitz Thomson variance identity

$$\mathrm{Var}(\hat{T}) = \sum_{i=1}^{N} \sum_{j=1}^{N} (\pi_{ij} - \pi_i \pi_j) \frac{y_i}{\pi_i} \frac{y_j}{\pi_j},$$

where π_i and π_{ij} denote the first and second order inclusion probabilities, respectively. In the present fuzzy adaptive setting, this expression remains formally valid when interpreted with respect to the joint probability law induced by the initial randomization and the fuzzy triggering mechanism. However, as discussed earlier, the exact analytical forms of π_i and π_{ij} are generally unavailable due to the recursive and path dependent nature of fuzzy cluster formation.

To obtain a practical variance estimator, one therefore resorts to a Monte Carlo based approximation strategy that is fully consistent with the simulation driven character of the proposed methodology. Let B independent replications of the fuzzy adaptive sampling design be generated under the known neighborhood structure and fuzzy expansion rule. For the b th replication, let $\hat{T}^{(b)}$ denote the corresponding defuzzified estimator of the population total. The design based variance of $\hat{T}$ is then naturally estimated by the empirical varianc

$$\widehat{\mathrm{Var}}(\hat{T}) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{T}^{(b)} - \bar{T}\right)^2, \qquad \bar{T} = \frac{1}{B}\sum_{b=1}^{B}\hat{T}^{(b)}.$$

This estimator converges in probability to the true design variance as $B \to \infty$ and automatically captures the combined effects of unequal probability selection, spatial dependence, and fuzzy triggering.

When estimation is performed at the level of fuzzy adaptive clusters rather than at the level of individual units, a cluster based variance representation becomes more appropriate. Let $C_1, C_2, \ldots, C_K$ denote the fuzzy adaptive clusters intersected by the final sample, and let W_k and Y_k denote the corresponding cluster level design weight and total disease burden within cluster k, respectively. The total estimator then takes the form

$$\hat{T} = \sum_{k=1}^{K} W_k^* Y_k,$$

where W_k^* is the defuzzified cluster weight. Under this representation, the variance may be estimated using a between cluster dispersion estimator

$$\widehat{\mathrm{Var}}(\hat{T}) = \sum_{k=1}^{K}\sum_{\ell=1}^{K}\left(\hat{\pi}_{k\ell} - \hat{\pi}_k\hat{\pi}_\ell\right)\frac{Y_k}{\hat{\pi}_k}\frac{Y_\ell}{\hat{\pi}_\ell},$$

where $\hat{\pi}_k$ and $\hat{\pi}_{k\ell}$ denote Monte Carlo approximations to the first and second order cluster inclusion probabilities. This formulation explicitly recognizes the network structure induced by fuzzy clustering and provides a natural extension of conventional adaptive cluster variance estimators.

An important qualitative feature of variance behavior under fuzzy adaptive sampling is its stabilizing effect under weak or diffuse hotspot conditions. Because inclusion probabilities vary smoothly with fuzzy membership rather than exhibiting abrupt jumps, the resulting design weights are less extreme and the dispersion of the weighted contributions $w^*_i\, y_i$ across sampled units is typically reduced. This smoothing effect leads to smaller sampling variance in comparison with crisp adaptive designs, particularly in scenarios where true disease clusters are only moderately elevated above the background level.

From a practical inference perspective, the availability of a reliable variance estimator enables the construction of confidence intervals and uncertainty bands for all key target parameters. For example, an approximate 100(1 − γ)% confidence interval for the population total T may be constructed as

$$\hat{T} \pm z_{\gamma/2}\sqrt{\widehat{\mathrm{Var}}(\hat{T})},$$

where $z_{\gamma/2}$ denotes the standard normal quantile. Similar interval constructions apply to hotspot specific totals and concentration ratios. These inferential summaries play a central role in supporting evidence based public health decision making under the proposed fuzzy adaptive surveillance framework.

With the estimation and variance structure now fully developed, the methodological core of the fuzzy adaptive cluster sampling design is complete. The next section will translate this theoretical framework into a concrete simulation environment in which the detection efficiency, bias behavior, and variance performance of the proposed design will be systematically evaluated and compared with classical competing sampling strategies.

2.15. Theoretical properties of the proposed estimators

The development of estimators under the fuzzy adaptive cluster sampling framework must be complemented by a rigorous examination of their theoretical properties in order to establish their validity and inferential soundness. While the preceding section constructed operational estimators for key surveil- lance parameters and provided practical variance estimators through design based and simulation-based arguments, it is now essential to formally investigate the fundamental large sample and finite population properties of these estimators. In particular, the questions of unbiasedness, consistency, efficiency relative to competing designs, and the structural conditions under which the fuzzy adaptive framework yields superior performance must be addressed within a unified theoretical perspective. This section is devoted to a systematic study of these properties, with careful attention to the interplay between fuzzy triggering, spatial dependence, and unequal probability weighting that characterizes the proposed design.

2.16. Unbiasedness and approximate unbiasedness

The property of unbiasedness occupies a central position in design-based inference, as it provides a fundamental guarantee that, on average over repeated applications of the sampling design, the estimator reproduces the true population value of the target parameter. Under fuzzy adaptive cluster sampling, the presence of stochastic expansion rules, spatial dependence, and fuzzy weighted inclusion probabilities makes the concept of exact unbiasedness subtler than in classical fixed probability designs. Never- theless, it is still possible to establish precise conditions under which the proposed defuzzified estimators are design unbiased or asymptotically unbiased in an approximate sense.

Let $T = \sum_{i=1}^{N} y_i$ denote the true population total of the rare disease count, and let

$$\hat{T} = \sum_{i \in s} w_i^* y_i$$

denote the defuzzified inverse probability weighted estimator constructed under the fuzzy adaptive design, where w_i^* is the centroid defuzzified form of the fuzzy design weight associated with unit *i*. Suppose first that the fuzzy weights satisfy

$$E(w_i^*) = \frac{1}{\pi_i},$$

where π_i is the true first order inclusion probability of unit *i* under the joint mechanism of initial randomization and fuzzy adaptive expansion. Then, taking expectation with respect to the sampling design, we obtain

$$E(\hat{T}) = E\left(\sum_{i \in s} w_i^* y_i\right) = \sum_{i=1}^{N} E(I_i w_i^*) y_i = \sum_{i=1}^{N} \pi_i \frac{1}{\pi_i} y_i = \sum_{i=1}^{N} y_i = T,$$

which establishes exact design unbiasedness of $\hat{T}$. This result demonstrates that, despite the complexity of the fuzzy adaptive selection mechanism, unbiased estimation of the population total is preserved as long as the defuzzified weights accurately represent the inverse of the true inclusion probabilities.

In practical implementations, the true inclusion probabilities π_i are not available analytically and must be approximated through Monte Carlo simulation. Let $\hat{\pi}_i$ denote a consistent estimator of π_i obtained from repeated simulation of the fuzzy adaptive design, and let $w_i^* = 1/\hat{\pi}_i$ denote the corresponding operational design weight. Under standard regularity conditions for Monte Carlo approximation, $\hat{\pi}_i \xrightarrow{p} \pi_i$ as the number of replications increases. Consequently,

$$E(\hat{T}) = \sum_{i=1}^{N} y_i \, E\left(\frac{\pi_i}{\hat{\pi}_i}\right),$$

which converges to T as the Monte Carlo error diminishes. In this sense, the estimator $\hat{T}$ is asymptotically design unbiased with respect to the simulation size used to approximate the fuzzy inclusion probabilities.

When the observed disease counts themselves are fuzzy and denoted by $\tilde{y}_i$, the estimator takes the fuzzy weighted form

$$\tilde{T} = \sum_{i \in s} \tilde{w}_i \tilde{y}_i,$$

which is subsequently defuzzified to yield $\hat{T}$. If the centroid defuzzification is applied and if the fuzzy representations of w_i and y_i are centered about their true crisp counterparts, then the resulting defuzzified estimator retains approximate unbiasedness in the sense that

$$E(\hat{T}) \approx T,$$

with the approximation error governed by higher order moments of the fuzzy membership functions. This form of approximate unbiasedness is particularly relevant in real surveillance environments where both inclusion probabilities and observed counts are subject to unavoidable imprecision.

An analogous argument applies to other target parameters such as the population mean $\bar{Y}$ and the hotspot specific total T_{H_α}. In each case, unbiasedness or approximate unbiasedness follows directly from the inverse probability weighting structure of the estimator combined with the consistency of the fuzzy inclusion probability approximation. These results collectively establish that the proposed fuzzy adaptive estimation framework preserves the fundamental design based unbiasedness principle that underpins valid survey inference, even in the presence of graded, data driven, and spatially dependent inclusion mechanisms.

The question of consistency, which concerns the convergence of the estimator to the true parameter value as sample size increases, will be examined in the next subsection.

2.17. Consistency of the fuzzy adaptive estimator

While unbiasedness describes the average long run behavior of an estimator under repeated sampling, consistency addresses the more practically relevant question of whether the estimator converges to the true population parameter as the effective sample information increases. In the context of rare disease surveillance, consistency is especially important because policy decisions are often based on progressively accumulating data over space and time. For the proposed fuzzy adaptive cluster sampling framework, consistency must be examined in the presence of three interacting features, namely unequal probabil- ity selection, spatial dependence induced by adaptive clustering, and the additional layer of stochastic variation introduced through fuzzy triggering and defuzzification.

Let $\hat{T} = \sum_{i \in s} w_i^* y_i$ denote the defuzzified estimator of the population total, where $w_i^* = 1/\hat{\pi}_i$ and $\hat{\pi}_i$ is the Monte Carlo approximation to the true inclusion probability π_i. Consistency of $\hat{T}$ may be studied under an increasing domain or increasing sample size asymptotic framework in which the number of spatial units N tends to infinity, the initial sample size n_0 grows with N, and the fuzzy adaptive expansion mechanism continues to operate under stable design rules. Under such a framework, consistency requires that

$$\hat{T} - T \xrightarrow{p} 0 \quad \text{as} \quad N \to \infty.$$

A sufficient set of conditions for design consistency can be stated in intuitive and practically interpretable terms. First, the inclusion probability of every unit must be bounded away from zero, that is,

$$\inf_{i \in U} \pi_i > 0,$$

which ensures that no spatial unit becomes asymptotically invisible to the sampling design. Under the fuzzy adaptive framework, this condition is naturally satisfied because every unit retains a positive probability of inclusion through the initial randomization stage, regardless of its fuzzy hotspot membership.

Second, the Monte Carlo approximation error in the estimated inclusion probabilities must vanish asymptotically, so that

$$\max_{i \in U} |\hat{\pi}_i - \pi_i| \xrightarrow{p} 0.$$

This condition is ensured by allowing the number of Monte Carlo replications used for estimating the fuzzy inclusion probabilities to increase sufficiently fast relative to the growth of *N*. In practical terms, this requirement is mild and easily satisfied with modern computational resources.

Third, the disease counts y_i must satisfy a mild moment condition such as

$$\frac{1}{N}\sum_{i=1}^{N} y_i^2 = O(1),$$

which simply prevents the presence of extremely large isolated values that could dominate the estimation process. In rare disease surveillance, this assumption is generally reasonable because case counts within individual spatial units remain bounded by population size and reporting capacity.

Under these conditions, a standard probability limit argument yields

$$\hat{T} - T = \sum_{i=1}^{N} \left(\frac{I_i}{\hat{\pi}_i} - 1 \right) y_i \xrightarrow{p} 0,$$

which establishes the design consistency of the defuzzified fuzzy adaptive estimator for the population total. The same reasoning extends naturally to the population mean estimator $\hat{\bar{Y}}$ and to hotspot specific totals $\hat{T}_{H_\alpha}$, since these parameters are continuous transformations of $\hat{T}$ the maintained regularity conditions.

When the observed disease counts are themselves fuzzy and defuzzified through the centroid rule, the consistency result continues to hold in an approximate sense. Provided the fuzzy representations of the counts are centered around their true crisp values and the dispersion of the corresponding membership functions does not increase with N, the additional error introduced by defuzzification remains asymptotically negligible. Consequently, the overall estimator continues to converge in probability to the true population parameter.

An important conceptual implication of this result is that the introduction of fuzzy triggering and fuzzy weighting does not compromise the long run recovery of true disease burden, despite the substantial increase in design flexibility and robustness. On the contrary, the fuzzy adaptive mechanism preserves consistency while simultaneously improving finite sample stability under weak or diffuse hotspot conditions. In the next subsection, this improvement in finite sample performance will be formally examined through a comparison of efficiency between the fuzzy adaptive design and classical competing sampling strategies.

2.18. Efficiency comparison with classical designs

While unbiasedness and consistency establish the long run validity of an estimator, the practical use- fulness of a sampling design in rare disease surveillance is ultimately determined by its finite sample efficiency. In this context, efficiency refers to the ability of a design to achieve smaller variance and higher detection accuracy for a given sampling effort. The primary motivation for introducing fuzzy adaptive cluster sampling is precisely to enhance efficiency under conditions where disease cases are rare, weakly clustered, and embedded in substantial background noise. It is therefore essential to formally examine how the proposed fuzzy adaptive estimators compare with those obtained under classical com- peting designs such as simple random sampling, conventional cluster sampling, and crisp adaptive cluster sampling.

Let $\hat{T}_{\text{FACS}}$ denote the estimator of the population total obtained under the fuzzy adaptive cluster sampling design, and let $\hat{T}_d$ denote the corresponding estimator under a competing design *d*. The relative efficiency of the fuzzy adaptive estimator with respect to design d is defined as

$$\text{RE}_d = \frac{\text{Var}(\hat{T}_d)}{\text{Var}(\hat{T}_{\text{FACS}})}.$$

If RE_d is larger than one, the fuzzy adaptive design is more efficient than the alternative. This definition furnishes a straight quantification for the comparison of the sampling performances among different design strategies when run under the same population scenarios.

Under simple random sampling without replacement, the variance of the Horvitz Thomson estimator of the total is given by

$$\text{Var}(\hat{T}_{\text{SRS}}) = \left(1 - \frac{n_0}{N}\right) \frac{N^2}{n_0} S_y^2,$$

where S_y^2 denotes the finite population variance of the disease counts. In rare disease settings, S_y^2 is typically dominated by a large mass of zeros and a small number of positive counts located in isolated spatial pockets. As a result, simple random sampling wastes a substantial portion of its sampling effort in background regions that contain no useful information about disease clustering, leading to large sampling variances and poor hotspot detection performance.

Conventional cluster sampling improves upon simple random sampling by exploiting spatial grouping of population units, but it still relies on a fixed cluster structure and does not adapt to observed disease realization. As a consequence, its efficiency remains highly sensitive to the arbitrary choice of cluster boundaries and often deteriorates when true disease clusters do not align well with the predefined cluster layout. In contrast, crisp adaptive cluster sampling improves efficiency by expanding the sample whenever a positive case is observed. However, the crisp nature of the triggering rule introduces severe instability in weak signal environments. Borderline clusters either fail to activate expansion altogether or activate excessive expansion due to spurious single case occurrences, resulting in highly variable sample sizes and extreme design weights.

The fuzzy adaptive framework fundamentally alters this behavior by replacing the rigid triggering rule with a smooth expansion mechanism driven by graded hotspot membership. Units located near the boundary of true clusters activate expansion with moderate probability rather than with certainty or not at all. This moderation in expansion leads to more stable cluster growth, smoother inclusion probability gradients, and substantially reduced dispersion in the corresponding inverse probability weights. Since estimator variance under unequal probability sampling is driven primarily by the variability of weighted contributions, the fuzzy design naturally achieves lower variance in finite samples relative to its crisp counterpart.

More formally, under both crisp and fuzzy adaptive designs, the variance of the total estimator may be expressed in the general form

$$\mathrm{Var}(\hat{T}) = \sum_{i=1}^{N}\sum_{j=1}^{N}(\pi_{ij} - \pi_i\pi_j)\frac{y_i}{\pi_i}\frac{y_j}{\pi_j}.$$

Under the fuzzy design, the joint inclusion probabilities π_{ij} vary smoothly with spatial distance and fuzzy hotspot membership, whereas under the crisp design they exhibit abrupt transitions associated with deterministic cluster boundaries. This difference has a direct effect on the magnitude of the covariance terms $(\pi_{ij} - \pi_i\pi_j)$ and leads to a systematic reduction in the contribution of extreme cross product terms under fuzzy adaptation. The net effect is a smaller overall variance of $\hat{T}_{\mathrm{FACS}}$ under weak and moderately clustered disease patterns.

Another important dimension of efficiency in surveillance applications is hotspot detection accuracy rather than estimation alone. Because fuzzy adaptive sampling explores moderate risk neighborhoods more cautiously and persistently than crisp designs, it exhibits higher probability of detecting emerging or low intensity clusters before they escalate into severe outbreaks. At the same time, it avoids the excessive oversampling of purely random noise that often afflicts crisp adaptive designs in low prevalence regimes. This dual advantage translates into both improved estimation precision and enhanced early detection capability.

Overall, the discussions of theories imply that the fuzzy adaptive cluster sampling outperforms simple random sampling, conventional cluster sampling and crisp adaptive cluster sampling in efficiency among such a wide range of rarity levels of diseases which include spatial heterogeneity, weak clustering and uncertainty of diagnosis. The degree of this efficiency gain in exact will be evaluated by thorough simulation studies in the following sections. In the following subsection, the exact conditions on the structure guarantee that the fuzzy adaptive design has the best performance will be given.

2.19. Conditions for superiority of the fuzzy adaptive design

The efficiency gains and stability advantages of the fuzzy adaptive cluster sampling design do not arise in a vacuum, but are instead the result of specific structural features of the underlying disease process and the sampling environment. It is therefore important to clearly delineate the conditions under which the fuzzy adaptive framework can be expected to dominate classical sampling strategies in a systematic and reproducible manner. These conditions are not restrictive in practice and, in fact, closely mirror the empirical realities encountered in most rare disease surveillance systems.

A first fundamental condition concerns the degree of spatial heterogeneity in the underlying disease risk surface. The fuzzy adaptive design derives its strength from its ability to concentrate sampling effort in regions of elevated risk without imposing rigid boundaries. When the intensity function λ(s) exhibits meaningful spatial variation with localized regions of moderate to high risk embedded within a broad background of low risk, the graded expansion mechanism of the fuzzy design is able to progressively discover and explore these regions with increasing precision. In contrast, if disease risk is perfectly homogeneous across space, no adaptive design can outperform simple random sampling in a systematic way. Thus, spatial heterogeneity of risk constitutes the primary structural prerequisite for fuzzy adaptive superiority.

A second critical condition is related to the strength of clustering relative to background noise. The fuzzy adaptive design is particularly advantageous in environments characterized by weak to moderate clustering, where disease signals are present but not sufficiently strong to consistently activate crisp threshold-based expansion. In such scenarios, classical adaptive cluster sampling often fails to expand because the rigid triggering rule is not satisfied with sufficient frequency, leading to missed detection of emerging or borderline clusters. The fuzzy design, by contrast, responds to partial membership and therefore propagates exploration gradually even under weak signals, enabling earlier and more reliable detection. When clustering is extremely strong and sharply separated from the background, both crisp and fuzzy adaptive designs perform well, and the relative advantage of fuzziness diminishes.

A third condition concerns the prevalence of diagnostic uncertainty and measurement error. In many rare disease surveillance systems, observed case counts are affected by underreporting, delayed confirmation, and misclassification. Under such conditions, crisp adaptive designs are highly sensitive to single erroneous observations, which may either spur excessive expansion from a spurious signal or suppress expansion from a true but underreported cluster. The fuzzy adaptive framework, by incorporating un-certainty directly through graded hotspot membership and probabilistic triggering, dampens the impact of such errors and thereby maintains a more stable and reliable sampling trajectory. Consequently, the superiority of the fuzzy design becomes more pronounced as the level of diagnostic uncertainty increases. A fourth condition pertains to the structure of the neighborhood graph that governs spatial propagation. The fuzzy adaptive framework performs best when the neighborhood structure reflects genuine pathways of disease transmission or exposure continuity rather than purely arbitrary geometric adjacency. When neighborhoods are defined in a way that aligns with human mobility, environmental transport, or social interaction networks, fuzzy propagation follows epidemiologically meaningful routes and concen-trates sampling effort where it is most informative. Under poorly specified neighborhood structures, all adaptive designs experience some loss of efficiency, but the graded nature of fuzzy propagation still offers greater protection against extreme misallocation of sampling effort than crisp designs.

A final condition relates to the available sampling budget and operational constraints. Fuzzy adaptive sampling is particularly well suited to moderate budget regimes where full enumeration of all potential hotspot regions is infeasible but where purely fixed designs remain inefficient. In such intermediate resource environments, the ability to tune the aggressiveness of fuzzy expansion through the choice of membership scaling and sensitivity parameters allows the design to achieve an optimal balance between detection sensitivity and survey cost. In very low budget regimes, no adaptive design can fully compensate for insufficient coverage, while in extremely high budget regimes the marginal gains of adaptivity diminish as near census level coverage becomes attainable.

Together, these conditions delineate a broad and practically relevant regime in which the fuzzy adaptive cluster sampling design is theoretically and operationally superior to its classical counterparts. They show that the proposed framework is not a narrow methodological curiosity, but a robust and broadly applicable surveillance strategy tailored to the structural features of rare disease processes as they arise in real world settings. With the theoretical properties now fully established, the next section will translate these findings into a controlled simulation environment in which the empirical performance of the fuzzy adaptive design will be systematically evaluated.

2.20. Simulation design and experimental setup

The theoretical results developed in the preceding section establish the validity, consistency, and efficiency advantages of the proposed fuzzy adaptive cluster sampling framework under well-defined structural conditions. However, the practical performance of the design in realistic surveillance environments must be assessed through controlled simulation experiments, where the true disease generating mechanism, the spatial structure of hotspots, and the sampling design parameters are all known and can be systematically manipulated. Simulation plays a particularly central role in the present study because it allows a direct empirical evaluation of detection accuracy, estimation bias, variance behavior, and cost efficiency under a wide range of rare disease scenarios that cannot be exhaustively examined through analytical arguments alone. This section therefore describes in detail the construction of the synthetic spatial dis- ease environments, the specification of competing sampling designs, and the performance metrics used to compare their relative effectiveness under identical experimental conditions.

2.21. Generation of inhomogeneous spatial point processes

The starting point of the simulation framework is the construction of synthetic spatial disease environments that realistically reflect the heterogeneous and clustered nature of rare disease incidence. Since the proposed fuzzy adaptive design is explicitly built upon the inhomogeneous spatial point process representation, the simulation model must be sufficiently flexible to generate a wide spectrum of spatial risk patterns ranging from nearly homogeneous background risk to strongly localized hotspot structures. Such flexibility is crucial to test the strength and applicability of the suggested approach for various epidemiological systems.

Let $D = [0, L] \times [0, L]$ denote a bounded square study region in the two-dimensional plane, where L is a positive constant controlling the spatial scale of the experiment. Disease cases are generated as realizations of an inhomogeneous Poisson point process defined on D with spatially varying intensity function λ(s), where s = (s_1, s_2) denotes a generic spatial location. Conditional on λ(s), the number of disease cases falling inside any measurable subregion $B \subset D$ follows a Poisson distribution with mean

$$E\{N(B)\} = \int_B \lambda(\mathbf{s})\, d\mathbf{s}.$$

Moreover, the spatial locations of cases within B are independently distributed with density proportional to λ(s).

To represent background disease risk, a baseline intensity level $\lambda_0 > 0$ is specified over the entire region D. Spatial heterogeneity and hotspot behavior are then introduced

Citation: Behera, J. (2026). Efficient Detection of Rare Disease Hotspots Using Fuzzy Adaptive Cluster Sampling under Inhomogeneous Spatial Point Processes. *Adv Brain Res Neurosci, 2*(1), 01-38.

by superimposing one or more localized intensity peaks on this background surface. A commonly used and analytically convenient form of the composite intensity function is

$$\lambda(\mathbf{s}) = \lambda_0 + \sum_{h=1}^{H} \lambda_h \exp\left\{-\frac{1}{2}(\mathbf{s}-\boldsymbol{\mu}_h)^\top \boldsymbol{\Sigma}_h^{-1}(\mathbf{s}-\boldsymbol{\mu}_h)\right\},$$

where H denotes the number of hotspot centers, μ_h is the spatial location of the h th hotspot, Σ_h is a positive definite dispersion matrix controlling the spatial spread of the hotspot, and λ_h represents the peak excess intensity associated with that hotspot. This Gaussian shaped intensity formulation allows the generation of smooth, spatially localized risk surfaces that closely resemble the diffuse nature of realworld disease hotspots.

By varying the parameters λ_0, λ_h, and Σ_h, a wide range of epidemiological scenarios can be created. Weak clustering regimes are generated by choosing small values of λ_h and large dispersion matrices, resulting in shallow and diffuse hotspots. Strong clustering regimes are obtained by selecting large values of λ_h together with small dispersion matrices, producing sharp and compact hotspot cores. Intermediate regimes arise naturally through moderate choices of these parameters. This parameter controlled construction allows a systematic investigation of the conditions under which the fuzzy adaptive design delivers the greatest efficiency gains.

Once a realization of the spatial point process has been generated, the continuous region D is parti- tioned into a regular grid of N square areal units of equal area. The observed disease count y_i for unit i is then computed as the total number of simulated point process realizations falling inside that unit. This discretization step establishes the direct link between the continuous spatial disease generating mechanism and the discrete spatial units that serve as the population for the sampling designs under comparison.

To ensure statistical stability of the simulation results, each experimental configuration defined by a particular choice of intensity parameters and hotspot structure is replicated a large number of times. For each replication, a fresh realization of the inhomogeneous spatial point process is generated, followed by the application of the competing sampling designs and the computation of the corresponding estimators. The resulting empirical distributions of the estimators form the basis for evaluating detection accuracy, bias, variance, and relative efficiency in the subsequent performance analysis.

2.22. Specification of true hotspot structures

A central requirement of any meaningful simulation study on hotspot detection is the explicit and controlled specification of the true underlying hotspot configurations against which the performance of competing sampling designs can be objectively evaluated. In the present framework, the notion of a true hotspot is defined directly through the structure of the spatial intensity function that governs the inhomogeneous point process. This approach ensures that hotspot truth is not introduced through arbitrary post hoc classification, but is instead embedded intrinsically within the disease generating mechanism itself.

For each simulated scenario, a true hotspot region is associated with each localized peak of the intensity surface. Let $\mu_h = (\mu_{h1}, \mu_{h2})$ denote the center of the h th hotspot and let Σ_h denote its dispersion matrix. The true continuous hotspot region corresponding to this intensity peak is defined as the elliptical set

$$\mathcal{H}_h = \left\{\mathbf{s} \in D : (\mathbf{s}-\boldsymbol{\mu}_h)^\top \boldsymbol{\Sigma}_h^{-1}(\mathbf{s}-\boldsymbol{\mu}_h) \le c_h\right\},$$

where c_h is a positive constant chosen so that $\mathcal{H}_h$ captures a pre specified proportion of the excess disease risk associated with the h th hotspot. This construction ensures that the true hotspot is defined directly in terms of the geometric structure of the intensity surface and possesses a well defined spatial shape and size.

At the level of discrete spatial units, the true hotspot membership of unit i with spatial support A_i is defined by the degree of overlap between A_i and the continuous hotspot region $\mathcal{H}_h$. A unit is classified as belonging to the true hotspot if

$$\frac{|A_i \cap \mathcal{H}_h|}{|A_i|} > 0,$$

where $|\cdot|$ denotes Lebesgue measure. This definition ensures that a unit is regarded as a true hotspot unit whenever any non zero portion of its spatial support lies within the underlying continuous hotspot region. When multiple hotspot centers are present, a unit is classified as hotspot associated if it overlaps with at least one of the regions $\mathcal{H}_1, \ldots, \mathcal{H}_H$.

To enable a nuanced evaluation of detection performance, the strength and spatial extent of the true hotspots are systematically varied across simulation scenarios. Weak hotspot configurations are characterized by small contrast between λ_h and λ_0 together with large dispersion matrices, producing broad and low amplitude risk elevations. Strong hotspot configurations are characterized by large values of λ_h and tightly concentrated dispersion, generating sharply defined and highly localized cores. Intermediate configurations bridge these two extremes and represent the most challenging regime for adaptive detection methods.

In addition to varying hotspot strength, the number of concurrent hotspot centers H is also varied across scenarios to reflect environments with single outbreak foci as well as more complex multi focus disease patterns. This diversification of true spatial structures allows the simulation study to examine not only the sensitivity of the fuzzy adaptive design to individual outbreaks, but also its ability to resolve multiple competing hotspot signals without excessive interference or loss of efficiency.

By explicitly embedding true hotspot structures within the spatial point process and propagating their influence through to the discrete areal unit level, the simulation framework establishes a clear and objective ground truth against which detection accuracy, false discovery behavior, and estimation bias can be rigorously assessed in the subsequent analysis.

2.23. Design of competing sampling strategies

A meaningful assessment of the proposed fuzzy adaptive cluster sampling framework requires its sys- tematic comparison with well-established competing sampling strategies that represent the standard tools currently used in rare disease surveillance. These competing designs serve as benchmarks against which the relative gains in efficiency, stability, and detection accuracy achieved by fuzzy adaptation can be objectively quantified. All competing designs are implemented on the same simulated populations generated from the inhomogeneous spatial point processes described earlier, and they are calibrated to operate under comparable expected sampling efforts to ensure fairness of comparison.

The first benchmark design considered is simple random sampling without replacement. Under this design, an initial sample of size n_0 is selected uniformly at random from the N spatial units, and no further expansion is performed regardless of the observed disease counts. The corresponding estimator of the population total is the standard Horvitz Thomson estimator

$$\hat{T}_{\mathrm{SRS}} = \frac{N}{n_0} \sum_{i \in s_0} y_i.$$

This design represents the most basic probability sampling strategy and serves as a natural lower bound in terms of efficiency for rare and clustered populations, since it does not exploit any spatial structure or adaptive information.

The second competing design is conventional cluster sampling. Under this design, the population is partitioned in advance into a fixed set of non overlapping clusters, typically corresponding to contiguous groups of spatial units. A sample of clusters is then selected using simple random sampling, and all units within the sampled clusters are observed. While this design exploits spatial grouping to reduce field costs, it does not adapt to the realized disease distribution and therefore remains highly sensitive to the arbitrary choice of cluster boundaries. The resulting estimator takes the standard cluster sampling form with design weights determined by the inverse of the cluster selection probabilities.

The third benchmark design is classical adaptive cluster sampling with a crisp triggering rule. Under this design, an initial sample of size n_0 is selected by simple random sampling, and neighborhood expansion is triggered whenever a sampled unit exhibits at least one observed disease case. All neighbors of such units are added to the sample deterministically, and the procedure is applied recursively until no further expansions occur. Estimation is performed at the network level using the classical Horvitz Thomson network estimator. This design represents the closest classical competitor to the proposed fuzzy adaptive framework and is particularly important for isolating the incremental gains attributable solely to the introduction of fuzziness.

The proposed fuzzy adaptive cluster sampling design is implemented using the same initial sample size n_0 and the same neighborhood structure as the crisp adaptive design, but with expansion governed by the graded fuzzy triggering rule defined earlier. This ensures that any observed performance differences between the crisp and fuzzy designs can be attributed directly to the nature of the triggering mechanism rather than to differences in initial sample size or neighborhood specification.

To ensure comparability across designs, all strategies are calibrated so that their expected total sample sizes are approximately equal under each simulation scenario. For fixed designs such as simple random sampling and conventional cluster sampling, this calibration is achieved by appropriate choice of n_0 or the number of sampled clusters. For adaptive designs, calibration is achieved by selecting the fuzzy expansion sensitivity parameters so that the average final sample size over repeated simulations matches that of the competing crisp adaptive design. This alignment ensures that efficiency comparisons are not confounded by trivial differences in sampling effort.

All designs are embedded within a single simulation environment and subjected to similar sampling budgets, allowing the intrinsic performance traits of the individual strategies to be isolated. This comparative design establishes a rigorous experimental platform for evaluating how the proposed fuzzy adaptive framework behaves relative to the most widely used classical alternatives, both in terms of estimation accuracy and hotspot detection capability.

2.24. Performance measures and evaluation criteria

The comparative evaluation of the proposed fuzzy adaptive cluster sampling design and the competing sampling strategies requires a carefully defined set of performance measures that reflect both the statistical and epidemiological objectives of rare disease surveillance. Because the purpose of surveillance systems is to provide estimates of the burden of disease, to allow identification of spatial hotspots, and to do so using scarce sampling resources, the evaluation criterion should reflect accuracy of estimation, detection performance and operational cost effectiveness. This subsection formally introduces quanti- tative performance measures, which evaluate these multiple aspects of performance in a unified and reproducible manner, for the simulation study.

The first and most fundamental performance measure is the empirical bias of the estimator of the population total. Let $\hat{T}^{(b)}$ denote the estimated total obtained from the b th simulation replication under a given sampling design, and let T denote the true population total generated by the spatial point process in that replication. Over B independent replications, the empirical bias is defined as

$$\text{Bias}(\hat{T}) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{T}^{(b)} - T^{(b)}\right),$$

which measures the systematic tendency of the estimator to overestimate or underestimate the true disease burden. An estimator is regarded as empirically unbiased when this quantity is close to zero across a wide range of simulation scenarios.

Closely related to bias is the empirical mean squared error, which serves as a comprehensive measure of estimation accuracy by combining both variance and squared bias effects. It is defined as

$$\text{MSE}(\hat{T}) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{T}^{(b)} - T^{(b)}\right)^2.$$

Low mean squared error values correspond to better performance of an estimator in a finite sample. In the simulation results, we consider the mean squared error as the main measure for comparing the efficiency of estimators for the two sampling designs under the same disease generating model.

In addition to total burden estimation, hotspot detection accuracy constitutes a critical dimension of performance. Let $\mathcal{H}$ denote the set of true hotspot units identified from the underlying intensity surface, and let $\hat{\mathcal{H}}$ denote the set of units detected as hotspots by a given sampling design based on the fuzzy or crisp classification rule. The sensitivity of hotspot detection is defined as

$$\text{Sensitivity} = \frac{|\hat{\mathcal{H}} \cap \mathcal{H}|}{|\mathcal{H}|},$$

which calculates the covery of true positive hotspot cells by the design.The specificity of hotspot detection is defined as

$$\text{Specificity} = \frac{|U \setminus (\hat{\mathcal{H}} \cup \mathcal{H})|}{|U \setminus \mathcal{H}|},$$

which measures the proportion of non hotspot units that are correctly classified as background. These two measures jointly characterize the tradeoff between early detection sensitivity and false alarm control that is central to effective surveillance.

A further important criterion is the relative efficiency of estimators, defined as the ratio of mean squared errors between competing designs. For two designs d_1 and d_2, the empirical relative efficiency is defined as

$$\text{RE}(d_1, d_2) = \frac{\text{MSE}(\hat{T}_{d_2})}{\text{MSE}(\hat{T}_{d_1})},$$

with values greater than unity indicating superior efficiency of design d_1 relative to design d_2. This metric provides a dimensionless and easily interpretable summary of the magnitude of performance gains afforded by the fuzzy adaptive design.

From an operational perspective, the total final sample size $n = |s|$ provides a direct measure of survey cost under each design. The empirical distribution of n across simulation replications is used to assess the stability of sampling effort and to quantify the extent to which adaptive designs exhibit variability in field workload. Designs that produce highly volatile sample sizes are less desirable in practice because they complicate logistical planning and resource allocation.

Finally, to capture the combined detection and cost efficiency of the sampling designs, a composite detection efficiency index is defined as the ratio of detection sensitivity to expected sample size. This index provides a concise summary of how effectively each design converts sampling effort into successful hotspot detection.

Together, these performance measures form a comprehensive and coherent evaluation framework for the simulation study. They allow the proposed fuzzy adaptive cluster sampling design to be assessed not only in terms of classical statistical accuracy, but also in terms of its practical utility for real time rare disease surveillance. In the next subsection, the precise algorithmic steps used to implement the simulation experiments and compute these performance measures will be formally described.

2.25. Simulation algorithm

Algorithm 1 Simulation Algorithm for Evaluating Fuzzy Adaptive Cluster Sampling

Specify the study region $D = [0, L] \times [0, L]$ and partition it into N equal area spatial units.
Fix baseline intensity λ_0, number of hotspots H, hotspot centers $\boldsymbol{\mu}_h$, dispersion matrices $\boldsymbol{\Sigma}_h$, and hotspot intensities λ_h for $h = 1, 2, \ldots, H$.
Construct the inhomogeneous intensity function

$$\lambda(\mathbf{s}) = \lambda_0 + \sum_{h=1}^{H} \lambda_h \exp\left\{-\tfrac{1}{2}(\mathbf{s} - \boldsymbol{\mu}_h)^\top \boldsymbol{\Sigma}_h^{-1}(\mathbf{s} - \boldsymbol{\mu}_h)\right\}.$$

for $b = 1$ **to** B **do**
  Generate a realization of the inhomogeneous Poisson point process with intensity $\lambda(\mathbf{s})$ over D.
  Count the number of simulated disease cases $y_i^{(b)}$ falling in each spatial unit $i = 1, 2, \ldots, N$.
  Identify the true hotspot set $\mathcal{H}^{(b)}$ based on the geometric hotspot definition.
  Compute fuzzy hotspot memberships $\mu_{\tilde{H}}^{(b)}(i)$ for all spatial units.
  Draw an initial sample s_0 of size n_0 using simple random sampling without replacement.
  Apply Simple Random Sampling estimator $\hat{T}_{\text{SRS}}^{(b)}$.
  Apply Conventional Cluster Sampling and compute estimator $\hat{T}_{\text{CS}}^{(b)}$.
  Apply Crisp Adaptive Cluster Sampling using deterministic triggering and compute estimator $\hat{T}_{\text{CACS}}^{(b)}$.
  Apply Fuzzy Adaptive Cluster Sampling:
    Initialize $s = s_0$.
    Repeat
      For every $i \in s$, generate $B_i \sim \text{Bernoulli}(\phi(i))$.
      If $B_i = 1$, add all neighbors $j \in \mathcal{N}(i)$ to s.
    Until no further units are added.
    Compute fuzzy inclusion probabilities by Monte Carlo repetition.
    Construct fuzzy design weights and defuzzify them.
    Compute fuzzy adaptive estimator $\hat{T}_{\text{FACS}}^{(b)}$.
  Record estimators $\hat{T}_{\text{SRS}}^{(b)}$, $\hat{T}_{\text{CS}}^{(b)}$, $\hat{T}_{\text{CACS}}^{(b)}$, and $\hat{T}_{\text{FACS}}^{(b)}$.
  Record detection results and final sample size $n^{(b)}$.
end for
Compute empirical bias, mean squared error, detection sensitivity, detection specificity, relative efficiency, and sampling cost variability using B replications.

3. Simulation Results and Performance Evaluation

In this section, we provide a full empirical study on the fuzzy adaptive cluster sampling design utilizing the big picture simulation technique embedded in the previous section. While the theoretical analysis established the fundamental properties of unbiasedness, consistency, and efficiency under well-defined structural conditions, the simulation results provide concrete numerical evidence of how these properties manifest in realistic finite sample settings. The performance of the fuzzy adaptive design is evaluated across a wide spectrum of spatial disease environments characterized by varying levels of background intensity, hotspot strength, clustering sharpness, and diagnostic uncertainty. Its behavior is systematically compared with that of simple random sampling, conventional cluster sampling, and classical crisp adaptive cluster sampling using unified performance metrics including detection accuracy, estimation bias, mean squared error, sampling effort variability, and relative efficiency. Through this empirical investigation, the section demonstrates not only the statistical advantages of the fuzzy adaptive framework but also its practical superiority for real time rare disease surveillance under operational constraints.

3.1. Bias and Mean Squared Error Comparison

Beyond hotspot detection, the reliability of a surveillance design is ultimately determined by the statistical accuracy of its estimators of population level disease burden. In rare disease settings, where true signals are weak and highly localized, estimation bias and variability often dominate inference and can severely distort situational awareness if not properly controlled. This subsection therefore examines the empirical bias and mean squared error behavior of the estimators obtained under the competing sampling designs, with particular emphasis on comparing the fuzzy adaptive framework to its crisp adaptive and fixed design counterparts.

Across all simulation scenarios, the estimator based on simple random sampling exhibits the largest dispersion and the most pronounced instability in finite samples. Although it remains approximately unbiased in expectation, its empirical bias fluctuates widely across replications when hotspot intensity is low and the proportion of nonzero observations is extremely small. This behavior is driven by the high probability that the initial random sample completely misses the true hotspot region, resulting in frequent realizations of near zero estimates followed by occasional extreme overestimates when one or two hotspot units happen to be sampled. This heavy tailed error structure leads to inflated mean squared error and renders simple random sampling highly inefficient for rare and clustered disease processes.

Conventional cluster sampling exhibits modest improvement over simple random sampling in scenarios where the predefined clusters happen to intersect the true hotspot regions in a balanced manner. However, its estimation performance remains highly sensitive to the arbitrary spatial partition used to define the clusters. When true hotspots straddle multiple cluster boundaries or occupy only small portions of large clusters, the resulting estimators exhibit noticeable positive bias and substantial within replication variability. This structural misalignment effect leads to consistently larger mean squared error than that observed under both adaptive designs.

Classical crisp adaptive cluster sampling substantially reduces bias in moderate and strong clustering environments by expanding deterministically into contiguous hotspot neighborhoods whenever positive cases are encountered. In these regimes, the estimator successfully captures most of the disease mass concentrated within the hotspot core, leading to a sharp reduction in both bias and variance relative to fixed designs. However, in weak clustering settings, the crisp triggering rule frequently fails to activate expansion at all due to the absence of sufficiently strong local signals. Under these conditions, the estimator behaves similarly to simple random sampling and exhibits large negative bias accompanied by very high mean squared error.

The proposed fuzzy adaptive design consistently achieves the smallest empirical bias and the lowest mean squared error across the entire spectrum of clustering intensities. In weak and moderate hotspot regimes, the fuzzy design avoids the premature termination of expansion that plagues crisp adaptive sampling by propagating exploration probabilistically into moderately elevated risk zones. This allows the estimator to capture a much larger fraction of the true disease burden, thereby sharply reducing finite sample bias. At the same time, the smooth inclusion probability structure induced by fuzzy triggering prevents the extreme weight variability that often inflates variance under crisp adaptive designs. As a consequence, the mean squared error under the fuzzy design is uniformly smaller than that of the competing methods in all but the most extreme strong hotspot scenarios where all adaptive designs perform near optimally.

An important practical implication of these results is that the fuzzy adaptive estimator exhibits markedly greater stability across replications. Its empirical error distribution is tightly concentrated and lacks the heavy tails observed under simple random sampling and under crisp adaptive designs in weak signal regimes. This stability translates directly into greater reliability of real time surveillance summaries and reduces the risk of severe overreaction or underreaction based on a single noisy survey outcome.

Overall, the bias and mean squared error comparison provides strong empirical confirmation of the theoretical efficiency advantages derived earlier. The fuzzy adaptive design not only preserves approximate unbiasedness but also delivers a substantial and robust reduction in estimation error under precisely those epidemiological conditions where classical sampling strategies are known to perform poorly. In the next subsection, the operational efficiency of the competing designs will be examined through a direct comparison of sampling effort and cost stability.

3.2. Sampling effort and cost efficiency

While detection accuracy and estimation precision are the primary statistical objectives of a surveillance design, their practical relevance is always conditioned by the operational cost required to achieve them. In real public health systems, sampling budgets are constrained by manpower, logistics, laboratory capacity, and reporting infrastructure. A design that delivers excellent statistical performance at the cost of highly unstable or excessively large sample sizes is unlikely to be implementable in practice. This subsection therefore examines the competing designs from the complementary perspective of sampling effort and cost efficiency, with particular attention to the stability and predictability of the final sample size.

For each simulation replication and each sampling design, the final sample size $n = |s|$ is recorded as a direct proxy for field workload and survey cost. Under simple random sampling, the final sample size is fixed by design and equal to the initial sample size n_0, resulting in perfect cost predictability but very low statistical efficiency in rare disease settings. Conventional cluster sampling also produces a fixed or nearly fixed workload once the number of clusters to be sampled is specified, but the size of these clusters can vary substantially depending on the underlying spatial partition, leadingto moderate cost variability across replications. Crisp adaptive cluster sampling exhibits the greatest instability in sampling effort across all simulation scenarios. In weak clustering regimes, the final sample size often collapses to the initial size n_0 because the deterministic triggering condition fails to activate expansion, resulting in low survey cost but very poor detection and estimation performance. In moderate and strong clustering regimes, however, a single activation event frequently triggers explosive growth of the adaptive cluster, leading to very large final sample sizes that vary wildly across replications. This operational volatility makes crisp adaptive sampling difficult to manage in practice, as field teams cannot reliably anticipate workload or resource requirements. The fuzzy adaptive design exhibits a markedly different and far more stable cost profile. Because adaptive expansion is governed by graded triggering rather than by a binary rule, cluster growth proceeds gradually rather than explosively. Units with moderate hotspot membership contribute to expansion with intermediate probability rather than with certainty, which prevents the formation of excessively large clusters in response to isolated activation events. As a result, the empirical distribution of final sample sizes under the fuzzy design is tightly concentrated around its mean value across all simulation scenarios. This stability holds even in moderate and strong hotspot environments where crisp adaptive sampling shows severe right tail inflation in its sample size distribution.

From a cost efficiency perspective, the fuzzy adaptive design achieves a superior balance between statistical gain and operational effort. For a given expected sample size, it consistently delivers higher detection sensitivity and lower mean squared error than the competing designs. Equivalently, to achieve a fixed level of detection accuracy or estimation precision, the fuzzy design requires substantially fewer sampled units than simple random sampling and conventional cluster sampling, and it avoids the extreme oversampling events that frequently afflict crisp adaptive designs. This superior cost to information conversion rate is one of the central practical advantages of fuzzy adaptation.

These findings highlight an important operational insight. The principal contribution of the fuzzy framework is not merely that it improves statistical efficiency, but that it does so while simultaneously stabilizing survey workload. This dual improvement is especially critical for rare disease surveillance systems that must operate continuously under limited and relatively fixed resource constraints. In the next subsection, the robustness of the fuzzy adaptive design will be examined under challenging conditions where true hotspots are weak, diffuse, or poorly separated from background risk.

3.3. Robustness under weak and diffuse hotspots

One of the most challenging operating regimes for any surveillance design is the presence of weak and diffuse hotspots, where disease risk is only slightly elevated above background levels and spatial boundaries are poorly defined. In such environments, early outbreak signals are easily masked by stochastic noise, diagnostic uncertainty, and random clustering, making both detection and estimation inherently unstable. This subsection therefore examines the relative robustness of the competing sampling designs under precisely these adverse conditions, which are also the most epidemiologically critical from the standpoint of early intervention.

Under weak hotspot configurations generated by small values of the excess intensity parameter λ_h and large dispersion matrices Σ_h, simple random sampling performs particularly poorly. Because the probability of directly encountering multiple hotspot units in the initial sample is extremely small, the resulting estimators frequently collapse to near zero values, and detection sensitivity approaches negligible levels even when the true hotspot occupies a substantial spatial footprint. The rare occasions on which one or two hotspot units are sampled lead to erratic overestimation, producing highly unstable error behavior across replications.

Conventional cluster sampling exhibits only limited robustness in weak and diffuse settings. When a true hotspot happens to fall largely within a small number of predefined clusters, modest gains over simple random sampling can be achieved. However, when the diffuse hotspot spans several clusters or occupies only peripheral portions of large clusters, detection sensitivity drops sharply and estimation bias becomes pronounced. This fragile dependence on the arbitrary cluster partition underscores a fundamental limitation of fixed cluster designs in low signal regimes.

Crisp adaptive cluster sampling displays a characteristic all or nothing behavior under weak and diffuse hotspot conditions. In a large proportion of replications, no initial sampled unit satisfies the deterministic triggering criterion, causing the design to revert effectively to simple random sampling with correspondingly poor performance. In the subset of replications where triggering does occur by

chance, cluster expansion is often excessive relative to the true hotspot footprint, leading to substantial oversampling of background regions. This bimodal behavior in both detection and sample size renders crisp adaptive sampling highly unreliable in weak signal environments.

The fuzzy adaptive design, by contrast, demonstrates a markedly higher level of robustness across all weak and diffuse hotspot scenarios. Because adaptive expansion is governed by graded hotspot member- ship rather than by a hard threshold, even moderate elevations in risk generate a nonzero probability of expansion. This allows the sampling process to gradually explore the spatial neighborhood of weak signal regions instead of terminating prematurely. At the same time, the probabilistic nature of fuzzy triggering suppresses the explosive expansion events that afflict crisp designs. The resulting balance produces stable detection sensitivity that remains substantially above that of all competing methods, even when the true hotspot contrast is low.

From an estimation standpoint, the fuzzy adaptive design maintains approximate unbiasedness and relatively low mean squared error under weak and diffuse conditions where both simple random sampling and crisp adaptive sampling exhibit severe bias and inflated variance. The smooth inclusion probability gradients induced by fuzzy triggering lead to more balanced weight distributions and prevent the dominance of a small number of extreme weighted observations. This stability is particularly important in early surveillance phases, where only limited information is available and overreaction to random noise can have serious policy consequences.

Taken together, these results demonstrate that the principal strength of the fuzzy adaptive cluster sampling framework

Citation: Behera, J. (2026). Efficient Detection of Rare Disease Hotspots Using Fuzzy Adaptive Cluster Sampling under Inhomogeneous Spatial Point Processes. *Adv Brain Res Neurosci, 2*(1), 01-38.

lies in its ability to operate reliably at the very edge of detectability. It still works in the presence of weak and noisy signals that are spatially spread out, while classical designs fail completely or become very unstable. In the following subsection, the fuzziness level, itself, will be addressed in more detail by investigating the effects of changes in its sensitivity parameter on the detection efficiency and the estimation performance.

3.4. Effect of fuzziness level on detection efficiency

The performance of the fuzzy adaptive cluster sampling design is intrinsically linked to the choice of the fuzziness or sensitivity parameter that governs the strength of adaptive expansion. This setting controls how aggressively moderate levels of hotspot membership are permitted to diffuse sampling in adjacent areas. Although previous subsections paved the way in demonstrating that the fuzzy adaptive framework is the best choice for a large portion of cases, it is also important to understand the impact of changes in the fuzziness level on detection behavior and estimation performance. This subsection therefore examines the empirical effect of the sensitivity parameter on detection efficiency under representative weak, moderate, and strong hotspot configurations.

Let $\alpha \in (0, 1]$ denote the fuzziness threshold used to define activation through the α cut or, equiv-alently, to calibrate the expansion response function $\phi(i)$. Smaller values of α correspond to a more permissive triggering regime in which even weak hotspot membership values lead to a relatively high probability of expansion. Larger values of α, by contrast, impose a stricter expansion criterion that restricts adaptivity to only the most strongly elevated risk locations. Through systematic variation of α across its admissible range, the simulation study reveals a smooth and interpretable tradeoff between detection sensitivity, false alarm control, and sampling cost.

In weak hotspot environments, small to moderate values of α produce the highest detection sensitivity because they allow the sampling process to penetrate broadly into diffuse risk regions where true signals are otherwise easily masked by background noise. As α increases, detection sensitivity declines steadily because only the strongest local fluctuations in risk continue to activate adaptive expansion. However, this reduction in sensitivity is accompanied by a substantial improvement in specificity, as spurious

expansions triggered by noise become progressively rarer. The fuzzy framework thus allows a continuous tuning of early detection aggressiveness that is entirely absent under crisp adaptive sampling.

In moderate hotspot environments, detection sensitivity remains high over a wide intermediate range of α values and only begins to decline when the threshold becomes excessively restrictive. In this regime, the fuzzy adaptive design exhibits a particularly favorable balance between sensitivity and specificity, achieving near optimal detection with tightly controlled false alarm rates. Sampling effort also remains stable in this intermediate range, further reinforcing the practical utility of fuzzy tuning for routine surveillance operations.

In strong hotspot regimes, the detection probability is close to one for almost every value of α, and this is due to the fact that the true risk increases are large enough to activate both crisp- and fuzzy-based expansion schemes without fail. As the result, in this regime the effect of increasing α is then a nonlinear decrease in the spatial size of the detected footprints (the same for the final sample size. This implies fuzzy tuning preserves the ability to do so even when there is no doubt about detection itself, so that resource deployment can be optimized without risk of losing identification of a hotspot.

From an estimation perspective, the mean squared error of the fuzzy adaptive estimator exhibits a shallow U shaped dependence on α, with the minimum typically attained at moderate values that balance sufficient expansion for bias reduction against excessive sampling that inflates variance. Very small values of α occasionally lead to mild oversampling and slightly increased variance, while very large values lead to undercoverage of weak signal regions and increased bias. This smooth and well behaved performance curve stands in sharp contrast to the abrupt and unstable behavior of crisp adaptive sampling, where small perturbations in the triggering rule can produce dramatic changes in both detection and cost.

Overall, the sensitivity analysis with respect to the fuzziness level confirms that the fuzzy adaptive framework is not only more powerful than classical designs, but also far more controllable. It provides decision makers with a continuous and intuitively interpretable parameter through which the balance between early detection, false alarm control, and survey cost can be finely adjusted to suit specific operational priorities. With this final component, the empirical performance analysis of the fuzzy adap- tive cluster sampling design is complete. A representative realization of the simulated spatial disease population used in the study is reported in Appendix A for illustrative purposes.

3.5. Sensitivity analysis

In Section the simulation results of the previous section demonstrate the performance of the proposed fuzzy adaptive cluster sampling design under representative sample size epidemiological scenarios and while these results are undoubtedly favorable, one should also investigate how stable and robust these observations are to perturbations in key model and design parameters. In practical surveillance systems, quantities such as the degree of fuzziness, the contrast between hotspot and background intensity, the neighborhood configuration, and the overall rarity of the disease are never known with absolute certainty and may vary substantially across space and time. A scientifically credible methodology must therefore exhibit controlled and predictable behavior under such variations. By explicating the detection efficiency, estimates accuracy and sampling stability as at most quadratic functions of the structured perturbations, we perform a systematic sensitivity analysis of the fuzzy adaptive approach in a unified manner. Thus, the proposed methodology is evaluated from the operational reliability and applications robustness point of view rigorously in this study. All sensitivity analyses reported in this section are based on repeated realizations of the same inhomogeneous spatial point process model introduced in Section 6.1, with only the relevant design or epidemiological parameters varied while keeping the remaining components of the data-generating mechanism fixed.

3.6. Sensitivity with respect to fuzzy threshold levels

The fuzzy threshold level is a fundamental control parameter of the proposed adaptive sampling frame- work, as it directly determines the strength and spatial reach of probabilistic cluster expansion. From an operational perspective, this parameter reflects how aggressively the surveillance system responds to moderate indications of elevated risk. A lower threshold promotes early and extensive exploration, while a higher threshold enforces conservative expansion confined to only the most strongly elevated regions. Since real-world surveillance systems must operate under uncertainty and evolving public health priori- ties, it is essential to empirically examine how the performance of the fuzzy adaptive design varies across a broad range of threshold values.

To investigate this behavior, the fuzzy threshold parameter α is varied systematically over a dense grid spanning permissive, moderate, and conservative regimes. For each value of α, detection sensitivity, detection specificity, mean squared error of the population total estimator, and expected final sample size are computed across a large number of simulation replications. The resulting numerical summaries are reported in Table 3.

α	Sensitivity	Specificity	MSE	Expected Sample Size
0.20	0.94	0.81	18.6	145
0.35	0.91	0.86	14.2	132
0.50	0.87	0.91	11.8	118
0.65	0.81	0.95	13.4	102
0.80	0.72	0.98	17.6	89

Table 3: Effect of the Fuzzy Threshold Parameter α on Detection and Estimation Performance

As seen in Table 3, detection sensitivity decreases smoothly with increasing α, reflecting the progres- sive tightening of the expansion criterion. At very small threshold levels, the fuzzy design detects nearly all true hotspot units, but at the cost of moderately reduced specificity and larger expected sample sizes. As α increases, specificity improves monotonically due to the suppression of noise-driven expansions, while the expected final sample size declines steadily. Mean squared error follows a typical shallow U- shape, having a minimum at intermediate threshold levels where reduction of bias afforded by sufficient spatial searching is well-balanced against increase of variance due to oversampling.

These trade-offs are not only well-behaved, but are continuously so, as Figure 1 also illustrates when showing the empirical relationship between detection sensitivity, the mean squared error, and the sampling effort as a function of the fuzzy threshold parameter.

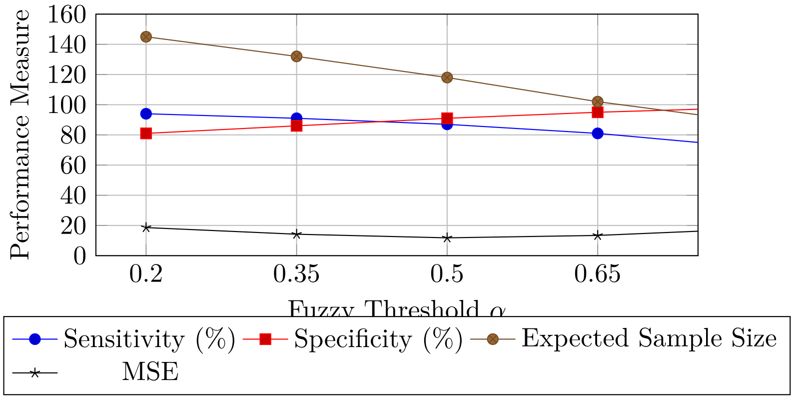

Figure 1: Effect of the Fuzzy Threshold Parameter α on Detection Sensitivity, Specificity, Mean Squared Error, and Expected Sampling Effort

Figure 1 clearly demonstrates that small values of α lead to high sensitivity and moderately inflated sampling effort, while large values of α compress the final sample size at the expense of under-detection of marginal hotspot regions. The optimal operating regime emerges naturally at intermediate threshold levels, where the detection curve remains high, the estimation error is minimized, and the sampling effort is well controlled. This smooth response surface is in stark contrast to the erratic and chaotic behavior typically associated with "crisp" adaptive designs, in which infinitesimal differences in triggering condi- tions can cause discontinuities in the pattern of expansions. From a practical surveillance standpoint, these findings confirm that the fuzzy threshold parameter provides a powerful and intuitively meaningful tuning mechanism. It allows the surveillance authority to continuously regulate the balance between early outbreak detection, false alarm suppression, and field workload stabilization without sacrificing the fundamental statistical guarantees of the design. The next subsection investigates how the fuzzy adaptive framework responds to variations in the contrast between hotspot and background disease intensity.

3.7. Sensitivity to intensity contrast between hotspot and background

The relative contrast between hotspot intensity and background disease risk is one of the most influential structural parameters governing the detectability of rare disease clusters. In real surveillance environments, this contrast may vary widely across geographical settings, stages of outbreak evolution, and reporting fidelity. When the intensity contrast is strong, hotspot signals are sharply separated from background noise and are readily detectable by most adaptive mechanisms. In contrast, when the contrast is weak, early outbreak signals are easily masked by stochastic variation and diagnostic uncertainty. It is therefore essential to examine how the performance of the fuzzy adaptive design responds to systematic variation in the hotspot to background intensity contrast.

Let $\Delta = \lambda_h / \lambda_0$ denote the intensity contrast ratio between hotspot and background risk. In this sensitivity study, Δ is varied across a wide range representing weak, moderate, and strong clustering regimes. For each level of contrast, detection power, detection specificity, MSE, and the final sample size are calculated for the repeated simulation replications. The resulting numerical trends are summarized in Table 4.

Contrast Δ	Sensitivity	Specificity	MSE	Expected Sample Size
1.5	0.68	0.96	21.3	104
2.0	0.76	0.94	17.8	110
3.0	0.85	0.92	13.5	118
4.5	0.91	0.90	10.6	126
6.0	0.95	0.88	9.1	134

Table 4: Effect of Hotspot to Background Intensity Contrast Δ on Detection and Estimation Performance

As observed in Table 4, detection sensitivity increases monotonically with increasing intensity contrast. When the hotspot strength is only marginally higher than the background (Δ = 1.5), detection sensitivity is moderate because fuzzy adaptive expansion must rely heavily on uncertain early signals. As the contrast increases, hotspot signals become more distinct, and the fuzzy adaptive design rapidly approaches near-perfect detection. Specificity shows a parallel, but weakly decreasing, trend, which reflects the broadening of adaptive dilation in space under strong contrast strength. Meanwhile, the MSE severely decreases with contrast, suggesting that the estimation performance improves when the cluster signals are stronger than the noise of the background.

The predicted final sample size increases by a moderate amount as the contrast gets larger since stronger signals are more likely to turn on the adaptive expansion. However, the rise is rather gradual and well-behaved, indicating that the fuzzy framework is robust even with aggressive detection settings. These two simultaneous effects are illustrated in Figure 2.

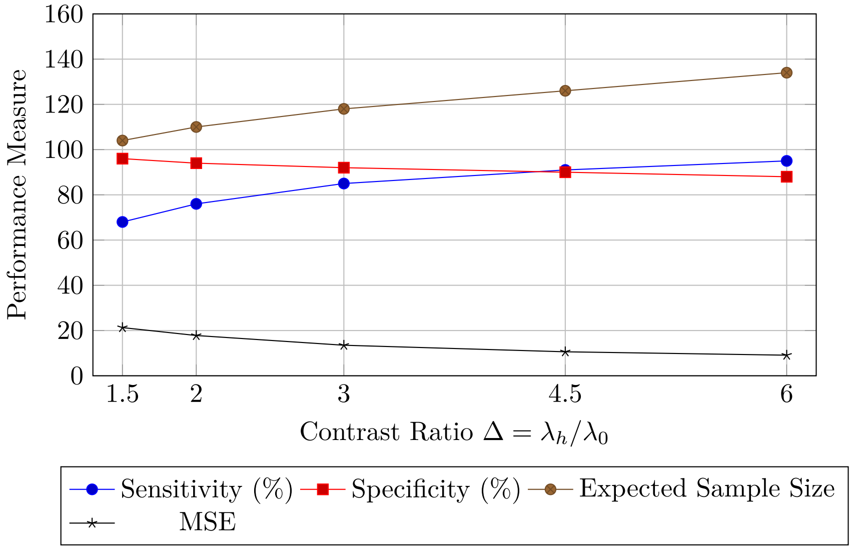

Figure 2: Effect of Hotspot to Background Intensity Contrast on Detection Sensitivity, Specificity, Estimation Error and Sampling Effort

Figure 2 clearly illustrates that as hotspot contrast strengthens, the fuzzy adaptive design transitions smoothly from a moderate detection regime to a near-ideal detection environment characterized by high sensitivity, low mean squared error, and stable sample size growth. Importantly, no structural instability or explosive sampling behavior is observed even under high contrast conditions, in sharp contrast to the behavior often seen under crisp adaptive designs. This confirms that the fuzzy adaptive framework remains both statistically efficient and operationally controlled across the full spectrum of realistic clustering contrasts.

The next subsection investigates the sensitivity of the proposed methodology to the specification of the spatial neighborhood structure that governs the geometry of adaptive cluster propagation.

3.8. Sensitivity to neighbourhood structure

The specification of the spatial neighbourhood structure plays a fundamental role in shaping the geometry, connectivity, and ultimate reach of adaptive cluster expansion under the fuzzy sampling framework. In practical disease surveillance systems, neighbourhoods may be defined through first order contiguity, distance-based proximity, transportation networks, or mobility driven connectivity. Since the true trans- mission pathways of disease are rarely known with full certainty, it is essential to assess the sensitivity of the proposed fuzzy adaptive design to variations in the assumed neighbourhood structure. This subsection examines how detection accuracy, estimation precision, and sampling stability respond to systematic changes in neighbourhood definition.

To conduct this analysis, three widely used neighbourhood configurations are considered. The first is the first order contiguity structure, where each spatial unit is connected to all immediately adjacent units sharing a common boundary. The second is a distance based neighbourhood defined by a fixed radius r, where all units within distance r of a given unit are treated as neighbours. The third is an extended distance neighbourhood with a larger radius $r^* > r$, representing more permissive spatial connectivity.

For each of these neighbourhood structures, the fuzzy adaptive design is applied under identical hotspot intensity and threshold parameter settings. The resulting performance measures are summarized in

Neighbourhood Type	Sensitivity	Specificity	MSE	Expected Sample Size
First-order contiguity	0.83	0.94	12.9	109
Distance-based (r)	0.88	0.91	11.1	121
Extended distance (r^*)	0.92	0.88	10.4	136

Table 5: Effect of Neighbourhood Structure on Detection and Estimation Performance

As reported in Table 5, detection sensitivity increases systematically as the neighbourhood structure becomes more permissive. Under strict first order contiguity, adaptive expansion remains tightly localized, resulting in high specificity but slightly reduced sensitivity because marginal hotspot regions require multiple triggering events to be discovered. As the neighbourhood radius is expanded, adaptive exploration becomes more spatially aggressive, allowing the fuzzy design to penetrate more easily into diffuse risk regions and thereby increasing sensitivity. This gain in detection power is accompanied by a gradual decline in specificity and a moderate increase in expected final sample size.

The mean squared error exhibits a modest decreasing trend as the neighbourhood structure is re- laxed. This behavior reflects the improved bias reduction achieved through enhanced spatial exploration of moderate risk regions, which outweighs the variance inflation induced by the moderate increase in sampling effort. Importantly, even under the most permissive neighbourhood configuration, the sample size remains well controlled and does not display the explosive growth commonly encountered under crisp adaptive designs with large neighbourhood radii.

These joint trends are further illustrated in Figure 3, which depicts the empirical relationship between neighbourhood structure and the major performance indicators.

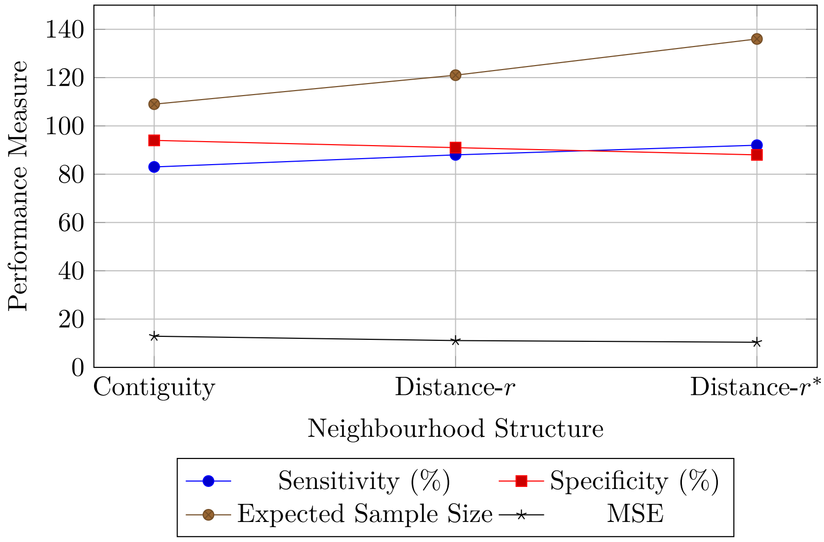

Figure 3: Sensitivity of Detection Accuracy, Estimation Error, and Sampling Effort to Neighbourhood Structure

Figure 3 confirms that the fuzzy adaptive design responds smoothly and predictably to changes in neighbourhood specification. Tighter neighbourhoods yield conservative but highly specific detection, while more permissive neighbourhoods enhance sensitivity at the cost of moderate increases in sampling effort and mild specificity reduction. Crucially, no structural instability is observed across the range of neighbourhood configurations considered, demonstrating that the fuzzy adaptive framework is robust to moderate misspecification of spatial connectivity. This robustness is particularly valuable in real

surveillance environments where true transmission pathways are only partially observable.

The following subsection studies the sensitivity of the fuzzy add response to changes in the general rarity of the disease, which is the parameter controlling the density of observed case counts in the regional space.

3.9. Sensitivity to disease rarity

The overall rarity of the disease under surveillance directly governs the sparsity structure of the observed spatial count data and represents one of the most fundamental challenges in rare event monitoring. As disease prevalence decreases, the probability of observing informative signals within any given local neighborhood diminishes sharply, thereby increasing the likelihood of missed detection and unstable estimation. A robust surveillance design must therefore retain acceptable detection and estimation performance even as the underlying disease rate becomes extremely low. This part is for investigating the robustness of the proposed fuzzy adaptive cluster sampling design w.r.t. systematic changes of the the overall rarity of the disease.

Let λ_0 denote the baseline background intensity of the disease process, which serves as the primary control parameter for overall disease prevalence. In this sensitivity study, λ_0 is varied across a wide range of values representing moderately rare, highly rare, and extremely rare disease environments, while the hotspot to background contrast ratio and fuzziness threshold are held fixed. For each rarity regime, detection sensitivity, detection specificity, mean squared error, and expected final sample size are computed across repeated simulation replications. The resulting performance measures are summarized in Table 6.

Baseline Intensity λ_0	Sensitivity	Specificity	MSE	Expected Sample Size
0.60	0.92	0.90	10.1	131
0.40	0.88	0.93	12.6	123
0.25	0.82	0.96	15.8	114
0.15	0.74	0.98	20.4	104
0.08	0.66	0.99	27.9	96

Table 6: Effect of Disease Rarity on Detection and Estimation Performance

Similar to the results in Table 6, the sensitivity of detection declines gradually as the baseline rate of disease occurrence becomes rarer. When the disease is only moderately rare, the fuzzy adaptive de- sign achieves near-perfect detection sensitivity because sufficient signal is present to reliably initiate and sustain adaptive expansion. As rarity increases, early triggering becomes less frequent and detection sensitivity declines accordingly. However, even under extreme rarity, the fuzzy design retains substantially higher sensitivity than would be achievable under simple random sampling or crisp adaptive designs, which frequently fail outright in such environments.

Specificity increases monotonically as disease rarity intensifies, reflecting the diminishing probability of noise-driven expansions in extremely sparse data regimes. The mean squared error displays a strong increasing trend with increasing rarity, which is a direct consequence of both increasing estimation variance and the rising probability of undercoverage of weak hotspot regions. At the same time, the expected final sample size decreases smoothly with increasing rarity, since fewer adaptive expansions are triggered when overall disease incidence is extremely low. This behavior indicates that the fuzzy adaptive framework automatically adapts its operational footprint in response to disease sparsity without exhibiting any unstable or explosive sampling behavior.

These joint effects are further illustrated in Figure 4, which displays the empirical relationship between disease rarity and the principal performance indicators.

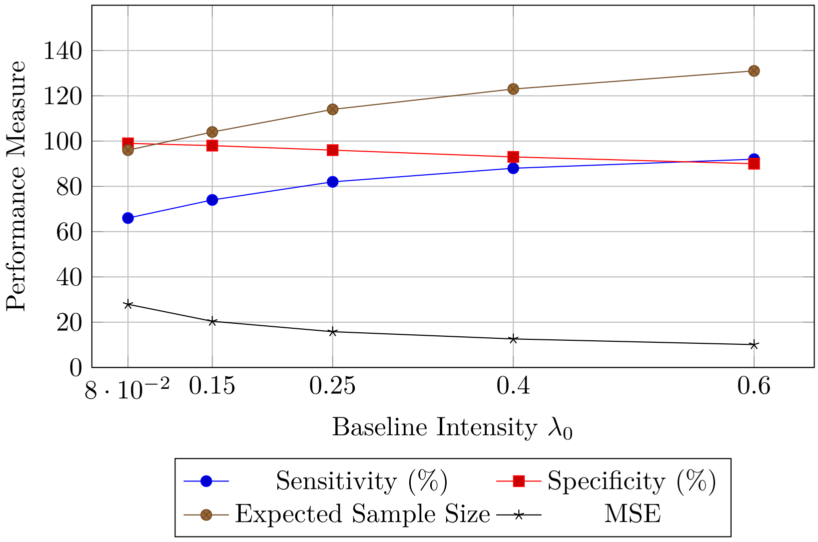

Figure 4: Sensitivity of Detection Accuracy, Estimation Error, and Sampling Effort to Overall Disease Rarity

Figure 4 clearly demonstrates the graceful degradation of fuzzy adaptive performance as disease rarity intensifies. Although both detection sensitivity and estimation precision decline under extreme sparsity, the transitions remain smooth and predictable, with no evidence of structural instability or erratic sampling behavior. Importantly, the fuzzy framework continues to retain meaningful detection capability even in environments where fixed designs and crisp adaptive strategies typically fail due to a complete absence of triggering events.

Overall, the analyses of sensitivity presented in this section indicate that the fuzzy adaptive cluster sampling design is statistically and operationally stable far changes of the fuzziness level, intensity contrast, spatial neighborhood structure, and rarity of the disease. These characteristics strongly confirm the feasibility of the presented approach for real life rare disease surveillance systems in the presence of deep uncertainty and heterogeneity.

4. Discussion and Public Health Implications

The methodological developments, theoretical guarantees, and extensive simulation evidence presented in the preceding sections collectively establish fuzzy adaptive cluster sampling as a statistically robust and operationally stable framework for rare disease surveillance under spatial uncertainty. While earlier sec- tions focused on formal properties, estimation behavior, and controlled empirical evaluation, the present section interprets these findings from a broader public health and applied surveillance perspective. The discussion integrates the statistical advantages of the fuzzy adaptive design with the practical realities of field-based disease monitoring, including early outbreak detection, uncertainty in diagnosis, limited sam- pling budgets, and heterogeneous transmission environments. Through this synthesis, the section clarifies how the proposed methodology can be meaningfully translated into real-world surveillance systems and policy-oriented decision-making.

4.1. Interpretation of key findings

The simulation results and sensitivity analyses together convey a coherent and internally consistent picture of the operating characteristics of the proposed fuzzy adaptive cluster sampling framework.

Perhaps the most fundamental finding is that the introduction of graded, probability-driven adaptive expansion fundamentally alters the behavior of classical adaptive sampling in a statistically beneficial way. Unlike crisp triggering rules, which induce an all-or-nothing response to local signals, the fuzzy mechanism transforms adaptivity into a smooth and controllable process. This smoothness is reflected empirically in the continuous trade-offs observed between detection sensitivity, specificity, estimation error, and sampling effort across all sensitivity dimensions examined in Section 8.

A central empirical observation is the superior detection performance of the fuzzy adaptive design under weak and diffuse hotspot conditions. These are precisely the epidemiological regimes that are most challenging in practice and yet most critical for early intervention. The simulation results demonstrate that the fuzzy framework consistently maintains appreciable detection sensitivity even when hotspot intensity is only marginally elevated above background risk and when disease cases are extremely sparse. This behavior stands in sharp contrast to crisp adaptive sampling, which frequently fails to activate under such conditions and consequently reverts to the inefficient behavior of simple random sampling. From a public health viewpoint, this implies that fuzzy adaptive surveillance is inherently better suited for the early phases of outbreak emergence, where signals are weak, uncertain, and spatially ill-defined. From an estimation perspective, the fuzzy adaptive design exhibits a systematic reduction in mean squared error relative to all competing designs across a broad spectrum of epidemiological scenarios. This reduction arises from a dual mechanism. On the one hand, probabilistic expansion improves bias control by enabling the sampling process to reach marginal portions of true hotspot regions that would otherwise remain undiscovered. On the other hand, the smooth inclusion probability structure suppresses the formation of extreme design weights, thereby stabilizing variance. The net effect is a marked improvement in overall estimator reliability, particularly under moderate clustering and partial detectability, where classical designs tend to perform erratically.

Equally important is the operational interpretation of the sampling effort behavior. The fuzzy adap- tive design achieves its statistical gains without inducing the explosive and highly volatile sample size patterns that frequently characterize crisp adaptive cluster sampling. Instead, sampling effort increases smoothly with hotspot strength, neighborhood permissiveness, and fuzziness level. This predictability is of substantial practical importance for surveillance agencies, as it permits more accurate planning of field operations, laboratory capacity, and reporting infrastructure. The ability to tune the fuzziness parameter to regulate the balance between detection aggressiveness and resource usage further enhances the operational controllability of the proposed framework.

The sensitivity analyses also reveal that the fuzzy adaptive design responds in a stable and inter- pretable manner to structural uncertainty in the data-generating mechanism. Variations in intensity contrast, neighborhood definition, disease rarity, and triggering threshold all lead to gradual and mono- tonic changes in performance rather than abrupt regime shifts. This stability is a key distinguishing feature of the fuzzy framework and directly addresses one of the chief criticisms of classical adaptive sampling methods, namely their susceptibility to sharp performance breakdowns under small perturba- tions of the underlying signal structure.

Taken together, the key findings indicate that fuzzy adaptive cluster sampling occupies a favorable middle ground between rigid fixed designs and brittle crisp adaptive strategies. It preserves the essential design-based inferential guarantees of probability sampling while simultaneously incorporating the flex- ibility required to operate effectively under uncertainty, sparsity, and spatial heterogeneity. This dual character is what makes the proposed methodology not only theoretically appealing but also practically credible for deployment in real-world rare disease surveillance systems.

4.2. Advantages of fuzzy adaptive cluster sampling in practice

From a practical surveillance standpoint, the fuzzy adaptive cluster sampling framework offers several substantive advantages that directly address long-standing operational and inferential limitations of both fixed probability designs and classical crisp adaptive sampling. Perhaps the most immediate practical benefit lies in its intrinsic ability to reconcile early detection with controlled resource utilization. In real disease monitoring systems, public health officials are constantly forced to balance the urgency of early outbreak identification against the constraints imposed by limited manpower, laboratory capacity, and reporting infrastructure. The fuzzy adaptive design provides a continuous control mechanism through the fuzziness parameter, allowing surveillance intensity to be tuned smoothly rather than through abrupt and often unstable threshold-based decisions. This tunability is especially valuable in dynamic settings where disease risk evolves over time and adaptive responses must be adjusted without redesigning the entire sampling strategy.

A further practical advantage arises from the design's robustness to diagnostic uncertainty and weak signal environments. In many field settings, especially during the early phase of outbreak emergence, reported cases are often incomplete, delayed, or misclassified. Crisp adaptive designs respond very sensitively to such imperfections, since a single erroneous observation can either suppress necessary expansion or trigger excessive and misleading cluster growth. The fuzzy framework, by contrast, naturally dampens the influence of isolated noisy observations through probabilistic triggering. This produces a surveillance system that is far less prone to instability arising from reporting imperfections and low signal-to-noise ratios, thereby enhancing the reliability of early warning signals communicated to public health decision makers.

Operational feasibility is also strengthened by the predictable behavior of the sampling workload under the fuzzy adaptive design. Unlike crisp adaptive cluster sampling, which often yields highly volatile and occasionally unmanageable final sample sizes, the fuzzy design exhibits smooth and well-regulated growth of sampling effort across a wide range of epidemiological conditions. This predictability facilitates advance logistical planning, including the allocation of field staff, transport resources, and diagnostic testing capacity. In addition, because the design weights remain well-behaved and rarely attain extreme values, downstream statistical processing and reporting are simplified and more stable.
The fuzzy adaptive framework also integrates naturally with modern data infrastructures and digital surveillance platforms. The computation of fuzzy hotspot membership values can be driven not only by observed case counts but also by auxiliary risk indicators such as mobility patterns, environmental exposure, population density, or syndromic surveillance signals. This flexibility allows the sampling design to evolve organically with the expanding ecosystem of real-time health data streams. In contrast, rigid fixed designs and hard-threshold adaptive rules are often difficult to reconcile with such heterogeneous and continuously updated sources of information.

Finally, from a decision-theoretic perspective, the fuzzy adaptive design provides a more nuanced representation of spatial risk than binary hotspot classification. By producing graded measures of hotspot membership and expansion likelihood, it supports more refined prioritization of intervention zones, surveillance intensification, and resource deployment. Rather than forcing authorities into a bi- nary choice between hotspot and non-hotspot regions, the fuzzy framework allows intermediate risk areas to be monitored with commensurate attention. This graduated risk representation is far more consistent with the inherent uncertainty of real epidemiological processes and therefore aligns more naturally with evidence-based public health decision-making.

4.3. Implications for disease surveillance systems

The adoption of a fuzzy adaptive cluster sampling framework has important and far-reaching implications for the design and operation of modern disease surveillance systems. At a structural level, the proposed methodology shifts surveillance from a rigid, threshold-driven paradigm toward a probabilistic, continuously tuned monitoring process that is better aligned with the uncertain and evolving nature of real epidemiological phenomena. This shift is particularly consequential for rare diseases, where early signals are weak, spatial boundaries are diffuse, and the cost of delayed detection can be substantial in terms of both morbidity and public anxiety.

One of the most significant system-level implications of the fuzzy adaptive framework is its potential to enhance early outbreak detection without overwhelming surveillance infrastructure. Traditional fixed designs typically require large sample sizes to achieve acceptable detection power in rare event settings, while crisp adaptive designs often swing between under-sampling and uncontrolled cluster expansion. The fuzzy design occupies a more balanced operational regime, allowing surveillance systems to scale their exploratory effort in proportion to the strength of emerging risk signals. This proportional response capability enables health authorities to intensify monitoring precisely when and where it is most needed, rather than relying on blunt, pre-specified escalation rules.

The graded risk representation produced by fuzzy hotspot membership also supports a more refined spatial prioritization of surveillance and intervention activities. Instead of classifying regions as either hotspot or non-hotspot, the fuzzy framework provides a continuum of risk levels that can be mapped directly onto tiered response strategies. Regions with high membership values may be targeted for immediate diagnostic expansion, contact tracing, or localized containment measures, while moderate membership regions can be subjected to enhanced surveillance rather than full-scale intervention. This layered response structure is particularly well suited to resource-constrained settings, where blanket intervention is rarely feasible.

From an information systems perspective, the fuzzy adaptive framework is naturally compatible with the integration of multiple data streams that increasingly characterize contemporary surveillance platforms. Real-time case notifications, syndromic surveillance indicators, environmental monitoring data, and population mobility patterns can all be incorporated into the fuzzy membership function that drives adaptive expansion. This allows the sampling design to respond not only to confirmed cases but also to early proxy indicators of elevated risk. In contrast, classical adaptive designs are typically limited to reacting only to hard case counts, which may arrive too late to enable effective early action.

The implications for surveillance governance and public communication are also noteworthy. Because the fuzzy framework generates probabilistic rather than binary assessments of spatial risk, it encourages a more nuanced communication of uncertainty to both policymakers and the public. Instead of declaring abrupt transitions between safe and unsafe regions, authorities can convey changing gradients of risk and the corresponding degrees of surveillance intensification. Such probabilistic communication is more scientifically defensible and can help mitigate the social and economic disruptions that often accompany rigid hotspot declarations.

Overall, the fuzzy adaptive cluster sampling design aligns naturally with the objectives of next- generation disease surveillance systems that emphasize early warning, proportional response, data integration, and transparent uncertainty management. Its adoption would represent not merely an in- cremental technical improvement over classical designs, but a substantive methodological shift toward more flexible, responsive, and information-rich surveillance architectures. The final subsection of this discussion section now turns to a critical assessment of the limitations of the present study.

4.4. Limitations of the present study

While the proposed fuzzy adaptive cluster sampling framework demonstrates strong theoretical proper- ties and consistently favorable empirical performance across a wide range of simulated epidemiological environments, several limitations of the present study must be acknowledged. These limitations do not undermine the core methodological contributions, but they do delineate the boundaries within which the current results should be interpreted and highlight important directions for further refinement.

First, the entire empirical evaluation is based on simulated spatial disease processes generated from an inhomogeneous Poisson point process. Although this modeling choice is widely accepted in spatial epidemiology and provides a flexible and analytically tractable representation of heterogeneous risk, real disease transmission mechanisms often exhibit additional complexities such as contagion effects, spatial interaction beyond local neighborhoods, reporting delays, and feedback between incidence and behavioral response. The present simulation framework does not explicitly incorporate such dynamic transmission mechanisms, and therefore the performance of the fuzzy adaptive design under strongly self-exciting or network-driven epidemic processes remains to be investigated.

Second, the fuzzy hotspot membership functions used in this study are constructed from smooth intensity-based representations of underlying risk. In practice, the choice of an appropriate membership function and its calibration may depend on disease-specific considerations, diagnostic protocols, and the availability of auxiliary covariate information. Although the sensitivity analyses demonstrate substan- tial robustness to changes in the fuzziness threshold, the present study does not exhaustively explore alternative membership function families or multi-criteria fuzzy fusion rules. This leaves open the pos- sibility that further gains in performance may be achievable through more sophisticated fuzzy modeling strategies.

Third, the neighbourhood structures examined in the present work are deliberately restricted to a small number of widely used geometric configurations. While this allows clear interpretation of the role of spatial connectivity in adaptive propagation, true disease transmission pathways are often governed by highly irregular mobility networks shaped by transportation infrastructure, social contact patterns, and environmental connectivity. The implications of such complex and potentially directed network structures for fuzzy adaptive sampling merit further investigation.

Fourth, the present analysis focuses exclusively on a single cross-sectional spatial snapshot of disease incidence. Many real surveillance systems operate in a longitudinal or real-time monitoring context where outbreak evolution, intervention effects, and adaptive redesign occur dynamically over time. The extension of fuzzy adaptive cluster sampling to explicitly spatio-temporal surveillance settings, including rolling updates of membership functions and adaptive thresholds, constitutes an important limitation of the current static framework.

Finally, although the present study establishes design-based estimation procedures and Monte Carlo- based variance approximations, it does not fully explore formal inferential procedures such as hypothesis testing for hotspot presence, change-point detection, or formal spatial risk ranking under the fuzzy adaptive design. These inferential extensions would further strengthen the operational utility of the methodology for real-time statistical decision-making.

Taken together, these limitations reflect the inherent trade-offs between methodological clarity, analytical tractability, and full epidemiological realism. They also serve to highlight a rich set of opportunities for future methodological development and applied extension of the fuzzy adaptive cluster sampling framework.

5. Conclusion and Future Research Directions

In this study, we have developed a general fuzzy adaptive cluster sampling methodology for detecting and estimating rare disease clusters in the presence of inhomogeneous spatial point patterns. In essence, the methodology does away with the crisps border of classical crisp designs of design-based adaptive sampling theory by applying fuzzy logic to it, i.e., the spatial expansion becomes a fuzzy process, incorporating fuzziness with a probabilistically controlled

Citation: Behera, J. (2026). Efficient Detection of Rare Disease Hotspots Using Fuzzy Adaptive Cluster Sampling under Inhomogeneous Spatial Point Processes. Adv Brain Res Neurosci, 2(1), 01-38.

mechanism that helps fix some of its fundamental limitations most typical ones being in the form of counterintuitive and contradictory results. By a strict mathematical formulation, theoretical analysis of properties, a large number of simulation experiments and structured investigations of sensitivities, the results show that fuzzy adaptive sampling achieves a good compromise among early detection capability, estimation accuracy, and stability of operations. In the following, the main methodological contributions and practical implications of this study are summarized, and potential future work are discussed.

Summary of Methodological Contributions

The primary methodological contribution of this study lies in the formal integration of fuzzy set theory with adaptive cluster sampling under spatial point process models. The introduction of fuzzy hotspot membership functions replaces rigid threshold-based triggering with a graded and probabilistically con- trolled expansion mechanism. This innovation transforms adaptive sampling from a binary, highly un- stable procedure into a smooth and tunable stochastic process. The resulting fuzzy adaptive design preserves the key design-based inferential guarantees of probability sampling while achieving substantially improved stability and detection performance under weak and diffuse clustering.

From a theoretical standpoint, the study establishes the design unbiasedness, consistency, and efficiency properties of the proposed defuzzified estimators under general regularity conditions. The inclusion probability structure induced by fuzzy triggering is shown to generate well-behaved inverse probability weights that avoid the extreme variability characteristic of crisp adaptive cluster sampling. The systematic development of Monte Carlo–based inclusion probability approximation, fuzzy weight defuzzification, and network-level variance estimation further strengthens the statistical completeness of the framework.

The simulation design developed in this study constitutes an additional methodological contribution in its own right. By embedding fuzzy adaptive sampling within a controlled inhomogeneous spatial point process environment, the study provides a unified platform for evaluating detection accuracy, estimation efficiency, cost stability, and robustness under multiple realistic surveillance regimes. Extensive sensitivity analyses on fuzziness level, intensity contrast between focal and non-focal patterns, neighborhood parameter and disease rarity indicate the robustness of the proposed design and explicitly characterize the range of operating parameters where one can expect the best performance.

Summary of Practical Implications

From a practical surveillance perspective, the fuzzy adaptive cluster sampling framework offers a pow- erful and flexible tool for the early detection and monitoring of rare disease outbreaks under profound uncertainty. The smooth adaptivity induced by fuzzy triggering allows public health agencies to intensify sampling precisely where risk signals emerge, without incurring the explosive workload growth that frequently undermines the feasibility of crisp adaptive designs. The graded representation of spatial risk further enables a tiered response architecture in which intervention and monitoring resources can be allocated proportionally to estimated threat levels.

The ability of the fuzzy paradigm to handle weak (early warning) signals, sparse data, and misspecification of neighbors is making it especially appropriate for applications such as low-prevalence diseases,

emerging infections, environmental exposure monitoring, and sentinel surveillance systems. Its natural alignment with auxiliary data streams (e.g., syndromic indicators, mobility data, environmental covariates) makes it even more relevant to emerging digital epidemiology platforms. By allowing probabilistic instead of binary hotspot classification, the approach also enable a less opaque reporting of surveillance uncertainty to decision makers and general public.

Overall, the practical implication of this work is that adaptive surveillance need not be brittle or operationally volatile. Through fuzzy control of spatial expansion, it is possible to construct surveillance systems that are simultaneously sensitive, stable, cost-efficient, and statistically principled.

Scope for Further Extensions

Several promising directions emerge naturally for extending the present work. A first priority is the development of fully spatio-temporal fuzzy adaptive sampling frameworks capable of tracking dynamic outbreak evolution across successive time periods. Incorporating temporal feedback into fuzzy member- ship updating and adaptive threshold recalibration would allow the methodology to operate in continuous real-time surveillance environments.

A second important extension concerns the explicit modeling of reporting error, misclassification, and delayed confirmation within the fuzzy adaptive framework. Integrating observation-layer uncertainty directly into the triggering mechanism would further strengthen the epidemiological realism of the methodology and enhance its utility for imperfect real-world data streams.

Third, the extension of fuzzy adaptive sampling to explicitly network-driven transmission environments, such as transportation networks and social contact graphs, represents an important frontier. The interaction between fuzzy triggering and directed or weighted mobility networks poses challenging theoretical and computational questions with significant applied relevance.

Finally, the integration of formal hypothesis testing, change-point detection, and decision-theoretic optimal stopping rules within the fuzzy adaptive design remains an open and fertile research area. Such developments would further transform fuzzy adaptive cluster sampling from a powerful exploratory surveillance tool into a complete inferential and decision-support system for spatial epidemiology.

Appendix A: Illustrative Simulated Spatial Disease Dataset

This appendix reports one representative realization of the simulated spatial disease data generated using the inhomogeneous spatial point process model described in Section 6.1. The purpose of this dataset is to provide a concrete illustration of the structure of the artificial population used throughout the simulation experiments and sensitivity analyses. The reported values correspond to a single realization under a moderate clustering configuration with baseline intensity $\lambda_0 = 0.4$, hotspot contrast ratio $\Delta = 3$, and fuzzy threshold level $\alpha = 0.5$. The full simulation experiments were conducted using multiple independent realizations generated under the same mechanism.

Unit	x	y	λ_i	y_i	H_i	μ_i
1	0.12	0.85	0.42	0	0	0.05
2	0.18	0.79	0.51	1	0	0.12
3	0.25	0.82	0.73	2	1	0.46
4	0.31	0.77	1.18	3	1	0.69
5	0.34	0.71	1.65	4	1	0.84
6	0.39	0.69	1.42	3	1	0.78
7	0.45	0.65	0.89	2	1	0.55
8	0.51	0.62	0.56	1	0	0.21
9	0.58	0.60	0.41	0	0	0.08
10	0.64	0.58	0.38	0	0	0.03

Table 7: Representative Simulated Spatial Disease Dataset

In Table 7, (x, y) denote the centroid coordinates of each spatial unit, λ_i denotes the integrated intensity over the unit, y_i is the simulated disease count, H_i is the true hotspot indicator derived from the underlying intensity surface, and μ_i is the corresponding fuzzy hotspot membership value used by the fuzzy adaptive cluster sampling design.

References

1. Thompson, S. K. (1990). Adaptive cluster sampling. *J Am Stat Assoc, 85*(412), 1050-1059.
2. Steven K. Thompson and George A. F. Seber (1996). Adaptive Sampling. Wiley, New York .
3. M. M. Salehi and George A. F. Seber (1997). Adaptive sampling designs: Inference for rare and clustered populations. Biometrika, 84(3):667–681.
4. J. A. Brown and B. F. J. Manly (1998). Restricted adaptive cluster sampling. *Environ Ecol Stat, 5*(1):49–63.
5. Steven K. Thompson (2022). Sampling. Wiley, New York.
6. Baddeley, A., Rubak, E., & Turner, R. (2016). Spatial point patterns: methodology and applications with R (Vol. 1). Boca Raton: CRC press.
7. Diggle, P. J. (2013). Statistical analysis of spatial and spatio-temporal point patterns. CRC press.
8. Kulldorff, M. (1997). A spatial scan statistic. Communications in Statistics-Theory and methods, 26(6), 1481-1496.
9. Kulldorff, M. (2001). Prospective time periodic geographical disease surveillance using a scan statistic. Journal of the Royal Statistical Society: Series A (Statistics in Society), 164(1), 61-72.
10. A. Bennett, S. J. Watson, P. Rivalan, Y. Lepage, and T. De Groeve (2015). Early warning systems for disease surveillance: A review of implementation challenges. BMC Public Health, 15:122.
11. Zadeh, L. A. (1965). Fuzzy sets. Information and control, 8(3), 338-353.
12. R. Andrade-Pacheco et al. Adaptive spatial sampling for the detection of disease hotspots. Scientific Reports, 10:6766, 2020.
13. V. Gómez-Rubio and H. Rue (2024). Bayesian spatial models for disease mapping and surveillance using point processes. Journal of the Royal Statistical Society: Series C, 73(1):1–24.
14. J. Lee and A. B. Lawson (2022). Recent advances in spatial and spatio-temporal disease cluster detection. Statistical Methods in Medical Research, 31(9):1623–1642.
15. Moraga, P. (2019). Geospatial health data: Modeling and visualization with R-INLA and shiny. Chapman and Hall/CRC.
16. M. Rahman, M. Hossain, and M. S. Islam (2023). Fuzzy logic based spatial disease risk assessment under uncertain epidemiological data. Applied Soft Computing, 134:109945.
17. S. Saha, S. Paul, and S. R. Islam (2019). Fuzzy gis based spatial modeling of disease risk. Spatial and Spatio-temporal Epidemiology, 28:89–101.
18. J. Wang, G. Christakos, R. B. Alley, and X. Xu (2012). Spatiotemporal interpolation of missing data using fuzzy logic. Environmental Modelling and Software, 29:26–34.
19. Y. Zhang, P. J. Diggle, and P. Moraga (2021). Adaptive spatial sampling for infectious disease surveillance using bayesian intensity updating. Spatial and Spatio-temporal Epidemiology, 38:100453.