



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원 저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리와 책임은 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)



Optimizing the maximum reported cluster size for multinomial-based spatial scan statistic

Jisu Moon

The Graduate School
Yonsei University
Department of Biostatistics and Computing

Optimizing the maximum reported cluster size for multinomial-based spatial scan statistic

A Dissertation Submitted to the
Department of Biostatistics and Computing
and the Graduate School of Yonsei University
in the partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Biostatistics and Computing

Jisu Moon

December 2022



This certifies that the dissertation of
Jisu Moon is approved.

Inkyung Jung: Thesis Supervisor

Chung Mo Nam: Thesis Committee Member #1

Sohee Park: Thesis Committee Member #2

Min Jin Ha: Thesis Committee Member #3

Goeun Park: Thesis Committee Member #4

The Graduate School
Yonsei University
December 2022

Contents

List of Tables	iii
List of Figures	x
Abstract	xi
1 Introduction	1
2 Method	4
2.1 Spatial scan statistic for multinomial model	4
2.2 Existing methods.....	5
2.3 Proposed method.....	11
2.4 Comparison with the ordinal model	14
3 Simulation study	17
3.1 Simulation setting	17
3.2 Data generation	21
3.3 Performance evaluation.....	23
3.4 Results.....	25
4 Application	35
4.1 Korea Community Health Survey data	35
4.2 Results	36



5 Discussion and Conclusion	40
Bibliography	43
Appendix	46
국문요약	90

List of Tables

Table 1. Simulation scenarios for the true cluster model and alternative hypothesis.....	20
Table 2. Multinomial model: Simulation results for the true cluster model (B) and alternative hypothesis (1) using elliptical windows	27
Table 3. Multinomial model: Simulation results for the true cluster model (B) and alternative hypothesis (2) using elliptical windows	28
Table 4. Multinomial model: Simulation results for the true cluster model (B) and alternative hypothesis (3) using elliptical windows	29
Table 5. Multinomial model: Simulation results for the true cluster model (B) and alternative hypothesis (4) using elliptical windows	30
Table 6. Ordinal model: Simulation results for true cluster model (B) with an alternative hypothesis (1) using elliptical windows.....	31
Table 7. Ordinal model: Simulation results for true cluster model (B) with an alternative hypothesis (2) using elliptical windows.....	32
Table 8. Ordinal model: Simulation results for true cluster model (B) with an alternative hypothesis (3) using elliptical windows.....	33
Table 9. Ordinal model: Simulation results for true cluster model (B) with an alternative hypothesis (4) using elliptical windows.....	34

Table 10. A summary of the significant spatial clusters identified using the multinomial-based spatial scan statistic with elliptical windows at the MRCS suggested by 1) Default setting, 2) SCIC ₁ , 3) SCIC ₂ , 4) Elbow method, 5) MCS-P, and 6) MCHS-P.....	39
--	----

Appendix

Table A1. Multinomial model: Simulation results for the true cluster model (A) and alternative hypothesis (1) using elliptical windows	46
Table A2. Multinomial model: Simulation results for the true cluster model (A) and alternative hypothesis (2) using elliptical windows	47
Table A3. Multinomial model: Simulation results for the true cluster model (A) and alternative hypothesis (3) using elliptical windows	48
Table A4. Multinomial model: Simulation results for the true cluster model (A) and alternative hypothesis (4) using elliptical windows	49
Table A5. Multinomial model: Simulation results for the true cluster model (C) and alternative hypothesis (1) using elliptical windows	50
Table A6. Multinomial model: Simulation results for the true cluster model (C) and alternative hypothesis (2) using elliptical windows	51
Table A7. Multinomial model: Simulation results for the true cluster model (C) and alternative hypothesis (3) using elliptical windows	52

Table A8. Multinomial model: Simulation results for the true cluster model (C) and alternative hypothesis (4) using elliptical windows	53
Table A9. Multinomial model: Simulation results for the true cluster model (D) and alternative hypothesis (1) using elliptical windows	54
Table A10. Multinomial model: Simulation results for the true cluster model (D) and alternative hypothesis (2) using elliptical windows	55
Table A11. Multinomial model: Simulation results for the true cluster model (D) and alternative hypothesis (3) using elliptical windows	56
Table A12. Multinomial model: Simulation results for the true cluster model (D) and alternative hypothesis (4) using elliptical windows	57
Table A13. Multinomial model: Simulation results for the true cluster model (D) and alternative hypothesis (5) using elliptical windows	58
Table A14. Multinomial model: Simulation results for the true cluster model (D) and alternative hypothesis (6) using elliptical windows	59
Table A15. Multinomial model: Simulation results for the true cluster model (D) and alternative hypothesis (7) using elliptical windows	60
Table A16. Multinomial model: Simulation results for the true cluster model (E) and alternative hypothesis (1) using elliptical windows	61

Table A17. Multinomial model: Simulation results for the true cluster model (E) and alternative hypothesis (2) using elliptical windows	62
Table A18. Multinomial model: Simulation results for the true cluster model (E) and alternative hypothesis (3) using elliptical windows	63
Table A19. Multinomial model: Simulation results for the true cluster model (E) and alternative hypothesis (4) using elliptical windows	64
Table A20. Multinomial model: Simulation results for the true cluster model (E) and alternative hypothesis (5) using elliptical windows	65
Table A21. Multinomial model: Simulation results for the true cluster model (E) and alternative hypothesis (6) using elliptical windows	66
Table A22. Multinomial model: Simulation results for the true cluster model (E) and alternative hypothesis (7) using elliptical windows	67
Table A23. Ordinal model: Simulation results for true cluster model (A) with an alternative hypothesis (1) using elliptical windows	68
Table A24. Ordinal model: Simulation results for true cluster model (A) with an alternative hypothesis (2) using elliptical windows	69
Table A25. Ordinal model: Simulation results for true cluster model (A) with an alternative hypothesis (3) using elliptical windows	70

Table A26. Ordinal model: Simulation results for true cluster model (A) with an alternative hypothesis (4) using elliptical windows.....	71
Table A27. Ordinal model: Simulation results for true cluster model (C) with an alternative hypothesis (1) using elliptical windows.....	72
Table A28. Ordinal model: Simulation results for true cluster model (C) with an alternative hypothesis (2) using elliptical windows.....	73
Table A29. Ordinal model: Simulation results for true cluster model (C) with an alternative hypothesis (3) using elliptical windows.....	74
Table A30. Ordinal model: Simulation results for true cluster model (C) with an alternative hypothesis (4) using elliptical windows.....	75
Table A31. Ordinal model: Simulation results for true cluster model (D) with an alternative hypothesis (1) using elliptical windows.....	76
Table A32. Ordinal model: Simulation results for true cluster model (D) with an alternative hypothesis (2) using elliptical windows.....	77
Table A33. Ordinal model: Simulation results for true cluster model (D) with an alternative hypothesis (3) using elliptical windows.....	78
Table A34. Ordinal model: Simulation results for true cluster model (D) with an alternative hypothesis (4) using elliptical windows.....	79

Table A35. Ordinal model: Simulation results for true cluster model (D) with an alternative hypothesis (5) using elliptical windows.....	80
Table A36. Ordinal model: Simulation results for true cluster model (D) with an alternative hypothesis (6) using elliptical windows.....	81
Table A37. Ordinal model: Simulation results for true cluster model (D) with an alternative hypothesis (7) using elliptical windows.....	82
Table A38. Ordinal model: Simulation results for true cluster model (E) with an alternative hypothesis (1) using elliptical windows.....	83
Table A39. Ordinal model: Simulation results for true cluster model (E) with an alternative hypothesis (2) using elliptical windows.....	84
Table A40. Ordinal model: Simulation results for true cluster model (E) with an alternative hypothesis (3) using elliptical windows.....	85
Table A41. Ordinal model: Simulation results for true cluster model (E) with an alternative hypothesis (4) using elliptical windows.....	86
Table A42. Ordinal model: Simulation results for true cluster model (E) with an alternative hypothesis (5) using elliptical windows.....	87
Table A43. Ordinal model: Simulation results for true cluster model (E) with an alternative hypothesis (6) using elliptical windows.....	88



Table A44. Ordinal model: Simulation results for true cluster model (E) with an alternative hypothesis (7) using elliptical windows.....	89
--	----



List of Figures

Figure 1. True cluster models considered in the simulation study	19
Figure 2. A map of the significant spatial clusters identified using the multinomial-based spatial scan statistic with elliptical windows at the MRCS suggested by 1) Default setting, 2) SCIC ₁ , 3) SCIC ₂ , 4) Elbow method, 5) MCS-P, and 6) MCHS-P	38

Abstract

Optimizing the Maximum Reported Cluster Size for multinomial-based Spatial Scan Statistic

Correctly identifying spatial disease cluster is a primary issue in public health and epidemiology, and the spatial scan statistic is widely used for detecting spatial disease clusters in spatial epidemiology and disease surveillance. The spatial scan statistic has been developed for various probability models, such as Bernoulli, Poisson, exponential, ordinal, normal and multinomial models. Specifying the maximum reported cluster size (MRCS) is vital when using the spatial scan statistic. Many studies have set the MRCS as the default setting, which is 50% of the total population. However, if we use the MRCS value as the default setting, the clusters larger than true clusters may be reported, which include less meaningful regions. Recently, an optimization measure using the Gini coefficient has been developed to choose an optimal MRCS value. However, it is difficult to clearly define the Gini coefficient for the multinomial-based spatial scan statistic because the Gini coefficient is inapplicable to nominal categorical data. In this study, we propose a spatial cluster information criterion (SCIC) for selecting the optimal MRCS value for the multinomial



model. Since the SCIC is a likelihood-based method, it can be applied to the spatial scan statistic regardless of the probability model. We conducted a simulation to evaluate the performance of the proposed method in the multinomial-based spatial scan statistic, as compared with that in the ordinal-based spatial scan statistic. Three evaluation metrics are used: Sensitivity, positive predicted value (PPV), and misclassification. We also conducted a real data analysis using the Korea Community Health Survey (KCHS) data to demonstrate the usefulness of the proposed method.

Keywords: Spatial cluster detection, Spatial scan statistic, Multinomial model, Maximum reported cluster size

Chapter 1

Introduction

In public health and epidemiology, the spatial scan statistic is a widely used method for identifying spatial clusters with significantly high or low risk of disease outcome. This method is based on the likelihood ratio test statistic for each scanning window to compare its inside and outside. The scanning window that maximizes the test statistic is identified as the most likely cluster. Secondary clusters are also identified as the most likely cluster. Statistical significance of the most likely cluster and secondary clusters is obtained using the Monte Carlo hypothesis testing. The spatial scan statistic has been developed for various probability models such as Poisson (Kulldorff, 1997), Bernoulli (Kulldorff, 1997), exponential (Cook et al., 2007), ordinal (Jung et al., 2007), normal (Kulldorff et al., 2009; Huang et al., 2009) and multinomial (Jung et al., 2010). SaTScan™ software is available to implement this method.

In SaTScan™, the scanning window shape and the maximum scanning window size (MSWS) should be specified by researchers. In many studies, the MSWS value was set as just the default setting (i.e., 50% of the total population). Riberiro and Costa (2012) found that spatial cluster detection results may vary depending on the MSWS value through the extensive simulation study. However, Han et al. (2016) argued that rerunning the analysis while changing the MSWS values could cause multiple testing problem. To overcome this

problem, Han et al. (2016) suggested rerunning the analysis with a fixed large MSWS value while changing the maximum reported cluster size (MRCS) values. Here, if we set the MRCS value as the default setting of 50%; the clusters larger than true clusters may be reported, which include less meaningful regions. Thus, selecting the optimal MRCS value is the major task.

Numerous studies have recently developed optimization criteria to select the optimal value of the MRCS. Han et al. (2016) proposed an optimization criterion using the Gini coefficient (Gini et al., 1912) and defined them for the Poisson-based spatial scan statistic. They showed that the proposed Gini coefficient tends to identify the correct clusters through the simulation study. However, the Gini coefficient should be differently defined according to different probability models. Kim and Jung (2017), Yoo and Jung (2018), and Lee et al. (2021) developed the Gini coefficient for the ordinal-, normal- and exponential-based spatial scan statistic, respectively, but that for the multinomial-based spatial scan statistic has not yet been developed. It seems that it is difficult to clearly define the Gini coefficient for the multinomial-based spatial scan statistic because it is inapplicable to nominal categorical data.

Other studies (Ma et al., 2017; Wang et al., 2020; Meysami et al., 2021) have suggested alternative criteria for selecting the optimal MRCS or MSWS, but all these studies evaluated the performance of their method only for the Poisson-based spatial scan statistic. Unlike the Gini coefficient, these methods might be extended to other probability models because they are likelihood-based optimization criteria. Thus, it is vital to evaluate whether

these existing methods can be used in other probability models other than the Poisson model.

This study proposes a spatial cluster information criterion (SCIC) inspired by the formulation of the Bayes Information Criterion (BIC) (Schwarz, 1978) to choose the optimal MRCS value for the multinomial-based spatial scan statistic. The SCIC can be defined for the spatial scan statistic regardless of the probability model because its approach is based on a likelihood ratio test statistic. Through a simulation study, we evaluated the performance of our proposed method applied to the multinomial- as well as ordinal-based spatial scan statistic. Then we compared the performance of the proposed method with that of the existing methods.

The rest of the study is organized as follows: In Chapter 2, we provide a brief review of the multinomial-based spatial scan statistic. We then describe several existing methods and our proposed method for selecting the optimal MRCS value for the multinomial-based spatial scan statistic. Also, we provide a brief review of the ordinal-based spatial scan statistic and describe above existing methods defined for the ordinal model. In Chapter 3, we evaluate the performance of our proposed method and other existing methods through a simulation study. In Chapter 4, we conduct a real data analysis to apply our proposed method. In Chapter 5, we discuss the results and conclude this paper.

Chapter 2

Method

2.1 Spatial scan statistic for multinomial model

The multinomial-based spatial scan statistic (Jung et al., 2010) is used to detect disease clusters with the statistically different disease-type distribution. Let p_k and q_k denote the probabilities of category k inside and outside the scanning window z , respectively. If we identify regions with different disease-type distributions, the null and alternative hypotheses are stated as

$$H_0: p_1 = q_1, \dots, p_K = q_K \text{ for all } z \in Z \quad v.s. \quad H_1: \text{not } H_0$$

where Z denotes the set of all scanning windows and K denotes the total number of categories. The likelihood ratio test statistic, given the scanning window z , is denoted as

$$\lambda_z = \frac{\prod_k \left\{ \left(\frac{\sum_{i \in z} c_{ik}}{\sum_k \sum_{i \in z} c_{ik}} \right)^{\sum_{i \in z} c_{ik}} \cdot \left(\frac{\sum_{i \notin z} c_{ik}}{\sum_k \sum_{i \notin z} c_{ik}} \right)^{\sum_{i \notin z} c_{ik}} \right\}}{\prod_k \left\{ \left(\frac{C_k}{C} \right)^{c_k} \right\}}$$

where c_{ik} is the number of cases belonging to category k inside the region i , C_k is the total number of cases belonging to category k in the whole study area and C is the total number of cases in the whole study area.



2.2 Existing methods

Elbow method

In unsupervised learning, the Elbow method (Thorndike, 1953), which works by finding the elbow point, is used to determine the optimal number of clusters. In order to choose the sensible value of the MRCS, Meysami et al. (2021) proposed an optimization criterion by adopting the method for selecting the optimal elbow point, as suggested by Delgado et al. (2015). For the multinomial model, the algorithm for computing the Elbow method is as follows:

(Step 1) For a given MRCS $m\%$, detect J_m significant clusters $Z_1^{(m)}, \dots, Z_{J_m}^{(m)}$ using the multinomial-based spatial scan statistic, where $m = 1, \dots, 50$. If no significant cluster is found, use the most likely cluster (MLC).

(Step 2) For each MRCS $m\%$, calculate the negative sum of the likelihood ratio test (LRT) statistic over all significant clusters. Let $\lambda_{Z_j^{(m)}}$ denotes the LRT for the multinomial-based spatial statistic given the j^{th} significant cluster $Z_j^{(m)}$, where $j = 1, \dots, J_m$.

$$-LRT(m) = -\sum_{j=1}^{J_m} \lambda_{Z_j^{(m)}}$$

(Step 3) To define the elbow plot, connect the points $(m, -LRT(m))$ for $m = 1, \dots, 50$.

(Step 4) For each MRCS $m\%$, calculate the orthogonal distance between each point $(m, -LRT(m))$ and the line connecting the first and last points.

(Step 5) Choose the MRCS, which maximizes the orthogonal distance as the optimal MRCS.

Maximum clustering set-proportion (MCS-P)

Ma et al. (2017) proposed an optimization criterion called the maximum clustering set-proportion (MCS-P) to select the optimal value of the MSWS for the Poisson-based spatial scan statistic. This was based on assumption that all detected significant clusters were homogeneous clusters with the same relative risks. However, due to the multiple testing problem, it may be inappropriate to analyze multiple times with different MSWS values and select the optimal result. In this paper, we apply the MCS-P to the multinomial model and use this criterion for selecting the optimal MRCS value, with the MSWS value fixed at 50%. For the multinomial model, the algorithm for computing the MCS-P is as follows:



(Step 1) For a given MRCS $m\%$, detect J_m significant clusters $Z_1^{(m)}, \dots, Z_{J_m}^{(m)}$ using the multinomial-based spatial scan statistic, where $m = 1, \dots, 50$.

(Step 2) For each MRCS $m\%$, define the union cluster set $Z_A^{(m)}$ by merging all detected significant clusters. The expression is as follows:

$$Z_A^{(m)} = \bigcup_{j=1}^{J_m} Z_j^{(m)}$$

where $Z_j^{(m)}$ is the j^{th} detected significant cluster, where $j = 1, \dots, J_m$.

(Step 3) For each MRCS $m\%$, define the union log-likelihood ratio test statistic $\log\lambda_{Z_A^{(m)}}$ given the union cluster set $Z_A^{(m)}$. The expression is as follows:

$$\log\lambda_{Z_A^{(m)}} = \sum_k \left\{ \left(C_k - \sum_{i \in Z_A^{(m)}} c_{ik} \right) \cdot \log \left(\frac{C_k - \sum_{i \in Z_A^{(m)}} c_{ik}}{C - \sum_{i \in Z_A^{(m)}} c_i} \right) + \sum_{i \in Z_A^{(m)}} c_{ik} \cdot \log \left(\frac{c_{ik}}{C_k} \right) \right\}$$

where C is the total number of cases in the whole study area, C_k is the total number of cases belonging to category k in the whole study area, c_i is the number of cases inside the region i and c_{ik} the number of cases belonging to category k inside the region i .

(Step 4) Choose the MRCS, which maximizes the $\log\lambda_{Z_A^{(m)}}$ as the optimal MRCS.



Maximum clustering heterogeneous set–proportion (MCHS-P)

Wang et al. (2020) proposed an optimization criterion called the maximum clustering heterogeneous set–proportion (MCHS-P) to select the optimal value of the MSWS. This was based on assumption that some detected significant clusters were heterogeneous clusters with the different relative risks. As discussed previously, we apply the MCS-P to the multinomial model and use this criterion for selecting the optimal MRCS value, with the MSWS value fixed at 50%. For the multinomial model, the algorithm for computing the MCHS-P is as follows:

(Step 1) For a given MRCS $m\%$, detect J_m significant clusters $Z_1^{(m)}, \dots, Z_{J_m}^{(m)}$ using the multinomial-based spatial scan statistic, where $m = 1, \dots, 50$.

(Step 2) For each MRCS $m\%$, define the heterogeneous cluster set $Z_B^{(m)}$ by merging J_m detected significant clusters into W_m ($W_m \leq J_m$) merged clusters according to their spatial contiguity. The expression is as follows:

$$Z_B^{(m)} = \{Z_{B_1}^{(m)}, \dots, Z_{B_{W_m}}^{(m)}\}$$

(Step 3) For each MRCS $m\%$, define the union log-likelihood ratio test statistic $\log\lambda_{Z_B^{(m)}}$ given the heterogeneous cluster set $Z_B^{(m)}$. The expression is as follows

$$\begin{aligned} \log\lambda_{Z_B^{(m)}} = & \sum_k \left\{ \sum_{i \in Z_{B_1}^{(m)}} c_{ik} \cdot \log \left(\frac{\sum_{i \in Z_{B_1}^{(m)}} c_{ik}}{\sum_{i \in Z_{B_1}^{(m)}} c_i} \right) + \dots + \sum_{i \in Z_{B_{W_m}}^{(m)}} c_{ik} \cdot \log \left(\frac{\sum_{i \in Z_{B_{W_m}}^{(m)}} c_{ik}}{\sum_{i \in Z_{B_{W_m}}^{(m)}} c_i} \right) \right. \\ & \left. + \left(C_k - \sum_{i \in Z_B^{(m)}} c_{ik} \right) \cdot \log \left(\frac{C_k - \sum_{i \in Z_B^{(m)}} c_{ik}}{C - \sum_{i \in Z_B^{(m)}} c_i} \right) \right\} + \sum_k C_k \cdot \log \left(\frac{C_k}{C} \right) \end{aligned}$$

where C is the total number of cases in the whole study area, C_k is the total number of cases belonging to category k in the whole study area, c_i is the number of cases inside the region i and c_{ik} the number of cases belonging to category k inside the region i .

(Step 4) Choose the MRCS, which maximizes the $\log\lambda_{Z_B^{(m)}}$ as the optimal MRCS.

2.3 Proposed method

Spatial cluster information criterion (SCIC)

We propose an optimization criterion called the spatial cluster information criterion (SCIC), inspired by the formulation of the Bayes information criterion (BIC) (Schwarz, 1978), to select the optimal MRCS value. In statistical modeling, the BIC is one of the most widely used criteria for selecting a suitable model. The BIC for a candidate model M_u is defined as follows:

$$BIC(M_u) = -2 \cdot \log L(\widehat{\theta}_u | y) + u \cdot \log (v)$$

where y is observed data, $L(\theta_u | y)$ is the likelihood of y given the model M_u , $\widehat{\theta}_u$ is the maximum likelihood estimation (MLE) of θ_u that maximizes the $L(\theta_u | y)$, u is the number of parameters in the model M_u , and v is the total number of observations. In the BIC equation, the second term is a penalty term, which penalizes additional parameters in the model. The model with the minimum BIC is selected as the suitable model (Neath and Cavanaugh, 2012).

We define the SCIC for selecting the optimal value of the MRCS. The first term of the SCIC is defined by using the sum of the LLR test statistic over all significant clusters. In the multinomial-based spatial scan statistic, the LLR test statistic for each scanning window is used to measure the degree of heterogeneity in the spatial distribution of the disease-type.

The larger the LLR test statistic, the more degree of heterogeneity in the distribution of the disease-type inside the scanning window compared with that outside the scanning window. However, as a large cluster is searched, the LLR test statistic tends to increase because of a rise in the number of cases inside the cluster. Hence, we design a penalty term to offset the increased LLR test statistic because of the large number of cases inside the cluster.

The second term, called the penalty term, is defined in two versions. The first version of the penalty term is defined as the value, calculated by multiplying the logarithm of the number of cases inside the significant clusters by the number of categories and the number of significant clusters. In the second version of the penalty term, we use the number of regions instead of the number of cases inside the significant clusters. We understand them as a functional relationship where the number of cases inside the significant clusters tends to increase with increasing the number of regions inside the significant clusters. Both versions are designed as optimization criteria with similar meanings. For the multinomial model, the algorithm for computing the SCIC is as follows:

(Step 1) For a given MRCS $m\%$, detect J_m significant clusters $Z_1^{(m)}, \dots, Z_{J_m}^{(m)}$ using the multinomial-based spatial scan statistic, where $m = 1, \dots, 50$.

(Step 2) For each MRCS $m\%$, calculate the SCIC for all significant clusters as follows:

$$SCIC_1(m) = -2 \sum_{j=1}^{J_m} \log(\lambda_{Z_j^{(m)}}) + K \cdot J_m \cdot \log(\tau^{(m)}) \quad (\text{Version 1})$$

$$SCIC_2(m) = -2 \sum_{j=1}^{J_m} \log(\lambda_{Z_j^{(m)}}) + K \cdot J_m \cdot \log(\delta^{(m)}) \quad (\text{Version 2})$$

where $\lambda_{Z_j^{(m)}}$ denotes the LRT statistic for the multinomial-based spatial statistic given the j^{th} significant cluster $Z_j^{(m)}$, K denotes the total number of categories, $\tau^{(m)}$ and $\delta^{(m)}$ denote the sum of the number of total cases and the sum of the number of regions inside all significant clusters, respectively. ($j = 1, \dots, J_m$)

(Step 3) Choose the MRCS, which minimizes the SCIC as the optimal MRCS.

2.4 Comparison with the ordinal model

Spatial scan statistic for ordinal model

The ordinal-based spatial scan statistic (Jung et al., 2007) is used to detect disease clusters, which have statistically high or low rates of higher-value category. Let p_k and q_k denote the probabilities of category k inside and outside the scanning window z , respectively. If we want to identify regions with high rates of higher-value category, the null and alternative hypotheses are stated as follows:

$$H_0: p_1 = q_1, \dots, p_K = q_K \text{ for all } z \in Z \text{ v.s. } H_1: \frac{p_1}{q_1} \leq \dots \leq \frac{p_K}{q_K} \text{ for some } z \in Z$$

where Z denotes the set of all scanning windows and K denotes the total number of categories. The likelihood ratio test statistic given the scanning window z is derived as

$$\lambda_z = \frac{\prod_k \left\{ \left(\frac{\sum_{i \in z} c_{ik}}{\sum_k \sum_{i \in z} c_{ik}} \right)^{\sum_{i \in z} c_{ik}} \cdot \left(\frac{\sum_{i \notin z} c_{ik}}{\sum_k \sum_{i \notin z} c_{ik}} \right)^{\sum_{i \notin z} c_{ik}} \right\}}{\prod_k \left\{ \left(\frac{C_k}{C} \right)^{c_k} \right\}}$$

where c_{ik} is the number of cases belonging to category k inside the region i , C_k is the total number of cases belonging to category k in the whole study area and C is the total number of cases in the whole study area.

MRCS optimization criteria in ordinal model

The Elbow method, MCS-P, MCHS-P, and SCIC described in Chapter 2.2 ~ 2.3 can be applied to the ordinal model similar to the multinomial model because these criteria are based on the likelihood ratio test statistic. We only change the probability model from the multinomial model to the ordinal model in each approach to apply these criteria to the ordinal model. In the MCS-P and MCHS-P approaches, both formulations of the union log-likelihood ratio test statistic for the ordinal model and the multinomial model are the same because the formulation of the LLR test statistic for the ordinal model is equal to that of the multinomial model.

In the ordinal model, we can use the Gini coefficient (Han et al., 2016) as the optimization criterion for selecting the MRCS value. In economics, the Gini coefficient, based on the Lorenz curve (Lorenz, 1905), is used to estimate wealth inequality of its distribution. In order to choose the optimal value of the MRCS, Kim and Jung (2017) proposed the optimization criterion for the ordinal-based spatial scan statistic by adopting the Gini coefficient, suggested by Han et al. (2016). For the ordinal model, the algorithm for computing the Gini coefficient is as follows:

(Step 1) For a given MRCS $m\%$, detect J_m significant clusters $Z_1^{(m)}, \dots, Z_{J_m}^{(m)}$ using the ordinal-based spatial scan statistic, where $m = 1, \dots, 50$.

(Step 2) For each MRCS $m\%$, define the x_j -coordinate as the cumulative percentages of weighted cases and the y_j -coordinate as the cumulative percentages of observed cases up to the j^{th} significant cluster. The coordinates are expressed as follows:

$$(x_j, y_j) = \left(\frac{\sum_k k \sum_{i \in \{ \cup_{a=1}^j z_a^{(m)} \}} c_{ik}}{\sum_k k \cdot C_k}, \frac{\sum_k \sum_{i \in \{ \cup_{a=1}^j z_a^{(m)} \}} c_{ik}}{\sum_k C_k} \right)$$

(Step 3) For each MRCS $m\%$, connect the points $(0, 0), (x_1, y_1), \dots, (x_{C_m}, y_{C_m}), (1, 1)$ to define the Lorenz curve.

(Step 4) For each MRCS $m\%$, calculate the Gini coefficient as twice the area between the Lorenz curve and the reference line $y = x$. The Gini coefficient is computed as follows:

$$Gini(m) = \sum_{j=1}^{J_m+1} y_j x_{j-1} - y_{j-1} x_j$$

where $x_0 = y_0 = 0$ and $x_{J_m+1} = y_{J_m+1} = 1$.

(Step 5) Choose the MRCS, which maximizes the Gini coefficient as the optimal MRCS.

Chapter 3

Simulation Study

3.1 Simulation setting

We conducted the simulation study to evaluate and compare the performance of the proposed method to that of other existing methods for the multinomial and ordinal model. We assumed 5 true cluster models with Seoul and Gyeonggi Province in South Korea, consisting of 69 districts, as the whole study region. Shapes and sizes of true clusters are shown in Figure 1. True cluster models (A) and (B) assumed one circular- and elliptical-shaped true cluster with 5 districts, which is 8% of the entire study region, respectively. True cluster model (C) assumed one irregular-shaped true cluster with 10 districts, which is 15% of the entire study region. True cluster models (D) and (E) assumed two circular- and elliptical-shaped true clusters with each 5 districts, respectively.

For each true cluster model, we considered various scenarios of the alternative hypothesis, assuming four categories, to investigate whether the performance of the proposed method is affected by the alternative hypothesis. We adopted the parameter setting for the alternative hypothesis, suggested in previous study (Jung et al., 2010). Since we set the alternative hypotheses to have an ordinal structure, it can be analyzed not only with the multinomial model but also with the ordinal model. Regarding the true cluster

model with two clusters, we included heterogeneous settings with different alternative hypotheses and homogeneous settings with the same alternative hypotheses for each cluster. The goal is to examine whether the proposed method also work well in more plausible heterogeneous settings, in which the relative risks of each category are different between the two clusters.

We considered 4 alternative hypotheses in the true cluster models with one cluster and two homogeneous clusters. Moreover, we considered 3 alternative hypotheses in the true cluster models with two heterogeneous clusters. Overall, we considered 26 scenarios in combination with the true cluster model and alternative hypothesis. Table 1 presents the simulation scenarios for the true cluster model and alternative hypothesis.

Spatial scan statistic used elliptical windows when searching spatial clusters in the simulation study. In SaTScan™, both circular and elliptical windows are used as the scan shape. However, we presented the simulation results obtained from the elliptical windows as the main results because Kulldorff et al. (2006) found that the spatial scan statistic with elliptic windows showed good performance in terms of the power when the shape of the true cluster is elliptical and circular.



Figure 1. True cluster models in the simulation study.

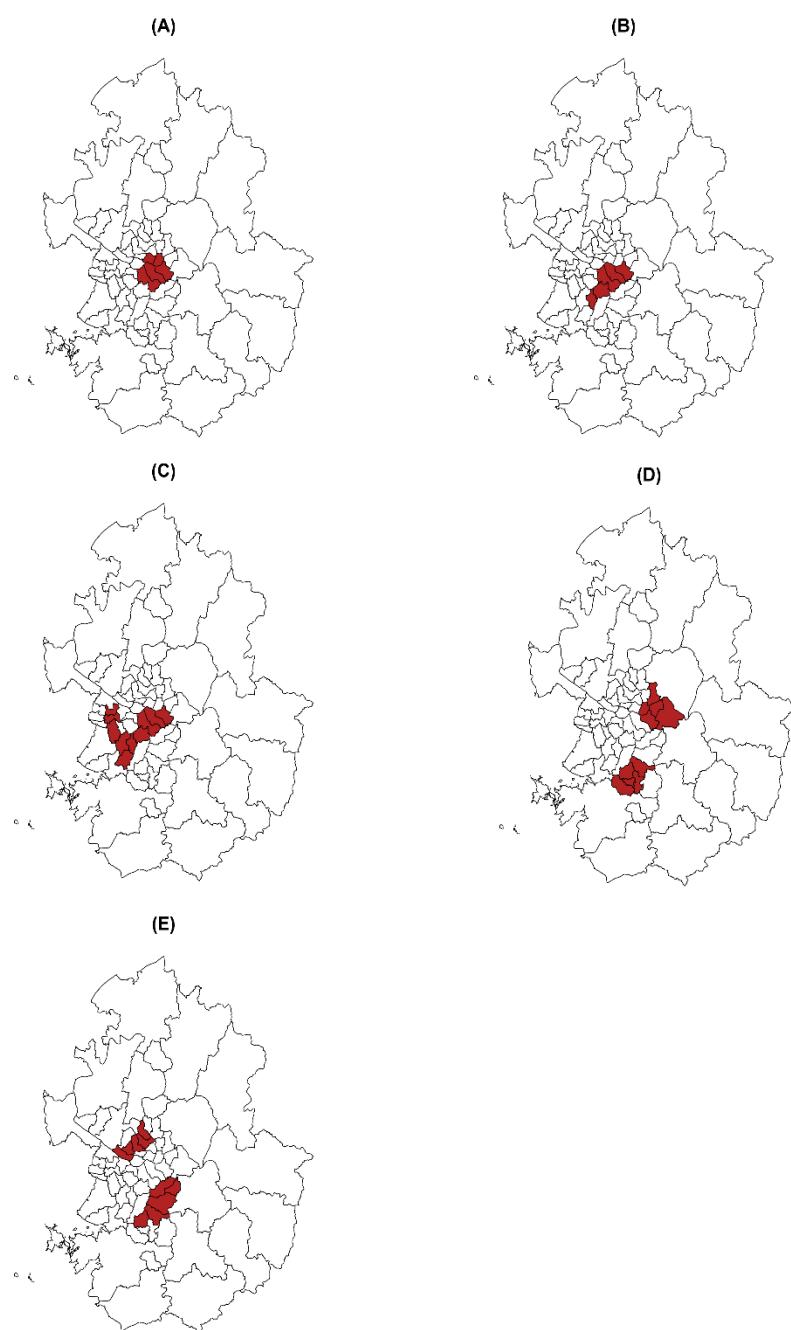


Table 1. Simulation scenarios for the true cluster model and alternative hypothesis.

Setting	True cluster model	Alternative hypothesis* (for cluster 1, for cluster 2)
Single cluster	(A) One circular-shaped cluster (8%)	(1) $(p^{(1)}=(0.05,0.15,0.35,0.45),$ -
	(B) One elliptic-shaped cluster (8%)	(2) $(p^{(1)}=(0.05,0.25,0.25,0.45),$ -
	(C) One irregular-shaped cluster (15%)	(3) $(p^{(1)}=(0.10,0.10,0.40,0.40),$ - (4) $(p^{(1)}=(0.15,0.15,0.15,0.55),$ -
Two homo- geneous clusters	(D) Two circular-shaped clusters (8% each)	(1) $(p^{(1)}=(0.05,0.15,0.35,0.45),$ $p^{(2)}=(0.05,0.15,0.35,0.45))$ (2) $(p^{(1)}=(0.05,0.25,0.25,0.45),$ $p^{(2)}=(0.05,0.25,0.25,0.45))$
	(E) Two elliptic-shaped clusters (8% each)	(3) $(p^{(1)}=(0.10,0.10,0.40,0.40),$ $p^{(2)}=(0.10,0.10,0.40,0.40))$ (4) $(p^{(1)}=(0.15,0.15,0.15,0.55),$ $p^{(2)}=(0.15,0.15,0.15,0.55))$
Two hetero- geneous clusters	(D) Two circular-shaped clusters (8% each)	(5) $(p^{(1)}=(0.05,0.15,0.35,0.45),$ $p^{(2)}=(0.05,0.25,0.25,0.45))$ (6) $(p^{(1)}=(0.05,0.15,0.35,0.45),$ $p^{(2)}=(0.10,0.10,0.40,0.40))$
	(E) Two elliptic-shaped clusters (8% each)	(7) $(p^{(1)}=(0.05,0.15,0.35,0.45),$ $p^{(2)}=(0.15,0.15,0.15,0.55))$

* $p = (0.25, 0.25, 0.25, 0.25)$ was assumed for the remaining areas.

3.2 Data generation

In the simulation study, one of the objectives was to examine whether the optimal MRCS selected using the proposed method is the same as the size of the true cluster. To represent the size of the true cluster, we defined the cluster size θ as a percentage of the number of districts belonging to the true cluster among the total 69 districts. However, in SaTScan™, MRCS (%) is defined using the number of cases, not the number of districts. If the MRCS value is specified to $\theta\%$, SaTScan™ only reports the clusters with the number of cases less than or equal to $\theta\%$ of the total number of cases. To adjust the number of cases inside the true cluster to $\theta\%$ of the total cases, we generated 1000 cases and then assigned $\theta\%$ of the 1000 cases to one of the districts inside the true cluster. The R software version 4.0.2 was used to generate simulated datasets.

If T clusters are in the true cluster model, we randomly assigned $n^{(t)}$ cases ($\theta\%$ of 1000 cases) to one of the districts inside the t^{th} cluster and $(1000 - T \cdot n^{(t)})$ cases to one of the districts outside the true clusters. The number of cases belonging to each category inside the t^{th} cluster $(n_1^{(t)}, n_2^{(t)}, n_3^{(t)}, n_4^{(t)})$ was generated using the following multinomial distribution

$$(n_1^{(t)}, n_2^{(t)}, n_3^{(t)}, n_4^{(t)}) \sim \text{Multinomial}(n^{(t)}, p^{(t)} = (p_1^{(t)}, p_2^{(t)}, p_3^{(t)}, p_4^{(t)}))$$

where $n^{(t)}$ is the total number of cases inside the t^{th} cluster, and $p^{(t)} =$



$(p_1^{(t)}, p_2^{(t)}, p_3^{(t)}, p_4^{(t)})$ is the alternative hypothesis for the t^{th} cluster under each scenario.

Then, we randomly assigned $n_1^{(t)}$ cases to category 1, $n_2^{(t)}$ cases to category 2, $n_3^{(t)}$ cases to category 3 and $n_4^{(t)}$ cases to category 4. The number of cases belonging to each category outside the true clusters ($n_1^{\text{out}}, n_2^{\text{out}}, n_3^{\text{out}}, n_4^{\text{out}}$) was generated using the following multinomial distribution:

$$(n_1^{\text{out}}, n_2^{\text{out}}, n_3^{\text{out}}, n_4^{\text{out}}) \sim \text{Multinomial}(n^{\text{out}}, p^{\text{out}} = (0.25, 0.25, 0.25, 0.25))$$

where n^{out} is the total number of cases outside the true clusters, and $p^{\text{out}} = (0.25, 0.25, 0.25, 0.25)$ is the alternative hypothesis under all scenarios. Then, we randomly assigned n_1^{out} cases to category 1, n_2^{out} cases to category 2, n_3^{out} cases to category 3 and n_4^{out} cases to category 4.

3.3 Performance evaluation

To evaluate the performance of the existing methods and proposed method for the multinomial-and ordinal-based spatial scan statistic, we repeatedly detected clusters with different category-type distributions by changing the MRCS values. In SaTScanTM, the MSWS value was fixed at 50%; whereas the MRCS value was set by changing it to 1%, 2%, 3%, 4%, 5%, 6%, 8%, 10%, 12%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, and 50%. Since SaTScanTM provides the Gini coefficient values over these 17 candidate MRCS values in the Bernoulli and Poisson model, we computed the SCICs, Gini coefficient (for the ordinal model), Elbow method, MCS-P and MCHS-P values over only 17 candidate MRCS values for consistency. Then, we compared the clusters detected using the optimal MRCS chosen by each method with the true clusters.

We repeated the simulation 1000 times for each alternative hypothesis under each true cluster model. For each method, the frequency, with which each candidate MRCS value was chosen as the optimal MRCS, was counted among 1000 simulated datasets. To compare the performance of the proposed method with other existing methods and default setting (i.e., MRCS value of 50%), we used the sensitivity, positive predicted value (PPV) and misclassification as the performance measures following the previous study (Costa et al., 2012). The evaluation metrics are described as follows:

$$\text{Sensitivity} = \frac{|A \cap B|}{|A|}$$

$$\text{PPV} = \frac{|A \cap B|}{|B|}$$

$$\text{Misclassification} = \frac{|(A \cup B) - (A \cap B)|}{|U|}$$

where $|\cdot|$ is a cardinality set, U denotes the set of all 69 districts, A and B denote the sets of districts in the true clusters and detected clusters, respectively.

The sensitivity is defined as the proportion of correctly detected districts among the districts inside the true cluster. The PPV is defined as the proportion of correctly detected districts among the districts inside the detected cluster. The misclassification is defined as the proportion of incorrectly detected districts among the districts inside the true cluster or the detected cluster. Higher values of the sensitivity, PPV, and lower values of the misclassification demonstrate better performance in detecting clusters more accurately. We calculated the averages of the sensitivity, PPV, and misclassification over 1000 simulated datasets using the MRCS value: 1) The values selected by SCIC₁, SCIC₂, Gini coefficient (only for the ordinal model), Elbow method, MCS-P, and MCHS-P and 2) The value of 50% (i.e., the default setting). SaTScan™ version 10.0 and R software version 4.0.2 including the ‘rsatscan’ package (Kleinman, 2015) were used to conduct the simulation.

3.4 Results

For each scenario, the overall detection accuracy is summarized, which is the average of the performance metrics (i.e., Sensitivity, PPV, and misclassification) over 1000 simulated datasets using the MRCS value: 1) the optimal MRCS and 2) the MRCS value of 50%. In all result tables, the overall detection accuracy at the optimal MRCS and the default setting are described as ‘overall’ and ‘default setting’, respectively.

Table 2 ~ 5 and Table A1 ~ A22 show the simulation results for the multinomial model. In the case of the true cluster model (A), (B), (D), and (E), all the five methods (i.e., SCICs, Elbow method, MCS-P, and MCHS-P) similarly most often choose the optimal MRCS equal to the size of the true cluster from 17 candidate MRCS values, regardless of the scenario for the alternative hypothesis (i.e., homo- and heterogeneous setting). However, in the case of the true cluster model (C), all the five methods similarly most often choose the optimal MRCS as 12%, which is smaller than the size of the true cluster (i.e., 30%), regardless of the scenario for the alternative hypothesis. If we use the optimal MRCS rather than the default setting, it tends to report informative multiple clusters smaller than the true irregular cluster rather than to detect one large cluster that contains the true irregular cluster.

In most scenarios, the overall sensitivity is slightly lower than that of the default setting. However, the performance is further improved in terms of the overall PPV compared with that of the default setting, regardless of the method. For each scenario, it seems that all five methods (i.e., SCICs, Elbow method, MCS-P, and MCHS-P) produce similar overall

detection accuracy in terms of Sensitivity, PPV, and misclassification. The overall sensitivity of the SCIC₁ tends to be similar to that of the SCIC₂; whereas the overall PPV of the SCIC₁ tends to be higher than that of the SCIC₂.

When using the alternative hypothesis (1) compared with using the alternative hypothesis (2), (3), and (4), the overall sensitivity and PPV seems to be higher and the frequency of selecting the optimal MRCS with the same size as the true cluster tends to be increased, regardless of the true cluster model. The reason might be that the detection power tends to be lower due to the lower heterogeneity of the distribution for the category-type when using the alternative hypothesis (2), (3), and (4) compared with using the alternative hypothesis (A).

Table 6 ~ 9 and Table A23 ~ A44 show the simulation results for the ordinal model. The results for the ordinal model have consistently similar trends to those for the multinomial model, except for the Gini coefficient. However, using the ordinal model rather than the multinomial model appears to improve the overall detection accuracy in terms of Sensitivity, PPV, and misclassification because the alternative hypotheses in the simulation have an ordinal structure. The SCIC₁ and SCIC₂ showed similar performance in terms of the overall sensitivity and PPV in most scenarios for the single cluster settings. However, it seems that the overall sensitivity of the SCIC₂ tends to be slightly higher than that of the SCIC₁ in part of the scenarios (i.e., mainly in the two clusters models).

Table 2. Multinomial model: Simulation results for the true cluster model (B) and alternative hypothesis (1) using elliptical windows.

		Maximum reported cluster size (MRCS)																			Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall		
SCIC ₁	Freq	NA	2	NA	9	32	61	547	123	85	39	20	4	3	1	NA	NA	1	927	927	
	Sen	NA	0.200	NA	0.444	0.556	0.695	0.897	0.927	0.972	0.949	0.890	0.850	0.933	0.000	NA	NA	1.000	0.878	0.888	
	PPV	NA	1.000	NA	1.000	0.927	0.939	0.965	0.773	0.655	0.519	0.356	0.237	0.254	0.000	NA	NA	0.088	0.869	0.803	
	Mis	NA	0.058	NA	0.040	0.035	0.027	0.011	0.026	0.043	0.071	0.133	0.210	0.208	0.406	NA	NA	0.754	0.026	0.044	
SCIC ₂	Freq	NA	2	NA	13	32	66	521	115	86	39	24	8	10	2	4	2	3	927	927	
	Sen	NA	0.200	NA	0.538	0.594	0.694	0.899	0.920	0.972	0.959	0.892	0.900	0.860	0.500	0.800	1.000	0.933	0.878	0.888	
	PPV	NA	1.000	NA	0.981	0.947	0.933	0.967	0.770	0.649	0.513	0.339	0.233	0.195	0.081	0.126	0.132	0.105	0.850	0.803	
	Mis	NA	0.058	NA	0.035	0.032	0.027	0.010	0.027	0.044	0.074	0.142	0.223	0.278	0.391	0.420	0.478	0.594	0.034	0.044	
Elbow	Freq	NA	1	2	24	45	69	531	117	80	27	20	5	4	1	1	NA	NA	927	927	
	Sen	NA	0.200	0.800	0.692	0.667	0.707	0.898	0.916	0.968	0.956	0.850	0.920	0.850	0.000	0.600	NA	NA	0.874	0.888	
	PPV	NA	1.000	1.000	0.976	0.944	0.928	0.958	0.764	0.651	0.501	0.282	0.216	0.199	0.000	0.094	NA	NA	0.868	0.803	
	Mis	NA	0.058	0.014	0.024	0.027	0.027	0.011	0.028	0.044	0.079	0.174	0.249	0.272	0.406	0.449	NA	NA	0.027	0.044	
MCS-P	Freq	NA	NA	3	25	49	60	486	129	77	56	27	5	5	3	2	NA	NA	927	927	
	Sen	NA	NA	0.333	0.576	0.624	0.677	0.893	0.929	0.977	0.971	0.933	0.880	1.000	0.667	1.000	NA	NA	0.872	0.888	
	PPV	NA	NA	0.333	0.939	0.861	0.916	0.973	0.787	0.679	0.555	0.411	0.285	0.271	0.145	0.192	NA	NA	0.856	0.803	
	Mis	NA	NA	0.068	0.035	0.034	0.029	0.010	0.023	0.036	0.059	0.104	0.168	0.197	0.309	0.304	NA	NA	0.026	0.044	
MCHS-P	Freq	NA	NA	2	19	44	60	484	131	80	58	32	6	6	3	2	NA	NA	927	927	
	Sen	NA	NA	0.800	0.674	0.677	0.690	0.895	0.919	0.975	0.972	0.925	0.867	0.967	0.667	1.000	NA	NA	0.882	0.888	
	PPV	NA	NA	1.000	1.000	0.910	0.910	0.968	0.770	0.663	0.549	0.390	0.273	0.257	0.145	0.192	NA	NA	0.848	0.803	
	Mis	NA	NA	0.014	0.024	0.029	0.029	0.010	0.027	0.040	0.061	0.115	0.179	0.208	0.309	0.304	NA	NA	0.028	0.044	

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table 3. Multinomial model: Simulation results for the true cluster model (B) and alternative hypothesis (2) using elliptical windows.

	1%	2%	3%	4%	5%	6%	8%	Maximum reported cluster size (MRCS)												Default Setting
								10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall		
SCIC ₁	Freq	NA	2	4	11	43	54	403	132	72	53	25	12	7	3	1	1	3	826	826
	Sen	NA	0.100	0.350	0.400	0.567	0.641	0.890	0.923	0.961	0.970	0.840	0.817	1.000	0.933	1.000	1.000	1.000	0.862	0.876
	PPV	NA	0.500	1.000	0.879	0.927	0.894	0.959	0.763	0.651	0.545	0.348	0.256	0.275	0.203	0.185	0.172	0.139	0.824	0.752
	Mis	NA	0.072	0.047	0.049	0.035	0.032	0.011	0.028	0.043	0.065	0.130	0.200	0.193	0.271	0.319	0.348	0.469	0.035	0.060
SCIC ₂	Freq	NA	2	3	10	40	53	392	130	68	51	26	15	11	8	4	5	8	826	826
	Sen	NA	0.100	0.333	0.380	0.585	0.649	0.890	0.932	0.956	0.969	0.838	0.853	0.945	0.825	0.800	0.960	0.975	0.865	0.876
	PPV	NA	0.500	1.000	0.867	0.938	0.910	0.960	0.772	0.646	0.544	0.340	0.240	0.244	0.155	0.131	0.146	0.118	0.805	0.752
	Mis	NA	0.072	0.048	0.051	0.034	0.031	0.011	0.027	0.044	0.066	0.134	0.220	0.223	0.346	0.402	0.414	0.567	0.046	0.060
Elbow	Freq	NA	1	3	11	51	57	412	129	70	41	20	13	8	4	3	1	NA	824	826
	Sen	NA	0.000	0.333	0.509	0.600	0.653	0.891	0.927	0.954	0.966	0.800	0.877	0.850	0.900	0.467	1.000	NA	0.859	0.876
	PPV	NA	0.000	1.000	0.879	0.913	0.908	0.955	0.761	0.648	0.539	0.312	0.240	0.219	0.164	0.083	0.161	NA	0.828	0.752
	Mis	NA	0.087	0.048	0.041	0.034	0.031	0.012	0.029	0.044	0.067	0.147	0.224	0.250	0.355	0.357	0.377	NA	0.035	0.060
MCS-P	Freq	NA	2	5	11	46	51	378	121	74	60	36	18	11	3	5	3	2	826	826
	Sen	NA	0.000	0.280	0.473	0.565	0.655	0.886	0.924	0.962	0.967	0.878	0.911	0.855	0.933	0.960	1.000	1.000	0.862	0.876
	PPV	NA	0.000	0.800	0.788	0.870	0.899	0.961	0.775	0.671	0.551	0.380	0.302	0.235	0.203	0.187	0.167	0.155	0.802	0.752
	Mis	NA	0.087	0.058	0.046	0.038	0.031	0.011	0.025	0.037	0.062	0.115	0.159	0.213	0.271	0.304	0.362	0.399	0.038	0.060
MCHS-P	Freq	NA	2	4	9	47	50	376	123	76	60	35	19	12	3	5	3	2	826	826
	Sen	NA	0.000	0.350	0.556	0.591	0.652	0.887	0.922	0.961	0.970	0.886	0.884	0.867	0.933	0.960	1.000	1.000	0.866	0.876
	PPV	NA	0.000	1.000	0.807	0.891	0.903	0.961	0.767	0.662	0.548	0.383	0.287	0.234	0.203	0.187	0.167	0.155	0.800	0.752
	Mis	NA	0.087	0.047	0.042	0.036	0.031	0.011	0.027	0.040	0.063	0.114	0.175	0.217	0.271	0.304	0.362	0.399	0.039	0.060

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table 4. Multinomial model: Simulation results for the true cluster model (B) and alternative hypothesis (3) using elliptical windows.

	1%	2%	3%	4%	5%	6%	8%	Maximum reported cluster size (MRCS)												Default Setting
								10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall		
SCIC ₁	Freq	NA	1	4	10	46	52	424	117	74	44	27	9	2	4	1	3	1	819	819
	Sen	NA	0.200	0.250	0.380	0.552	0.650	0.881	0.916	0.924	0.973	0.874	0.911	1.000	0.750	0.400	0.733	0.800	0.850	0.868
	PPV	NA	1.000	0.750	0.917	0.889	0.911	0.954	0.756	0.620	0.532	0.336	0.270	0.278	0.173	0.080	0.094	0.105	0.827	0.755
	Mis	NA	0.058	0.062	0.048	0.039	0.031	0.012	0.029	0.050	0.068	0.145	0.193	0.188	0.279	0.377	0.531	0.507	0.036	0.055
SCIC ₂	Freq	NA	1	4	13	46	52	416	112	73	41	28	11	4	6	5	6	1	819	819
	Sen	NA	0.200	0.250	0.523	0.578	0.654	0.878	0.913	0.921	0.976	0.857	0.909	1.000	0.833	0.880	0.833	0.800	0.850	0.868
	PPV	NA	1.000	0.750	0.946	0.903	0.900	0.951	0.754	0.618	0.533	0.322	0.251	0.236	0.161	0.140	0.112	0.105	0.814	0.755
	Mis	NA	0.058	0.062	0.037	0.036	0.031	0.013	0.030	0.051	0.068	0.152	0.212	0.246	0.341	0.394	0.493	0.507	0.042	0.055
Elbow	Freq	NA	1	5	18	56	57	429	105	73	33	21	11	2	3	3	2	NA	819	819
	Sen	NA	0.200	0.320	0.611	0.632	0.653	0.882	0.897	0.915	0.964	0.800	0.964	1.000	0.667	0.800	0.600	NA	0.844	0.868
	PPV	NA	1.000	0.800	0.961	0.899	0.891	0.949	0.738	0.609	0.513	0.295	0.262	0.225	0.118	0.113	0.094	NA	0.830	0.755
	Mis	NA	0.058	0.055	0.030	0.033	0.032	0.013	0.032	0.053	0.074	0.163	0.208	0.268	0.372	0.473	0.449	NA	0.036	0.055
MCS-P	Freq	NA	3	5	13	50	56	384	104	80	53	41	14	7	5	2	2	NA	819	819
	Sen	NA	0.133	0.160	0.492	0.616	0.632	0.874	0.919	0.940	0.962	0.927	0.957	0.971	0.800	0.700	0.600	NA	0.850	0.868
	PPV	NA	0.667	0.400	0.808	0.844	0.879	0.955	0.774	0.650	0.545	0.400	0.317	0.253	0.180	0.133	0.094	NA	0.804	0.755
	Mis	NA	0.068	0.078	0.046	0.036	0.034	0.012	0.025	0.042	0.063	0.108	0.153	0.211	0.278	0.348	0.449	NA	0.037	0.055
MCHS-P	Freq	NA	1	3	11	47	54	381	107	81	54	46	17	7	5	3	2	NA	819	819
	Sen	NA	0.200	0.267	0.655	0.647	0.644	0.873	0.920	0.936	0.967	0.904	0.953	0.971	0.800	0.800	0.600	NA	0.861	0.868
	PPV	NA	1.000	0.667	0.982	0.888	0.895	0.950	0.762	0.645	0.536	0.375	0.303	0.253	0.180	0.144	0.094	NA	0.799	0.755
	Mis	NA	0.058	0.063	0.026	0.033	0.032	0.013	0.027	0.044	0.066	0.124	0.166	0.211	0.278	0.353	0.449	NA	0.039	0.055

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table 5. Multinomial model: Simulation results for the true cluster model (B) and alternative hypothesis (4) using elliptical windows.

		Maximum reported cluster size (MRCS)																	Default Setting	
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%		
SCIC ₁	Freq	NA	6	8	21	32	69	428	111	72	47	28	12	7	6	1	1	1	850	850
	Sen	NA	0.167	0.325	0.381	0.519	0.661	0.884	0.901	0.942	0.962	0.914	0.800	0.886	0.633	0.800	1.000	1.000	0.839	0.860
	PPV	NA	0.833	0.813	0.921	0.863	0.938	0.950	0.750	0.636	0.525	0.380	0.256	0.198	0.139	0.160	0.161	0.082	0.823	0.733
	Mis	NA	0.063	0.054	0.048	0.042	0.028	0.013	0.030	0.046	0.070	0.119	0.188	0.273	0.324	0.319	0.377	0.812	0.037	0.068
SCIC ₂	Freq	NA	7	8	26	35	69	401	101	65	44	27	21	16	11	5	6	8	850	850
	Sen	NA	0.171	0.350	0.446	0.549	0.658	0.884	0.895	0.935	0.959	0.911	0.819	0.900	0.727	0.800	0.900	1.000	0.835	0.860
	PPV	NA	0.857	0.813	0.913	0.864	0.938	0.951	0.744	0.634	0.513	0.363	0.228	0.191	0.146	0.130	0.131	0.114	0.792	0.733
	Mis	NA	0.062	0.053	0.045	0.040	0.028	0.012	0.032	0.046	0.074	0.135	0.222	0.289	0.331	0.406	0.442	0.576	0.054	0.068
Elbow	Freq	NA	5	9	35	51	74	423	99	65	39	18	16	10	4	1	1	NA	850	850
	Sen	NA	0.160	0.422	0.560	0.608	0.654	0.887	0.905	0.935	0.949	0.911	0.800	0.800	0.750	0.800	1.000	NA	0.833	0.860
	PPV	NA	0.800	0.833	0.914	0.869	0.932	0.947	0.755	0.638	0.493	0.320	0.216	0.164	0.173	0.160	0.161	NA	0.829	0.733
	Mis	NA	0.064	0.047	0.037	0.036	0.029	0.013	0.029	0.045	0.081	0.167	0.232	0.310	0.283	0.319	0.377	NA	0.038	0.068
MCS -P	Freq	NA	4	6	25	41	66	380	106	86	56	39	13	12	9	4	3	NA	850	850
	Sen	NA	0.100	0.333	0.568	0.600	0.630	0.878	0.900	0.944	0.971	0.949	0.923	0.800	0.867	0.950	1.000	NA	0.848	0.860
	PPV	NA	0.500	0.833	0.921	0.878	0.908	0.952	0.761	0.648	0.545	0.424	0.308	0.220	0.194	0.184	0.163	NA	0.801	0.733
	Mis	NA	0.072	0.053	0.035	0.036	0.032	0.013	0.028	0.042	0.062	0.099	0.156	0.220	0.272	0.308	0.372	NA	0.039	0.068
MCHS -P	Freq	NA	3	5	21	41	64	372	106	86	59	41	18	15	12	4	3	NA	850	850
	Sen	NA	0.133	0.320	0.562	0.610	0.644	0.882	0.900	0.942	0.966	0.946	0.889	0.827	0.817	0.950	1.000	NA	0.853	0.860
	PPV	NA	0.667	0.800	0.915	0.882	0.926	0.950	0.758	0.644	0.529	0.409	0.276	0.213	0.178	0.184	0.163	NA	0.787	0.733
	Mis	NA	0.068	0.055	0.036	0.035	0.030	0.013	0.029	0.043	0.068	0.107	0.187	0.236	0.292	0.308	0.372	NA	0.044	0.068

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table 6. Ordinal model: Simulation results for true cluster model (B) with an alternative hypothesis (1) using elliptical windows.

	1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Maximum reported cluster size (MRCS)		Default Setting
																		Overall		
SCIC ₁	Freq	1	3	4	18	37	68	585	130	80	36	14	3	1	1	1	NA	NA	982	982
	Sen	0.200	0.200	0.300	0.433	0.584	0.685	0.900	0.943	0.978	0.961	0.943	0.800	1.000	1.000	1.000	NA	NA	0.874	0.891
	PPV	1.000	1.000	0.875	0.924	0.973	0.950	0.980	0.794	0.683	0.544	0.406	0.258	0.278	0.238	0.200	NA	NA	0.899	0.848
	Mis	0.058	0.058	0.054	0.044	0.031	0.026	0.009	0.022	0.035	0.062	0.108	0.179	0.188	0.232	0.290	NA	NA	0.020	0.029
SCIC ₂	Freq	NA	5	5	28	51	75	566	120	80	32	14	3	1	1	1	NA	NA	982	982
	Sen	NA	0.280	0.360	0.529	0.667	0.704	0.901	0.945	0.975	0.963	0.943	0.800	1.000	1.000	1.000	NA	NA	0.871	0.891
	PPV	NA	1.000	0.900	0.917	0.948	0.935	0.981	0.795	0.682	0.545	0.399	0.258	0.278	0.238	0.200	NA	NA	0.899	0.848
	Mis	NA	0.052	0.049	0.038	0.028	0.026	0.009	0.022	0.035	0.062	0.113	0.179	0.188	0.232	0.290	NA	NA	0.020	0.029
Gini	Freq	NA	NA	1	31	80	63	459	100	106	80	37	7	6	5	4	2	1	982	982
	Sen	NA	NA	0.600	0.690	0.830	0.743	0.888	0.926	0.979	0.978	0.968	0.886	1.000	1.000	1.000	1.000	1.000	0.893	0.891
	PPV	NA	NA	1.000	0.967	0.943	0.931	0.978	0.784	0.674	0.562	0.408	0.283	0.274	0.224	0.196	0.170	0.152	0.845	0.848
	Mis	NA	NA	0.029	0.024	0.017	0.024	0.010	0.024	0.036	0.058	0.107	0.170	0.193	0.252	0.297	0.355	0.406	0.029	0.029
Elbow	Freq	NA	3	9	72	118	81	506	106	64	15	5	1	NA	2	NA	NA	NA	982	982
	Sen	NA	0.333	0.533	0.742	0.812	0.731	0.897	0.934	0.959	0.947	0.920	0.800	NA	1.000	NA	NA	NA	0.866	0.891
	PPV	NA	1.000	0.889	0.954	0.933	0.909	0.975	0.786	0.664	0.537	0.384	0.267	NA	0.223	NA	NA	NA	0.910	0.848
	Mis	NA	0.048	0.037	0.021	0.019	0.026	0.009	0.023	0.039	0.064	0.119	0.174	NA	0.254	NA	NA	NA	0.019	0.029
MSC-P	Freq	NA	1	5	48	96	64	465	120	78	59	27	6	4	4	4	1	NA	982	982
	Sen	NA	0.400	0.640	0.733	0.850	0.731	0.890	0.940	0.977	0.973	0.956	0.867	1.000	1.000	1.000	1.000	NA	0.887	0.891
	PPV	NA	1.000	0.900	0.951	0.944	0.921	0.975	0.796	0.678	0.557	0.408	0.278	0.276	0.225	0.196	0.179	NA	0.867	0.848
	Mis	NA	0.043	0.029	0.022	0.015	0.026	0.010	0.022	0.036	0.059	0.107	0.171	0.192	0.250	0.297	0.333	NA	0.025	0.029
MSHC-P	Freq	NA	2	10	57	92	64	462	117	76	57	27	5	4	4	4	1	NA	982	982
	Sen	NA	0.400	0.580	0.740	0.835	0.716	0.887	0.937	0.976	0.972	0.956	0.840	1.000	1.000	1.000	1.000	NA	0.879	0.891
	PPV	NA	1.000	0.950	0.970	0.928	0.925	0.977	0.794	0.677	0.556	0.405	0.268	0.276	0.225	0.196	0.179	NA	0.870	0.848
	Mis	NA	0.043	0.032	0.020	0.018	0.026	0.010	0.022	0.036	0.059	0.109	0.177	0.192	0.250	0.297	0.333	NA	0.026	0.029

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table 7. Ordinal model: Simulation results for true cluster model (B) with an alternative hypothesis (2) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall	
SCIC ₁	Freq	NA	1	6	15	47	68	472	154	83	51	20	5	9	3	1	2	1	938	938
	Sen	NA	0.200	0.367	0.413	0.566	0.662	0.894	0.940	0.961	0.965	0.950	0.880	0.978	0.667	1.000	1.000	0.800	0.868	0.883
	PPV	NA	1.000	1.000	0.956	0.938	0.935	0.968	0.793	0.666	0.556	0.399	0.298	0.269	0.145	0.179	0.182	0.114	0.858	0.796
	Mis	NA	0.058	0.046	0.044	0.034	0.028	0.010	0.022	0.038	0.059	0.109	0.159	0.195	0.304	0.333	0.326	0.464	0.028	0.042
SCIC ₂	Freq	NA	1	5	18	54	73	468	152	76	52	18	5	9	3	1	2	1	938	938
	Sen	NA	0.200	0.400	0.433	0.607	0.688	0.895	0.937	0.966	0.965	0.944	0.880	0.933	0.667	1.000	1.000	0.800	0.867	0.883
	PPV	NA	1.000	0.950	0.968	0.936	0.925	0.968	0.792	0.670	0.558	0.400	0.298	0.257	0.145	0.179	0.182	0.114	0.860	0.796
	Mis	NA	0.058	0.046	0.043	0.032	0.028	0.010	0.023	0.037	0.059	0.108	0.159	0.205	0.304	0.333	0.326	0.464	0.027	0.042
Gini	Freq	NA	1	4	18	58	66	376	119	108	70	58	23	13	7	7	6	4	938	938
	Sen	NA	0.200	0.350	0.722	0.710	0.694	0.884	0.924	0.972	0.960	0.962	0.939	0.985	0.829	0.943	1.000	0.950	0.883	0.883
	PPV	NA	1.000	1.000	0.989	0.957	0.929	0.966	0.778	0.672	0.558	0.420	0.310	0.281	0.185	0.179	0.170	0.144	0.796	0.796
	Mis	NA	0.058	0.047	0.021	0.024	0.027	0.011	0.025	0.037	0.059	0.101	0.158	0.185	0.277	0.319	0.355	0.417	0.041	0.042
Elbow	Freq	NA	1	7	39	79	76	460	146	73	35	12	2	3	2	1	2	NA	938	938
	Sen	NA	0.200	0.486	0.697	0.714	0.697	0.893	0.929	0.940	0.954	0.933	0.800	0.600	0.800	1.000	1.000	NA	0.861	0.883
	PPV	NA	1.000	0.964	0.974	0.924	0.910	0.964	0.780	0.647	0.548	0.400	0.228	0.176	0.174	0.179	0.182	NA	0.872	0.796
	Mis	NA	0.058	0.039	0.023	0.025	0.028	0.010	0.024	0.042	0.061	0.107	0.225	0.251	0.290	0.333	0.326	NA	0.025	0.042
MSC-P	Freq	NA	1	6	25	68	69	403	145	84	55	34	19	10	6	6	4	3	938	938
	Sen	NA	0.200	0.467	0.712	0.744	0.699	0.889	0.934	0.964	0.967	0.947	0.958	0.940	0.767	0.933	1.000	0.933	0.879	0.883
	PPV	NA	1.000	0.958	0.992	0.953	0.906	0.966	0.784	0.664	0.560	0.411	0.323	0.261	0.175	0.179	0.174	0.141	0.826	0.796
	Mis	NA	0.058	0.041	0.021	0.022	0.028	0.011	0.024	0.038	0.058	0.104	0.150	0.201	0.280	0.316	0.344	0.420	0.035	0.042
MSHC-P	Freq	NA	2	11	32	66	68	398	143	82	54	34	19	10	6	6	4	3	938	938
	Sen	NA	0.300	0.545	0.731	0.733	0.697	0.888	0.933	0.963	0.970	0.947	0.958	0.940	0.767	0.933	1.000	0.933	0.875	0.883
	PPV	NA	1.000	0.955	0.992	0.945	0.902	0.967	0.784	0.663	0.561	0.411	0.323	0.261	0.175	0.179	0.174	0.141	0.827	0.796
	Mis	NA	0.051	0.036	0.020	0.023	0.029	0.010	0.024	0.039	0.058	0.104	0.150	0.201	0.280	0.316	0.344	0.420	0.035	0.042

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table 8. Ordinal model: Simulation results for true cluster model (B) with an alternative hypothesis (3) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall	
SCIC ₁	Freq	NA	3	5	15	51	67	499	129	75	48	15	5	2	1	4	NA	1	920	920
	Sen	NA	0.200	0.360	0.400	0.569	0.672	0.882	0.926	0.971	0.963	0.947	0.920	1.000	1.000	1.000	NA	1.000	0.856	0.882
	PPV	NA	1.000	1.000	0.844	0.948	0.929	0.965	0.785	0.673	0.546	0.406	0.299	0.278	0.227	0.189	NA	0.152	0.870	0.797
	Mis	NA	0.058	0.046	0.049	0.034	0.027	0.011	0.024	0.037	0.062	0.106	0.162	0.188	0.246	0.312	NA	0.406	0.026	0.042
SCIC ₂	Freq	NA	3	5	16	53	71	497	124	76	46	15	5	3	1	4	NA	1	920	920
	Sen	NA	0.200	0.360	0.488	0.592	0.676	0.880	0.926	0.971	0.961	0.947	0.880	1.000	1.000	1.000	NA	1.000	0.856	0.882
	PPV	NA	1.000	1.000	0.894	0.942	0.923	0.964	0.782	0.676	0.548	0.406	0.298	0.258	0.227	0.189	NA	0.152	0.870	0.797
	Mis	NA	0.058	0.046	0.042	0.033	0.028	0.011	0.024	0.036	0.061	0.106	0.159	0.213	0.246	0.312	NA	0.406	0.026	0.042
Gini	Freq	NA	NA	3	12	55	62	388	107	107	80	46	19	15	7	9	5	5	920	920
	Sen	NA	NA	0.333	0.583	0.738	0.668	0.874	0.914	0.981	0.985	0.965	0.947	0.960	0.971	1.000	0.960	0.920	0.883	0.882
	PPV	NA	NA	1.000	0.944	0.966	0.918	0.964	0.768	0.678	0.566	0.415	0.310	0.263	0.218	0.189	0.167	0.137	0.796	0.797
	Mis	NA	NA	0.048	0.033	0.021	0.029	0.012	0.027	0.036	0.057	0.103	0.157	0.199	0.255	0.312	0.351	0.426	0.042	0.042
Elbow	Freq	NA	NA	7	35	81	76	475	127	70	32	9	2	1	2	2	1	NA	920	920
	Sen	NA	NA	0.457	0.657	0.711	0.687	0.880	0.910	0.963	0.950	0.911	1.000	1.000	1.000	1.000	1.000	NA	0.852	0.882
	PPV	NA	NA	1.000	0.933	0.939	0.917	0.960	0.767	0.662	0.530	0.377	0.292	0.278	0.227	0.179	0.172	NA	0.878	0.797
	Mis	NA	NA	0.039	0.028	0.025	0.027	0.012	0.027	0.039	0.066	0.118	0.188	0.188	0.246	0.333	0.348	NA	0.024	0.042
MSC-P	Freq	NA	1	4	27	66	67	412	122	90	60	36	15	6	4	6	NA	4	920	920
	Sen	NA	0.200	0.350	0.652	0.764	0.672	0.877	0.923	0.976	0.990	0.950	0.947	0.967	1.000	1.000	NA	0.950	0.874	0.882
	PPV	NA	1.000	1.000	0.942	0.961	0.921	0.960	0.775	0.674	0.564	0.405	0.318	0.258	0.231	0.190	NA	0.141	0.830	0.797
	Mis	NA	0.058	0.047	0.028	0.020	0.028	0.012	0.025	0.036	0.057	0.107	0.152	0.205	0.243	0.309	NA	0.424	0.033	0.042
MSHC-P	Freq	NA	2	8	13	48	65	440	125	89	60	36	15	5	4	6	NA	4	920	920
	Sen	NA	0.100	0.450	0.523	0.675	0.665	0.880	0.923	0.975	0.990	0.950	0.947	0.960	1.000	1.000	NA	0.950	0.872	0.882
	PPV	NA	0.500	1.000	0.921	0.956	0.926	0.963	0.776	0.674	0.564	0.405	0.318	0.267	0.231	0.190	NA	0.141	0.831	0.797
	Mis	NA	0.072	0.040	0.038	0.027	0.028	0.011	0.025	0.036	0.057	0.107	0.152	0.194	0.243	0.309	NA	0.424	0.033	0.042

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table 9. Ordinal model: Simulation results for true cluster model (B) with an alternative hypothesis (4) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall	
SCIC ₁	Freq	NA	7	7	22	44	80	487	123	80	42	17	8	3	1	1	NA	1	923	923
	Sen	NA	0.200	0.371	0.418	0.568	0.660	0.886	0.920	0.958	0.967	0.918	0.900	1.000	0.000	0.800	NA	1.000	0.846	0.873
	PPV	NA	1.000	0.929	0.961	0.947	0.930	0.963	0.778	0.671	0.552	0.395	0.292	0.261	0.000	0.143	NA	0.147	0.869	0.803
	Mis	NA	0.058	0.048	0.044	0.034	0.028	0.011	0.025	0.038	0.060	0.109	0.165	0.208	0.391	0.362	NA	0.420	0.026	0.040
SCIC ₂	Freq	NA	7	9	29	50	78	483	120	76	41	17	7	4	1	NA	NA	1	923	923
	Sen	NA	0.200	0.400	0.441	0.600	0.667	0.888	0.920	0.961	0.966	0.929	0.886	0.950	0.000	NA	NA	1.000	0.843	0.873
	PPV	NA	1.000	0.944	0.956	0.950	0.936	0.964	0.776	0.674	0.550	0.396	0.292	0.241	0.000	NA	NA	0.147	0.873	0.803
	Mis	NA	0.058	0.045	0.043	0.032	0.028	0.011	0.025	0.037	0.061	0.109	0.164	0.225	0.391	NA	NA	0.420	0.026	0.040
Gini	Freq	NA	2	1	32	49	62	390	99	106	84	53	12	13	7	5	4	4	923	923
	Sen	NA	0.200	0.400	0.663	0.706	0.674	0.873	0.901	0.974	0.967	0.940	0.917	0.923	0.857	0.960	1.000	1.000	0.871	0.873
	PPV	NA	1.000	1.000	0.971	0.962	0.938	0.964	0.757	0.675	0.558	0.409	0.309	0.247	0.192	0.181	0.168	0.154	0.803	0.803
	Mis	NA	0.058	0.043	0.026	0.024	0.027	0.012	0.029	0.036	0.059	0.105	0.155	0.212	0.271	0.319	0.359	0.399	0.040	0.040
Elbow	Freq	NA	4	10	53	79	86	474	111	61	29	9	3	3	NA	NA	NA	NA	922	923
	Sen	NA	0.200	0.500	0.657	0.699	0.672	0.889	0.912	0.944	0.966	0.911	0.800	0.667	NA	NA	NA	NA	0.840	0.873
	PPV	NA	1.000	1.000	0.963	0.940	0.916	0.961	0.766	0.664	0.535	0.386	0.239	0.163	NA	NA	NA	NA	0.889	0.803
	Mis	NA	0.058	0.036	0.027	0.026	0.029	0.011	0.027	0.039	0.064	0.113	0.203	0.275	NA	NA	NA	NA	0.023	0.040
MSC-P	Freq	NA	2	5	42	64	72	408	115	89	59	36	11	7	6	3	2	2	923	923
	Sen	NA	0.200	0.440	0.657	0.719	0.672	0.882	0.911	0.957	0.969	0.944	0.873	0.943	0.833	0.933	1.000	1.000	0.860	0.873
	PPV	NA	1.000	1.000	0.967	0.938	0.935	0.964	0.770	0.666	0.550	0.412	0.293	0.243	0.188	0.176	0.161	0.154	0.835	0.803
	Mis	NA	0.058	0.041	0.027	0.024	0.028	0.011	0.027	0.038	0.061	0.103	0.163	0.219	0.273	0.324	0.377	0.399	0.033	0.040
MSHC-P	Freq	NA	4	14	57	67	73	388	111	84	58	36	11	7	6	3	2	2	923	923
	Sen	NA	0.300	0.529	0.698	0.731	0.679	0.880	0.910	0.957	0.969	0.944	0.873	0.943	0.833	0.933	1.000	1.000	0.854	0.873
	PPV	NA	0.917	1.000	0.964	0.937	0.928	0.965	0.768	0.663	0.550	0.410	0.293	0.243	0.188	0.176	0.161	0.154	0.837	0.803
	Mis	NA	0.054	0.034	0.024	0.024	0.028	0.011	0.027	0.039	0.061	0.104	0.163	0.219	0.273	0.324	0.377	0.399	0.034	0.040

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Chapter 4

Application

4.1 Korea Community Health Survey data

We used the Korea Community Health Survey (KCHS) data to illustrate the usefulness of the proposed method. The Korea Disease Control and Prevention Agency has conducted the KCHS annually since 2008 to obtain community-based health statistics. This health survey was conducted at 253 community health centers. The questionnaire addressed health behaviors, self-reported health indicators, and demographic characteristics. In the 2019 KCHS data, the reason for starting to drink was used as the nominal categorical variable. If a subject of the survey had never drunk, the subject was excluded from the analysis. The reason for starting to drink was divided into four categories; 1) recommended by people, 2) out of curiosity, 3) to promote friendship, and 4) other reasons. In Seoul and Gyeonggi province, we searched for regions that had different distributions of the reason for starting to drink among males in their 20s and 30s, using the multinomial-based spatial scan statistic with elliptical windows.

4.2 Results

The detected clusters differed according to the method for optimizing the MRCS value. A map of the significant spatial clusters detected by each method is shown in Figure 2. A summary of the significant spatial clusters detected by each method is presented in Table 10.

The proposed method (i.e., SCIC₁ and SCIC₂) selected 10% smaller than the default setting as the optimal MRCS. Three large clusters were detected when using the default setting; whereas six smaller clusters, which seem more meaningful, were detected when using the proposed method. Cluster 1 detected using the SCICs belongs to Cluster 1 detected using the default setting. Cluster 2 detected using SCICs belongs to Cluster 2 detected using the default setting. Clusters 3, 4, and 5 detected using the SCICs belong to Cluster 3 detected using the default setting. Thus, more meaningful small clusters were found when using the proposed method rather than the default setting. Here, Cluster 4 detected using SCICs is part of Cluster 3 detected using the default setting. However, this is a hidden smaller informative cluster with the highest RR in Category 3, not Category 1. Cluster 6 detected using the SCICs consist of newly detected regions that have not been identified in the default setting. In terms of RRs, the difference between the largest and smallest among RRs by category is larger when the proposed method is used than the default method. A large difference between RRs means that a more meaningful cluster was searched with greater heterogeneity of distribution for category-type.

The Elbow method selected 4% smaller than the default setting as the optimal MRCS.

The MCS-P and MCHS-P selected 2% smaller than the default setting as the optimal MRCS.

Similarly, the spatial clusters were identified using the existing methods as one of the following clusters: smaller clusters belonging to clusters identified using the default setting, smaller clusters partially overlapping those identified using the default setting, or smaller clusters identified in new regions without any overlap with those identified using the default setting. Thus, the spatial clusters identified by the existing methods (i.e., Elbow, MCS-P, and MCHS-P) are more informative and interpretable than the clusters in the default setting. However, the clusters searched by the existing methods are mostly very small clusters consisting of one or two regions. In the case of using the MCHS-P, it seems that it is difficult to consider them as clusters because some detected clusters consisting of one region are remote without adjacent clusters.

Figure 2. A map of the significant spatial clusters identified using the multinomial-based spatial scan statistic with elliptical windows at the MRCS suggested by 1) Default setting, 2) SCIC₁, 3) SCIC₂, 4) Elbow method, 5) MCS-P, and 6) MCHS-P.

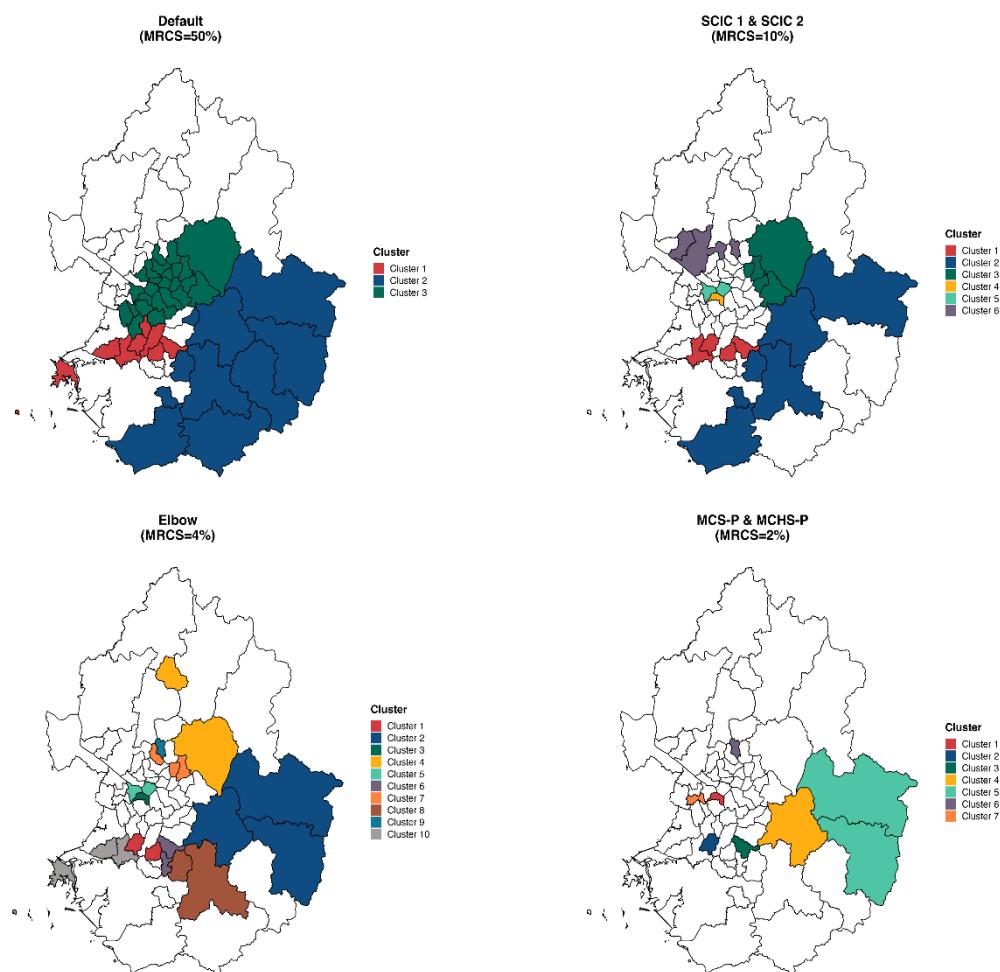




Table 10. A summary of the significant spatial clusters identified using the multinomial-based spatial scan statistic with elliptical windows at the MRCS suggested by 1) Default setting, 2) SCIC₁, 3) SCIC₂, 4) Elbow method, 5) MCS-P, and 6) MCHS-P.

	MRCS	Cluster	Districts	LLR	p-value	Obs	RR of each category
Default	50	1	7	48.655	<0.001	933	(0.68, 1.24, 1.45, 1.16)
		2	10	38.363	<0.001	1,200	(0.98, 1.60, 0.71, 0.91)
		3	25	40.119	<0.001	3,096	(1.19, 0.70, 0.87, 1.10)
SCIC ₁ , SCIC ₂	10	1	4	50.148	<0.001	501	(0.57, 1.24, 1.59, 1.47)
		2	6	37.323	<0.001	798	(0.91, 1.76, 0.72, 0.96)
		3	5	28.589	<0.001	694	(1.30, 0.77, 0.67, 0.75)
		4	1	19.396	<0.001	126	(0.87, 0.27, 1.83, 0.43)
		5	2	19.119	<0.001	237	(1.40, 0.69, 0.55, 0.61)
		6	3	17.032	0.015	385	(0.76, 1.00, 1.50, 0.80)
Elbow	4	1	2	26.842	<0.001	240	(0.55, 1.25, 1.67, 1.07)
		2	3	22.751	<0.001	274	(0.83, 2.01, 0.72, 0.87)
		3	1	19.396	<0.001	126	(0.87, 0.27, 1.83, 0.43)
		4	2	23.128	<0.001	318	(1.38, 0.80, 0.53, 0.63)
		5	2	19.119	<0.001	237	(1.40, 0.69, 0.55, 0.61)
		6	2	17.539	0.002	269	(0.75, 0.72, 1.57, 1.46)
		7	3	15.558	0.016	322	(1.28, 0.98, 0.63, 0.45)
		8	2	13.309	0.017	220	(1.12, 1.45, 0.49, 1.09)
		9	1	12.712	0.025	108	(0.65, 0.69, 1.82, 1.19)
		10	2	12.000	0.046	299	(0.72, 1.44, 1.20, 1.18)
MCS-P, MCHS-P	2	1	1	19.396	<0.001	126	(0.87, 0.27, 1.83, 0.43)
		2	1	19.383	<0.001	130	(0.51, 1.15, 1.83, 0.99)
		3	1	19.061	<0.001	116	(0.48, 1.04, 1.73, 2.09)
		4	1	14.115	0.011	139	(0.67, 2.05, 0.92, 1.06)
		5	2	12.870	0.022	135	(1.00, 1.89, 0.51, 0.68)
		6	1	12.712	0.025	108	(0.65, 0.69, 1.82, 1.19)
		7	1	11.991	0.039	109	(0.73, 1.32, 1.49, 0.00)

Districts; number of districts, LLR; log-likelihood ratio, Obs; number of observations, RR; relative risk

Chapter 5

Discussion and Conclusion

To select the optimal MRCS value instead of using the default setting, several optimization criteria have been developed such as the Gini coefficient (Han et al., 2016; Kim and Jung, 2017; Yoo and Jung, 2018; Lee et al., 2021), MCS-P (Ma et al., 2017), MCHS-P (Wang et al., 2020), and Elbow method (Meysami et al., 2021). However, the Gini coefficient for the multinomial model has not been developed. The other optimization criteria (i.e., MCS-P, MCHS-P and Elbow method) have been developed and evaluated only for the Poisson model. Thus, we have proposed the SCIC to choose the optimal MRCS value for the multinomial-based spatial scan statistic.

We have evaluated the performance of the proposed method (i.e., SCICs) in various scenarios. In many of the scenarios including the two heterogeneous clusters setting, similar to real cases, both results for the multinomial-and ordinal- model robustly showed that: 1) the SCICs mostly selected the MRCS value, which was the same as the size of the true cluster as the optimal MRCS over 1000 iterations, and 2) the detection accuracy at the optimal MRCS outperformed that of the default setting. We have also evaluated the performance of the existing methods (i.e., Gini, Elbow, MCS-P, and MCHS-P) under various scenarios. For each scenario, the overall detection accuracy using the proposed methods was comparable to those of other existing methods. This might be because these

methods are defined based on the likelihood.

However, the existing methods have some limitations. The Gini coefficient cannot be applied to all probability models because it is difficult to clearly define it for the multinomial model. The Elbow method assumes that the sum of the LRT statistic of the significant clusters is monotonically increasing as a function of the MRCS values. However, for example, if multiple significant clusters are detected when using small MRCS values, the sum of the LRT statistic of the significant clusters may first increase and then decrease further. In this case, the elbow point may not be found properly. The MCS-P and MCHS-P should differently define the union log-likelihood ratio test statistic for each probability model. Moreover, the MCHS-P has a long computation time because it needs to calculate the spatial contiguity matrix.

The SCICs can be easily extended to all probability models because it is defined based on the likelihood. Furthermore, the SCICs are computationally efficient because the criteria are directly calculated without any further modification of the test statistics. Therefore, we believe that it would be helpful to use the SCICs when selecting the optimal value of the MRCS for the multinomial- and ordinal- based spatial scan statistic. Using the SCICs, we may identify more meaningful and interpretable clusters than when using the default setting. We believe that it appears to be more appropriate to use the SCIC_1 among the two versions of the SCIC because both SCICs include regional information (e.g., area-level code). The SCIC_1 includes more information, for geographic information and number of cases (e.g., disease outbreaks) in each area. We obtained the simulation results of the multinomial



model, which showed that the SCIC₁ had better performance in terms of the PPV. However, in the simulation results of the ordinal model, the overall sensitivity and PPV between the SCIC₁ and SCIC₂ were similar in the case of the single cluster setting. The overall sensitivity of SCIC₂ was slightly higher in the case of the two clusters setting. However, the differences in overall sensitivity between the SCIC₁ and SCIC₂ are small in values, which do not appear to be significant.

Future studies will evaluate the performance of the SCIC for other probability models as well. We recommend future studies that develop the R package to implement the proposed method for practical research purposes.

Bibliography

Kulldorff M. A spatial scan statistic. *Communications in Statistics-Theory and methods* 1997; 26 (6):1481–1496.

Kulldorff M, Huang L, Konty K. A scan statistic for continuous data based on the normal probability model. *International Journal of Health Geographics* 2009; 8:58.

Huang L, Tiwari RC, Zou Z, Kulldorff M, Feuer EJ. Weighted normal spatial scan statistic for heterogeneous population data. *Journal of the American Statistical Association* 2009; 104(487):886–898.

Cook AJ, Gold DR, Li Y. Spatial cluster detection for censored outcome data. *Biometrics* 2007; 63 (2):540–549.

Jung I, Kulldorff M, Klassen AC. A spatial scan statistic for ordinal data. *Statistics in Medicine* 2007; 26 (7):1594–1607.

Jung I, Kulldorff M, Richard OJ. A spatial scan statistic for multinomial data. *Statistics in Medicine* 2010; 29(18):1910.

Ribeiro SHR, Costa MA. Optimal selection of the spatial scan parameters for cluster detection: a simulation study. *Spatial and Spatio-temporal Epidemiology* 2012; 3(2):107–120.

Han J, Zhu L, Kulldorff M, Hostovich S, Stinchcomb DG, Tatalovich Z, Lewis DR, Feuer EJ. Using Gini coefficient to determining optimal cluster reporting sizes for spatial scan statistics. *International Journal of Health Geographics* 2016; 15:27.

Gini C. Variabilità e mutabilità. Reprinted in Memorie di metodologica statistica (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi 1912.

Kim S, Jung I. Optimizing the maximum reported cluster size in the spatial scan statistic for ordinal data. *PLoS ONE* 2017; 12:e0182234.

Yoo H, Jung I. Optimizing the maximum reported cluster size for normal-based spatial scan statistics. *Communications for Statistical Applications and Methods* 2018; 25:373–383.

Lee S, Moon J, Jung I. Optimizing the maximum reported cluster size in the spatial scan statistic for survival data. *International Journal of Health Geographics* 2021; 20:33.

Meysami M, French JP, Lipner EM. Estimating the optimal population upper bound for scan methods in retrospective disease surveillance. *Biometrical Journal* 2021; 63:1633-1651.

Ma Y, Yin F, Zhang T, Zhou XA, Li X. Selection of the Maximum Spatial Cluster Size of the Spatial Scan Statistic by Using the Maximum Clustering Set-Proportion Statistic. *PLoS ONE* 2017; 11(1): e0147918.

Wang W, Zhang T, Yin F, Xiao X, Chen S, Zhang X, Li X, Ma Y. Using the maximum clustering heterogeneous set-proportion to select the maximum window size for the spatial scan statistic. *Scientific Reports* 2020; 10:4900.

Schwarz G. Estimating the dimension of a model. *The Annals of Statistics* 1978, 6:461–464.

Thorndike RL, Who belongs in the family? *Psychometrika* 1953; 18:267–276.

Neath AA, Cavanaugh JE. The Bayesian information criterion: background, derivation, and applications. *WIREs Computational Statistics* 2012; 4:199-203.

Lorenz MO. Methods of measuring the concentration of wealth. *Publications of the American Statistical Association* 1905; 9(70):209–219.

Costa MA, Assunção RM, Kulldorff M. Constrained spanning tree algorithms for irregularly-shaped spatial clustering. *Computational Statistics and Data Analysis* 2012; 56:1771-1783.

Kleinman K. Rsatscan: Tools, Classes, and Methods for Interfacing with SaTScan Stand-Alone Software. <https://CRAN.R-project.org/package=rsatscan/>, 2015.

Kulldorff M, Huang L, Pickle L, Duczmal L. An elliptic spatial scan statistic. *Statistics in Medicine* 2006; 25:3929–3943.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Appendix

Table A1. Multinomial model: Simulation results for the true cluster model (A) and alternative hypothesis (1) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting	
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall		
SCIC ₁	Freq	NA	NA	2	10	47	33	543	180	63	35	25	2	7	3	2	1	1	954	954	
	Sen	NA	NA	0.400	0.480	0.562	0.685	0.940	0.969	0.971	0.994	0.976	1.000	0.971	1.000	1.000	1.000	1.000	0.918	0.945	
	PPV	NA	NA	1.000	0.940	0.933	0.968	0.976	0.791	0.637	0.536	0.371	0.279	0.224	0.176	0.156	0.167	0.135	0.871	0.826	
	Mis	NA	NA	0.043	0.042	0.035	0.025	0.007	0.022	0.049	0.069	0.130	0.188	0.248	0.377	0.391	0.362	0.464	0.025	0.042	
SCIC ₂	Freq	NA	1	3	15	44	37	519	166	65	33	27	6	11	8	4	8	7	954	954	
	Sen	NA	0.200	0.467	0.573	0.600	0.686	0.941	0.967	0.972	0.994	0.978	0.967	0.982	1.000	1.000	0.975	0.943	0.919	0.945	
	PPV	NA	1.000	1.000	0.947	0.954	0.967	0.977	0.789	0.639	0.529	0.352	0.231	0.215	0.169	0.154	0.140	0.118	0.849	0.826	
	Mis	NA	0.058	0.039	0.035	0.032	0.025	0.006	0.023	0.048	0.071	0.141	0.237	0.271	0.366	0.399	0.438	0.522	0.037	0.042	
Elbow	Freq	NA	NA	5	27	46	39	559	160	63	33	9	7	3	1	1	1	NA	954	954	
	Sen	NA	NA	0.560	0.689	0.609	0.692	0.946	0.963	0.975	0.964	1.000	0.971	0.933	1.000	1.000	1.000	1.000	NA	0.916	0.945
	PPV	NA	NA	0.933	0.963	0.954	0.947	0.972	0.784	0.641	0.476	0.300	0.231	0.169	0.116	0.147	0.167	NA	0.882	0.826	
	Mis	NA	NA	0.035	0.025	0.032	0.027	0.007	0.024	0.048	0.089	0.172	0.238	0.353	0.551	0.420	0.362	NA	0.024	0.042	
MCS-P	Freq	NA	NA	6	16	29	31	548	185	65	34	25	3	5	3	NA	3	1	954	954	
	Sen	NA	NA	0.400	0.613	0.607	0.684	0.945	0.964	0.972	0.994	0.992	0.800	1.000	0.933	NA	1.000	1.000	0.926	0.945	
	PPV	NA	NA	0.833	0.988	0.932	0.946	0.985	0.797	0.677	0.566	0.414	0.247	0.256	0.222	NA	0.163	0.161	0.882	0.826	
	Mis	NA	NA	0.048	0.029	0.032	0.026	0.005	0.021	0.037	0.057	0.104	0.184	0.212	0.242	NA	0.372	0.377	0.021	0.042	
MCHS-P	Freq	NA	NA	2	14	24	29	542	188	70	39	28	4	6	3	1	3	1	954	954	
	Sen	NA	NA	0.700	0.614	0.633	0.710	0.943	0.962	0.974	0.995	0.993	0.950	0.900	0.933	0.800	1.000	1.000	0.932	0.945	
	PPV	NA	NA	1.000	0.943	0.939	0.957	0.977	0.786	0.656	0.540	0.401	0.249	0.223	0.222	0.129	0.163	0.161	0.865	0.826	
	Mis	NA	NA	0.022	0.032	0.031	0.024	0.006	0.023	0.043	0.066	0.111	0.217	0.258	0.242	0.406	0.372	0.377	0.025	0.042	

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A2. Multinomial model: Simulation results for the true cluster model (A) and alternative hypothesis (2) using elliptical windows.

		Maximum reported cluster size (MRCS)																			Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall		
SCIC ₁	Freq	NA	1	2	10	40	52	458	146	90	42	33	2	4	4	3	NA	2	889	889	
	Sen	NA	0.000	0.200	0.360	0.575	0.658	0.924	0.962	0.971	0.981	0.970	1.000	1.000	0.450	1.000	NA	1.000	0.898	0.933	
	PPV	NA	0.000	0.500	0.850	0.942	0.933	0.967	0.785	0.650	0.541	0.403	0.236	0.255	0.098	0.192	NA	0.143	0.844	0.796	
	Mis	NA	0.087	0.072	0.054	0.034	0.029	0.008	0.023	0.044	0.065	0.112	0.254	0.217	0.337	0.304	NA	0.435	0.029	0.046	
SCIC ₂	Freq	NA	NA	2	10	40	57	438	142	88	38	37	5	9	8	7	4	4	889	889	
	Sen	NA	NA	0.200	0.440	0.580	0.660	0.928	0.962	0.970	0.979	0.962	0.960	0.978	0.800	1.000	1.000	1.000	0.902	0.933	
	PPV	NA	NA	0.500	0.950	0.931	0.934	0.974	0.782	0.642	0.536	0.372	0.222	0.211	0.157	0.163	0.125	0.138	0.827	0.796	
	Mis	NA	NA	0.072	0.043	0.035	0.029	0.007	0.024	0.046	0.066	0.134	0.255	0.282	0.322	0.400	0.525	0.453	0.039	0.046	
Elbow	Freq	NA	NA	3	14	43	58	467	145	88	31	23	3	4	5	3	1	NA	888	889	
	Sen	NA	NA	0.333	0.514	0.595	0.672	0.927	0.957	0.961	0.968	0.948	0.933	0.950	0.760	1.000	1.000	NA	0.896	0.933	
	PPV	NA	NA	0.667	0.879	0.931	0.926	0.966	0.773	0.633	0.520	0.329	0.203	0.182	0.152	0.161	0.094	NA	0.848	0.796	
	Mis	NA	NA	0.058	0.041	0.034	0.028	0.008	0.026	0.047	0.071	0.163	0.271	0.315	0.316	0.440	0.696	NA	0.031	0.046	
MCS-P	Freq	NA	1	4	8	29	40	438	158	89	52	45	3	7	7	5	1	2	889	889	
	Sen	NA	0.000	0.000	0.350	0.586	0.610	0.933	0.965	0.980	0.985	0.978	1.000	0.943	0.686	1.000	1.000	1.000	0.911	0.933	
	PPV	NA	0.000	0.000	0.688	0.917	0.812	0.982	0.804	0.691	0.561	0.423	0.474	0.267	0.152	0.191	0.179	0.143	0.829	0.796	
	Mis	NA	0.087	0.101	0.062	0.035	0.040	0.006	0.020	0.034	0.058	0.101	0.121	0.193	0.300	0.307	0.333	0.435	0.030	0.046	
MCHS-P	Freq	NA	1	1	6	26	37	436	162	95	54	47	2	7	7	5	1	2	889	889	
	Sen	NA	0.000	0.000	0.400	0.592	0.643	0.933	0.964	0.981	0.985	0.970	1.000	0.943	0.686	1.000	1.000	1.000	0.920	0.933	
	PPV	NA	0.000	0.000	0.750	0.915	0.850	0.976	0.788	0.672	0.557	0.415	0.294	0.267	0.152	0.191	0.179	0.143	0.823	0.796	
	Mis	NA	0.087	0.101	0.056	0.035	0.036	0.007	0.023	0.038	0.059	0.105	0.174	0.193	0.300	0.307	0.333	0.435	0.031	0.046	

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A3. Multinomial model: Simulation results for the true cluster model (A) and alternative hypothesis (3) using elliptical windows.

	1%	2%	3%	4%	5%	6%	8%	Maximum reported cluster size (MRCS)													Default Setting
								10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall			
SCIC ₁	Freq	NA	1	4	4	50	53	435	165	78	46	25	10	2	1	3	NA	2	879	879	
	Sen	NA	0.200	0.350	0.400	0.572	0.679	0.921	0.953	0.956	0.983	0.968	0.900	1.000	1.000	1.000	NA	1.000	0.895	0.928	
	PPV	NA	1.000	0.917	1.000	0.938	0.936	0.969	0.767	0.653	0.543	0.364	0.281	0.265	0.238	0.137	NA	0.152	0.845	0.802	
	Mis	NA	0.058	0.051	0.043	0.035	0.027	0.009	0.027	0.042	0.066	0.140	0.177	0.210	0.232	0.488	NA	0.406	0.030	0.043	
SCIC ₂	Freq	NA	1	5	8	46	51	428	166	73	43	25	13	2	7	2	3	6	879	879	
	Sen	NA	0.200	0.400	0.650	0.578	0.686	0.921	0.949	0.953	0.981	0.992	0.908	1.000	1.000	1.000	0.933	0.933	0.897	0.928	
	PPV	NA	1.000	0.933	0.950	0.940	0.954	0.968	0.763	0.649	0.541	0.364	0.267	0.265	0.187	0.113	0.133	0.126	0.834	0.802	
	Mis	NA	0.058	0.046	0.029	0.034	0.026	0.009	0.028	0.043	0.066	0.141	0.193	0.210	0.333	0.572	0.444	0.481	0.036	0.043	
Elbow	Freq	NA	NA	7	11	52	51	446	165	81	32	16	13	NA	3	1	NA	NA	878	879	
	Sen	NA	NA	0.543	0.691	0.592	0.686	0.926	0.950	0.951	0.981	0.988	0.908	NA	1.000	1.000	NA	NA	0.896	0.928	
	PPV	NA	NA	0.952	0.945	0.929	0.917	0.966	0.767	0.626	0.522	0.342	0.243	NA	0.194	0.185	NA	NA	0.850	0.802	
	Mis	NA	NA	0.035	0.026	0.034	0.028	0.009	0.027	0.049	0.073	0.150	0.222	NA	0.343	0.319	NA	NA	0.028	0.043	
MCS -P	Freq	NA	1	7	13	34	35	435	159	93	52	27	12	4	3	1	1	2	879	879	
	Sen	NA	0.000	0.343	0.523	0.535	0.640	0.925	0.965	0.963	0.985	0.978	0.917	1.000	1.000	1.000	1.000	1.000	0.904	0.928	
	PPV	NA	0.000	0.714	0.738	0.865	0.888	0.975	0.801	0.673	0.559	0.426	0.311	0.276	0.234	0.185	0.161	0.152	0.837	0.802	
	Mis	NA	0.087	0.056	0.047	0.040	0.033	0.008	0.020	0.037	0.059	0.098	0.155	0.192	0.237	0.319	0.377	0.406	0.027	0.043	
MCHS -P	Freq	NA	NA	2	10	29	32	435	164	94	54	32	15	5	3	1	1	2	879	879	
	Sen	NA	NA	0.600	0.720	0.586	0.681	0.926	0.963	0.964	0.985	0.950	0.920	1.000	1.000	1.000	1.000	1.000	0.919	0.928	
	PPV	NA	NA	1.000	0.940	0.928	0.940	0.968	0.785	0.664	0.552	0.393	0.295	0.249	0.234	0.185	0.161	0.152	0.830	0.802	
	Mis	NA	NA	0.029	0.025	0.034	0.027	0.008	0.023	0.040	0.061	0.116	0.171	0.241	0.237	0.319	0.377	0.406	0.030	0.043	

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A4. Multinomial model: Simulation results for the true cluster model (A) and alternative hypothesis (4) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall	
SCIC ₁	Freq	NA	4	6	11	44	48	452	160	94	43	23	7	7	2	1	NA	1	903	903
	Sen	NA	0.200	0.300	0.400	0.536	0.654	0.916	0.948	0.962	0.958	0.922	1.000	0.857	0.800	1.000	NA	1.000	0.883	0.920
	PPV	NA	1.000	0.750	1.000	0.866	0.953	0.966	0.773	0.655	0.526	0.375	0.256	0.224	0.159	0.192	NA	0.147	0.842	0.785
	Mis	NA	0.058	0.058	0.043	0.040	0.028	0.009	0.026	0.041	0.070	0.124	0.219	0.224	0.341	0.304	NA	0.420	0.029	0.051
SCIC ₂	Freq	NA	4	7	12	42	42	431	157	91	41	29	12	12	8	3	6	6	903	903
	Sen	NA	0.200	0.314	0.433	0.538	0.662	0.916	0.946	0.954	0.956	0.931	0.967	0.883	0.925	0.933	0.967	0.933	0.885	0.920
	PPV	NA	1.000	0.786	0.983	0.877	0.976	0.965	0.770	0.652	0.520	0.345	0.242	0.192	0.151	0.164	0.137	0.123	0.814	0.785
	Mis	NA	0.058	0.056	0.042	0.039	0.026	0.009	0.027	0.042	0.073	0.146	0.228	0.281	0.402	0.353	0.447	0.495	0.043	0.051
Elbow	Freq	NA	3	8	13	52	50	458	148	96	35	21	10	6	2	1	NA	NA	903	903
	Sen	NA	0.200	0.400	0.554	0.573	0.672	0.923	0.942	0.950	0.937	0.914	0.960	0.800	1.000	1.000	NA	NA	0.883	0.920
	PPV	NA	1.000	0.875	0.985	0.873	0.948	0.967	0.764	0.639	0.504	0.302	0.246	0.138	0.119	0.192	NA	NA	0.842	0.785
	Mis	NA	0.058	0.047	0.033	0.038	0.027	0.008	0.028	0.046	0.077	0.170	0.225	0.353	0.536	0.304	NA	NA	0.032	0.051
MCS-P	Freq	NA	3	7	3	30	41	445	158	106	49	37	7	8	1	2	3	3	903	903
	Sen	NA	0.133	0.286	0.800	0.593	0.620	0.916	0.952	0.958	0.967	0.951	0.971	0.875	1.000	1.000	1.000	1.000	0.901	0.920
	PPV	NA	0.667	0.714	0.933	0.878	0.902	0.970	0.795	0.676	0.545	0.422	0.322	0.223	0.217	0.189	0.183	0.147	0.830	0.785
	Mis	NA	0.068	0.060	0.019	0.036	0.033	0.009	0.022	0.037	0.062	0.102	0.151	0.232	0.261	0.312	0.333	0.420	0.030	0.051
MCHS-P	Freq	NA	1	3	3	29	37	443	162	109	52	38	8	8	1	3	3	3	903	903
	Sen	NA	0.200	0.267	0.800	0.593	0.654	0.916	0.951	0.960	0.965	0.953	0.975	0.875	1.000	0.933	1.000	1.000	0.908	0.920
	PPV	NA	1.000	0.667	0.933	0.874	0.931	0.965	0.786	0.666	0.533	0.403	0.312	0.229	0.217	0.169	0.183	0.147	0.822	0.785
	Mis	NA	0.058	0.063	0.019	0.036	0.030	0.009	0.024	0.039	0.066	0.110	0.161	0.221	0.261	0.343	0.333	0.420	0.032	0.051

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A5. Multinomial model: Simulation results for the true cluster model (C) and alternative hypothesis (1) using elliptical windows.

	Freq	Maximum reported cluster size (MRCS)																			Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall		
SCIC ₁	Freq	NA	3	1	6	10	34	148	213	396	73	49	19	6	10	8	5	2	983	983	
	Sen	NA	0.100	0.200	0.300	0.330	0.426	0.576	0.677	0.738	0.714	0.739	0.779	0.850	0.830	0.988	0.720	0.900	0.683	0.715	
	PPV	NA	1.000	1.000	1.000	0.980	0.955	0.953	0.931	0.926	0.703	0.533	0.463	0.415	0.336	0.289	0.195	0.259	0.869	0.738	
	Mis	NA	0.130	0.116	0.101	0.099	0.086	0.067	0.056	0.050	0.095	0.147	0.175	0.215	0.317	0.395	0.496	0.406	0.076	0.128	
SCIC ₂	Freq	NA	4	1	9	24	58	146	191	342	63	50	31	15	23	8	15	3	983	983	
	Sen	NA	0.100	0.200	0.344	0.496	0.571	0.615	0.691	0.743	0.725	0.724	0.739	0.787	0.791	0.925	0.767	0.833	0.692	0.715	
	PPV	NA	1.000	1.000	0.950	0.978	0.963	0.939	0.925	0.922	0.702	0.480	0.374	0.318	0.274	0.240	0.201	0.185	0.833	0.738	
	Mis	NA	0.130	0.116	0.098	0.075	0.065	0.063	0.055	0.050	0.096	0.173	0.230	0.290	0.370	0.457	0.484	0.565	0.093	0.128	
Elbow	Freq	NA	1	2	14	61	104	179	212	294	60	26	20	5	5	NA	NA	NA	983	983	
	Sen	NA	0.100	0.300	0.379	0.579	0.614	0.617	0.683	0.737	0.703	0.692	0.680	0.700	0.920	NA	NA	NA	0.670	0.715	
	PPV	NA	1.000	1.000	0.968	0.970	0.945	0.928	0.916	0.907	0.634	0.413	0.325	0.239	0.248	NA	NA	NA	0.873	0.738	
	Mis	NA	0.130	0.101	0.092	0.064	0.062	0.065	0.058	0.054	0.117	0.211	0.264	0.365	0.446	NA	NA	NA	0.075	0.128	
MCS-P	Freq	NA	4	4	17	45	97	145	180	340	59	31	23	17	11	4	2	4	983	983	
	Sen	NA	0.025	0.050	0.335	0.509	0.630	0.626	0.699	0.755	0.703	0.726	0.822	0.865	0.936	0.975	0.950	1.000	0.693	0.715	
	PPV	NA	0.250	0.250	0.871	0.984	0.960	0.946	0.943	0.945	0.759	0.625	0.534	0.462	0.444	0.380	0.311	0.299	0.890	0.738	
	Mis	NA	0.152	0.159	0.101	0.072	0.058	0.061	0.050	0.043	0.076	0.104	0.138	0.167	0.194	0.236	0.312	0.341	0.064	0.128	
MCHS-P	Freq	NA	1	1	14	42	93	140	183	329	67	39	32	20	13	4	2	3	983	983	
	Sen	NA	0.100	0.200	0.400	0.543	0.617	0.644	0.711	0.745	0.691	0.723	0.772	0.845	0.900	0.975	0.650	1.000	0.699	0.715	
	PPV	NA	1.000	1.000	0.976	0.980	0.937	0.933	0.936	0.925	0.709	0.574	0.463	0.424	0.414	0.380	0.194	0.301	0.866	0.738	
	Mis	NA	0.130	0.116	0.089	0.068	0.062	0.060	0.051	0.048	0.093	0.127	0.180	0.206	0.217	0.236	0.500	0.338	0.072	0.128	

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A6. Multinomial model: Simulation results for the true cluster model (C) and alternative hypothesis (2) using elliptical windows.

		Maximum reported cluster size (MRCS)																			Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall		
SCIC ₁	Freq	NA	NA	2	7	30	41	131	180	319	92	58	26	20	8	9	14	1	938	938	
	Sen	NA	NA	0.100	0.271	0.313	0.417	0.507	0.642	0.692	0.676	0.714	0.808	0.825	0.900	0.811	0.864	1.000	0.640	0.693	
	PPV	NA	NA	0.500	0.943	0.956	0.902	0.931	0.916	0.920	0.664	0.582	0.460	0.400	0.353	0.262	0.256	0.313	0.829	0.712	
	Mis	NA	NA	0.145	0.110	0.102	0.092	0.079	0.062	0.055	0.106	0.122	0.183	0.222	0.274	0.367	0.410	0.319	0.090	0.135	
SCIC ₂	Freq	NA	NA	1	7	31	52	131	171	289	82	59	34	28	17	14	20	2	938	938	
	Sen	NA	NA	0.200	0.271	0.352	0.465	0.523	0.640	0.701	0.674	0.710	0.765	0.814	0.800	0.736	0.825	0.750	0.646	0.693	
	PPV	NA	NA	1.000	0.943	0.960	0.908	0.922	0.911	0.921	0.651	0.552	0.402	0.389	0.282	0.220	0.233	0.224	0.800	0.712	
	Mis	NA	NA	0.116	0.110	0.096	0.085	0.078	0.063	0.054	0.110	0.137	0.219	0.227	0.344	0.428	0.440	0.428	0.104	0.135	
Elbow	Freq	NA	NA	2	12	45	76	160	187	288	80	40	23	15	2	4	3	NA	937	938	
	Sen	NA	NA	0.100	0.333	0.420	0.507	0.527	0.634	0.692	0.665	0.715	0.722	0.733	0.750	0.700	0.967	NA	0.620	0.693	
	PPV	NA	NA	0.500	0.967	0.948	0.906	0.906	0.902	0.911	0.627	0.536	0.364	0.324	0.292	0.257	0.333	NA	0.840	0.712	
	Mis	NA	NA	0.145	0.099	0.088	0.080	0.079	0.066	0.057	0.120	0.147	0.246	0.288	0.319	0.330	0.285	NA	0.087	0.135	
MCS-P	Freq	NA	1	6	13	42	69	113	169	292	75	62	31	25	9	15	12	4	938	938	
	Sen	NA	0.000	0.067	0.238	0.374	0.512	0.536	0.649	0.695	0.661	0.715	0.787	0.840	0.922	0.907	0.958	0.950	0.644	0.693	
	PPV	NA	0.000	0.250	0.769	0.907	0.916	0.940	0.920	0.927	0.713	0.605	0.516	0.458	0.401	0.343	0.332	0.288	0.829	0.712	
	Mis	NA	0.159	0.159	0.118	0.096	0.078	0.073	0.060	0.053	0.090	0.110	0.139	0.168	0.213	0.265	0.286	0.348	0.084	0.135	
MCHS-P	Freq	NA	NA	1	9	33	63	121	166	288	80	65	32	35	10	16	14	5	938	938	
	Sen	NA	NA	0.200	0.322	0.430	0.510	0.550	0.644	0.692	0.664	0.717	0.781	0.811	0.870	0.894	0.886	0.860	0.655	0.693	
	PPV	NA	NA	1.000	0.956	0.954	0.891	0.923	0.902	0.910	0.686	0.591	0.481	0.404	0.376	0.335	0.303	0.257	0.808	0.712	
	Mis	NA	NA	0.116	0.101	0.087	0.081	0.074	0.064	0.057	0.100	0.118	0.166	0.217	0.233	0.274	0.318	0.386	0.092	0.135	

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A7. Multinomial model: Simulation results for the true cluster model (C) and alternative hypothesis (3) using elliptical windows.

		Maximum reported cluster size (MRCS)																			Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall		
SCIC ₁	Freq	NA	NA	3	15	36	37	133	181	322	85	65	33	10	9	13	7	3	952	952	
	Sen	NA	NA	0.100	0.247	0.325	0.403	0.506	0.625	0.714	0.661	0.720	0.742	0.830	0.844	0.885	0.786	1.000	0.635	0.696	
	PPV	NA	NA	0.667	0.933	0.951	0.946	0.934	0.927	0.922	0.666	0.546	0.437	0.359	0.344	0.299	0.198	0.323	0.834	0.699	
	Mis	NA	NA	0.140	0.112	0.100	0.092	0.078	0.063	0.052	0.105	0.142	0.194	0.270	0.275	0.350	0.532	0.304	0.091	0.139	
SCIC ₂	Freq	NA	NA	3	19	45	47	122	178	302	71	63	43	16	12	16	11	4	952	952	
	Sen	NA	NA	0.100	0.268	0.369	0.462	0.511	0.631	0.718	0.669	0.722	0.719	0.763	0.792	0.825	0.773	0.975	0.637	0.696	
	PPV	NA	NA	0.667	0.937	0.936	0.951	0.929	0.913	0.916	0.646	0.509	0.392	0.292	0.304	0.258	0.182	0.298	0.808	0.699	
	Mis	NA	NA	0.140	0.109	0.095	0.083	0.078	0.065	0.053	0.112	0.159	0.222	0.322	0.307	0.399	0.563	0.344	0.103	0.139	
Elbow	Freq	NA	NA	2	21	61	73	157	197	285	65	50	28	9	1	2	NA	NA	951	952	
	Sen	NA	NA	0.100	0.290	0.438	0.503	0.512	0.631	0.704	0.657	0.712	0.743	0.678	0.700	0.950	NA	NA	0.613	0.696	
	PPV	NA	NA	0.500	0.933	0.948	0.936	0.910	0.907	0.894	0.617	0.442	0.362	0.210	0.333	0.365	NA	NA	0.840	0.699	
	Mis	NA	NA	0.145	0.106	0.085	0.078	0.080	0.066	0.060	0.120	0.192	0.251	0.417	0.246	0.246	NA	NA	0.089	0.139	
MCS-P	Freq	NA	NA	4	13	41	69	105	167	309	80	66	34	19	16	13	8	8	952	952	
	Sen	NA	NA	0.050	0.277	0.446	0.488	0.517	0.637	0.721	0.644	0.732	0.785	0.863	0.925	0.977	0.950	0.988	0.657	0.696	
	PPV	NA	NA	0.250	0.831	0.962	0.936	0.925	0.916	0.932	0.697	0.616	0.518	0.439	0.420	0.373	0.326	0.312	0.833	0.699	
	Mis	NA	NA	0.159	0.113	0.082	0.079	0.078	0.062	0.049	0.094	0.106	0.138	0.185	0.197	0.243	0.293	0.319	0.083	0.139	
MCHS-P	Freq	NA	NA	2	12	38	67	102	155	306	83	76	42	25	14	13	9	8	952	952	
	Sen	NA	NA	0.100	0.292	0.455	0.503	0.539	0.630	0.714	0.655	0.737	0.740	0.800	0.929	0.977	0.889	0.988	0.661	0.696	
	PPV	NA	NA	0.500	0.900	0.967	0.934	0.921	0.899	0.915	0.666	0.578	0.455	0.383	0.417	0.373	0.299	0.312	0.807	0.699	
	Mis	NA	NA	0.145	0.107	0.082	0.077	0.076	0.066	0.054	0.104	0.126	0.183	0.234	0.199	0.243	0.337	0.319	0.092	0.139	

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A8. Multinomial model: Simulation results for the true cluster model (C) and alternative hypothesis (4) using elliptical windows.

	1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Maximum reported cluster size (MRCS)		Default Setting
																		Overall		
SCIC ₁	Freq	NA	3	6	9	30	34	156	170	298	79	62	38	20	13	17	9	5	949	949
	Sen	NA	0.067	0.217	0.267	0.343	0.371	0.510	0.638	0.704	0.662	0.703	0.803	0.835	0.808	0.894	0.889	0.980	0.639	0.707
	PPV	NA	0.667	0.917	1.000	0.958	0.911	0.940	0.931	0.913	0.668	0.542	0.460	0.406	0.302	0.294	0.247	0.237	0.819	0.649
	Mis	NA	0.140	0.118	0.106	0.099	0.097	0.077	0.060	0.055	0.103	0.141	0.181	0.220	0.314	0.357	0.436	0.499	0.097	0.164
SCIC ₂	Freq	NA	2	6	10	35	49	149	147	246	74	66	44	38	27	23	24	9	949	949
	Sen	NA	0.100	0.217	0.290	0.391	0.459	0.521	0.641	0.708	0.676	0.692	0.786	0.768	0.763	0.874	0.808	0.944	0.648	0.707
	PPV	NA	1.000	0.917	1.000	0.957	0.928	0.933	0.927	0.906	0.645	0.504	0.425	0.321	0.261	0.270	0.210	0.221	0.766	0.649
	Mis	NA	0.130	0.118	0.103	0.092	0.084	0.076	0.060	0.057	0.111	0.159	0.205	0.296	0.360	0.386	0.485	0.514	0.124	0.164
Elbow	Freq	NA	2	6	17	57	76	186	178	240	81	52	30	18	2	3	NA	NA	948	949
	Sen	NA	0.100	0.183	0.376	0.430	0.492	0.530	0.638	0.700	0.657	0.694	0.743	0.772	0.900	0.900	NA	NA	0.612	0.707
	PPV	NA	1.000	1.000	0.957	0.950	0.917	0.921	0.913	0.888	0.609	0.469	0.355	0.295	0.410	0.277	NA	NA	0.829	0.649
	Mis	NA	0.130	0.118	0.093	0.087	0.081	0.077	0.063	0.062	0.124	0.177	0.256	0.333	0.203	0.386	NA	NA	0.093	0.164
MCS-P	Freq	NA	2	4	19	47	59	150	148	245	79	67	45	32	12	22	10	8	949	949
	Sen	NA	0.100	0.100	0.289	0.423	0.531	0.559	0.647	0.714	0.652	0.716	0.784	0.863	0.917	0.968	0.950	0.975	0.657	0.707
	PPV	NA	1.000	0.500	0.753	0.945	0.944	0.953	0.933	0.934	0.709	0.605	0.522	0.462	0.404	0.376	0.327	0.294	0.824	0.649
	Mis	NA	0.130	0.145	0.114	0.087	0.073	0.069	0.058	0.050	0.090	0.110	0.136	0.166	0.209	0.238	0.291	0.344	0.086	0.164
MCHS-P	Freq	NA	1	2	12	40	53	139	150	238	83	77	54	41	18	23	11	7	949	949
	Sen	NA	0.100	0.300	0.367	0.450	0.528	0.564	0.639	0.706	0.648	0.705	0.776	0.805	0.783	0.939	0.955	0.971	0.663	0.707
	PPV	NA	1.000	1.000	0.850	0.967	0.926	0.935	0.911	0.911	0.667	0.565	0.477	0.403	0.331	0.360	0.325	0.294	0.785	0.649
	Mis	NA	0.130	0.101	0.100	0.083	0.075	0.071	0.062	0.056	0.104	0.130	0.169	0.216	0.269	0.255	0.295	0.344	0.099	0.164

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A9. Multinomial model: Simulation results for the true cluster model (D) and alternative hypothesis (1) using elliptical windows.

	1%	2%	3%	4%	5%	6%	8%	Maximum reported cluster size (MRCS)												Default Setting
								10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall		
SCIC ₁	Freq	NA	NA	NA	3	15	26	397	143	74	38	47	108	43	24	14	25	31	988	988
	Sen	NA	NA	NA	0.200	0.400	0.469	0.789	0.858	0.868	0.766	0.813	0.963	0.965	0.929	0.886	0.980	0.903	0.829	0.835
	PPV	NA	NA	NA	1.000	0.949	0.944	0.978	0.852	0.756	0.614	0.589	0.582	0.462	0.333	0.235	0.235	0.182	0.774	0.568
	Mis	NA	NA	NA	0.116	0.091	0.081	0.034	0.043	0.061	0.114	0.125	0.125	0.195	0.323	0.438	0.503	0.599	0.106	0.241
SCIC ₂	Freq	NA	NA	NA	1	13	26	384	147	76	45	36	61	36	32	28	40	63	988	988
	Sen	NA	NA	NA	0.200	0.431	0.477	0.801	0.900	0.878	0.840	0.850	0.944	0.883	0.859	0.821	0.875	0.892	0.834	0.835
	PPV	NA	NA	NA	1.000	0.920	0.949	0.973	0.843	0.749	0.558	0.535	0.454	0.331	0.270	0.221	0.211	0.190	0.720	0.568
	Mis	NA	NA	NA	0.116	0.090	0.080	0.033	0.042	0.062	0.129	0.152	0.204	0.296	0.377	0.442	0.502	0.563	0.143	0.241
Elbow	Freq	NA	NA	NA	3	27	33	598	160	66	37	35	16	10	2	1	NA	NA	988	988
	Sen	NA	NA	NA	0.467	0.441	0.494	0.826	0.873	0.885	0.827	0.846	0.894	0.830	0.750	1.000	NA	NA	0.816	0.835
	PPV	NA	NA	NA	1.000	0.951	0.944	0.966	0.827	0.694	0.517	0.443	0.378	0.293	0.187	0.208	NA	NA	0.870	0.568
	Mis	NA	NA	NA	0.077	0.086	0.079	0.031	0.048	0.078	0.147	0.195	0.250	0.336	0.493	0.551	NA	NA	0.058	0.241
MCS-P	Freq	NA	NA	3	6	20	25	496	208	116	40	25	16	17	14	1	NA	1	988	988
	Sen	NA	NA	0.067	0.383	0.380	0.496	0.816	0.890	0.902	0.898	0.892	0.938	0.965	0.929	1.000	NA	1.000	0.831	0.835
	PPV	NA	NA	0.333	1.000	0.883	0.932	0.982	0.880	0.775	0.683	0.716	0.682	0.560	0.431	0.435	NA	0.294	0.891	0.568
	Mis	NA	NA	0.155	0.089	0.096	0.079	0.029	0.033	0.052	0.075	0.071	0.080	0.123	0.194	0.188	NA	0.348	0.044	0.241
MCHS-P	Freq	NA	NA	NA	2	11	15	440	215	124	53	51	27	26	15	6	2	1	988	988
	Sen	NA	NA	NA	0.450	0.473	0.453	0.808	0.899	0.902	0.885	0.851	0.948	0.885	0.893	0.567	0.550	1.000	0.841	0.835
	PPV	NA	NA	NA	1.000	0.894	0.887	0.978	0.859	0.767	0.622	0.568	0.512	0.418	0.452	0.198	0.298	0.294	0.841	0.568
	Mis	NA	NA	NA	0.080	0.087	0.092	0.031	0.037	0.055	0.101	0.131	0.166	0.217	0.198	0.420	0.326	0.348	0.060	0.241

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A10. Multinomial model: Simulation results for the true cluster model (D) and alternative hypothesis (2) using elliptical windows.

		Maximum reported cluster size (MRCS)																	Default Setting	
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%		
SCIC ₁	Freq	NA	NA	3	8	25	36	371	115	76	46	44	114	59	24	8	9	18	956	956
	Sen	NA	NA	0.200	0.175	0.324	0.414	0.702	0.782	0.776	0.728	0.805	0.928	0.959	0.921	0.838	0.944	0.950	0.753	0.778
	PPV	NA	NA	1.000	0.875	0.920	0.980	0.969	0.827	0.704	0.623	0.598	0.595	0.480	0.358	0.260	0.193	0.195	0.778	0.616
	Mis	NA	NA	0.116	0.125	0.103	0.087	0.047	0.056	0.080	0.108	0.119	0.112	0.177	0.272	0.388	0.576	0.581	0.100	0.202
SCIC ₂	Freq	NA	NA	4	7	23	37	341	122	86	46	34	75	52	29	33	22	45	956	956
	Sen	NA	NA	0.175	0.171	0.357	0.430	0.728	0.785	0.814	0.785	0.788	0.904	0.929	0.852	0.824	0.805	0.836	0.760	0.778
	PPV	NA	NA	1.000	0.857	0.913	0.973	0.962	0.825	0.710	0.590	0.519	0.544	0.407	0.284	0.217	0.179	0.180	0.723	0.616
	Mis	NA	NA	0.120	0.126	0.098	0.085	0.044	0.056	0.075	0.120	0.158	0.141	0.235	0.350	0.454	0.551	0.569	0.138	0.202
Elbow	Freq	NA	NA	2	8	30	40	470	146	86	45	46	46	20	11	3	2	NA	955	956
	Sen	NA	NA	0.200	0.175	0.350	0.420	0.731	0.771	0.806	0.760	0.796	0.889	0.900	0.718	0.933	0.500	NA	0.729	0.778
	PPV	NA	NA	1.000	0.875	0.924	0.968	0.962	0.809	0.679	0.553	0.505	0.510	0.357	0.254	0.201	0.238	NA	0.824	0.616
	Mis	NA	NA	0.116	0.125	0.099	0.087	0.044	0.061	0.086	0.136	0.163	0.165	0.284	0.357	0.546	0.290	NA	0.081	0.202
MCS-P	Freq	NA	1	5	9	16	29	433	152	105	50	34	58	38	15	6	4	1	956	956
	Sen	NA	0.000	0.040	0.178	0.275	0.397	0.715	0.820	0.839	0.758	0.747	0.916	0.947	0.953	0.883	0.925	1.000	0.750	0.778
	PPV	NA	0.000	0.200	0.778	0.729	0.947	0.977	0.848	0.747	0.634	0.584	0.614	0.524	0.436	0.372	0.325	0.303	0.832	0.616
	Mis	NA	0.159	0.162	0.127	0.118	0.091	0.044	0.047	0.064	0.097	0.114	0.098	0.136	0.186	0.237	0.290	0.333	0.068	0.202
MCHS-P	Freq	NA	NA	1	5	11	26	376	155	120	61	51	66	44	20	11	5	4	956	956
	Sen	NA	NA	0.200	0.160	0.300	0.400	0.711	0.819	0.853	0.774	0.763	0.898	0.909	0.910	0.745	0.840	0.625	0.763	0.778
	PPV	NA	NA	1.000	0.800	0.779	0.895	0.973	0.842	0.735	0.617	0.532	0.569	0.464	0.386	0.262	0.289	0.174	0.792	0.616
	Mis	NA	NA	0.116	0.130	0.115	0.095	0.045	0.049	0.067	0.104	0.141	0.125	0.177	0.233	0.353	0.333	0.500	0.083	0.202

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A11. Multinomial model: Simulation results for the true cluster model (D) and alternative hypothesis (3) using elliptical windows.

	1%	2%	3%	4%	5%	6%	8%	Maximum reported cluster size (MRCS)												Default Setting
								10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall		
SCIC ₁	Freq	NA	1	6	4	22	41	333	96	72	45	63	140	62	20	18	15	21	959	959
	Sen	NA	0.100	0.183	0.200	0.323	0.439	0.671	0.725	0.800	0.709	0.794	0.946	0.913	0.925	0.883	0.947	0.852	0.746	0.773
	PPV	NA	1.000	1.000	1.000	0.906	0.970	0.972	0.810	0.730	0.622	0.597	0.579	0.447	0.338	0.259	0.213	0.175	0.748	0.572
	Mis	NA	0.130	0.118	0.116	0.106	0.083	0.051	0.064	0.072	0.113	0.118	0.125	0.197	0.314	0.409	0.523	0.603	0.117	0.228
SCIC ₂	Freq	NA	1	5	4	21	37	335	105	86	48	60	68	57	21	35	28	48	959	959
	Sen	NA	0.100	0.180	0.250	0.343	0.462	0.703	0.779	0.844	0.777	0.797	0.882	0.877	0.852	0.823	0.846	0.810	0.751	0.773
	PPV	NA	1.000	1.000	1.000	0.917	0.969	0.967	0.809	0.726	0.581	0.518	0.474	0.372	0.265	0.223	0.201	0.171	0.705	0.572
	Mis	NA	0.130	0.119	0.109	0.102	0.080	0.047	0.060	0.071	0.124	0.156	0.184	0.257	0.384	0.448	0.507	0.589	0.149	0.228
Elbow	Freq	NA	1	6	9	29	41	488	131	88	47	48	41	17	5	6	1	NA	958	959
	Sen	NA	0.100	0.183	0.311	0.341	0.444	0.729	0.767	0.810	0.760	0.777	0.793	0.676	0.700	1.000	1.000	NA	0.717	0.773
	PPV	NA	1.000	1.000	0.978	0.917	0.956	0.958	0.785	0.685	0.548	0.453	0.385	0.256	0.232	0.315	0.196	NA	0.817	0.572
	Mis	NA	0.130	0.118	0.101	0.101	0.084	0.045	0.066	0.087	0.143	0.185	0.242	0.346	0.368	0.360	0.594	NA	0.086	0.228
MCS-P	Freq	NA	NA	7	5	28	28	414	171	122	38	41	34	42	17	9	1	2	959	959
	Sen	NA	NA	0.114	0.180	0.311	0.421	0.720	0.796	0.841	0.766	0.756	0.903	0.921	0.906	1.000	0.900	0.500	0.746	0.773
	PPV	NA	NA	0.571	0.600	0.893	0.972	0.981	0.835	0.757	0.655	0.615	0.612	0.510	0.424	0.392	0.321	0.167	0.839	0.572
	Mis	NA	NA	0.141	0.133	0.105	0.086	0.043	0.051	0.062	0.092	0.105	0.100	0.145	0.194	0.227	0.290	0.464	0.068	0.228
MCHS-P	Freq	NA	NA	3	2	13	26	367	164	127	57	58	44	58	23	12	1	4	959	959
	Sen	NA	NA	0.200	0.250	0.331	0.404	0.712	0.817	0.836	0.781	0.771	0.857	0.881	0.835	0.800	0.900	0.500	0.758	0.773
	PPV	NA	NA	1.000	1.000	0.962	0.952	0.973	0.837	0.746	0.588	0.530	0.532	0.446	0.369	0.309	0.321	0.149	0.790	0.572
	Mis	NA	NA	0.116	0.109	0.100	0.090	0.045	0.050	0.065	0.118	0.145	0.148	0.197	0.244	0.307	0.290	0.507	0.085	0.228

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A12. Multinomial model: Simulation results for the true cluster model (D) and alternative hypothesis (4) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall	
SCIC ₁	Freq	NA	1	9	13	31	36	313	57	47	36	79	156	86	25	15	22	45	971	971
	Sen	NA	0.100	0.200	0.215	0.358	0.442	0.672	0.665	0.700	0.647	0.780	0.950	0.956	0.936	0.820	0.914	0.909	0.746	0.781
	PPV	NA	1.000	1.000	0.974	0.952	0.964	0.969	0.797	0.678	0.571	0.587	0.580	0.469	0.354	0.227	0.206	0.182	0.711	0.516
	Mis	NA	0.130	0.116	0.115	0.097	0.084	0.052	0.074	0.095	0.128	0.117	0.124	0.192	0.296	0.443	0.540	0.604	0.140	0.271
SCIC ₂	Freq	NA	NA	6	11	31	36	310	67	53	36	60	93	66	32	31	43	96	971	971
	Sen	NA	NA	0.200	0.236	0.403	0.450	0.711	0.770	0.777	0.736	0.750	0.906	0.888	0.847	0.700	0.879	0.848	0.750	0.781
	PPV	NA	NA	1.000	0.970	0.957	0.956	0.958	0.792	0.700	0.541	0.488	0.503	0.400	0.271	0.191	0.195	0.174	0.648	0.516
	Mis	NA	NA	0.116	0.112	0.091	0.085	0.048	0.065	0.084	0.136	0.165	0.170	0.242	0.374	0.465	0.550	0.597	0.191	0.271
Elbow	Freq	NA	1	9	17	38	47	500	88	62	39	70	58	31	4	4	3	NA	971	971
	Sen	NA	0.200	0.211	0.218	0.358	0.477	0.731	0.722	0.787	0.767	0.734	0.831	0.868	0.950	0.550	1.000	NA	0.706	0.781
	PPV	NA	1.000	1.000	0.951	0.929	0.952	0.956	0.741	0.643	0.506	0.446	0.432	0.348	0.327	0.291	0.415	NA	0.803	0.516
	Mis	NA	0.116	0.114	0.115	0.099	0.081	0.045	0.081	0.100	0.155	0.189	0.217	0.292	0.322	0.333	0.213	NA	0.094	0.271
MCS-P	Freq	NA	3	6	12	34	38	428	143	81	44	43	61	49	15	9	4	1	971	971
	Sen	NA	0.067	0.183	0.225	0.321	0.447	0.719	0.801	0.830	0.768	0.742	0.941	0.969	0.967	0.878	0.975	1.000	0.741	0.781
	PPV	NA	0.667	0.833	0.868	0.912	0.940	0.977	0.831	0.744	0.618	0.603	0.641	0.533	0.497	0.373	0.394	0.313	0.836	0.516
	Mis	NA	0.140	0.123	0.118	0.103	0.085	0.043	0.052	0.066	0.101	0.110	0.087	0.132	0.157	0.238	0.232	0.319	0.069	0.271
MCHS-P	Freq	NA	NA	2	7	23	33	374	133	95	56	67	82	60	20	10	6	3	971	971
	Sen	NA	NA	0.200	0.257	0.322	0.424	0.723	0.820	0.827	0.788	0.727	0.855	0.920	0.880	0.780	0.717	0.667	0.753	0.781
	PPV	NA	NA	1.000	1.000	0.972	0.868	0.969	0.824	0.718	0.592	0.498	0.525	0.478	0.389	0.319	0.241	0.184	0.774	0.516
	Mis	NA	NA	0.116	0.108	0.100	0.095	0.044	0.053	0.076	0.112	0.164	0.152	0.173	0.241	0.293	0.394	0.512	0.092	0.271

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A13. Multinomial model: Simulation results for the true cluster model (D) and alternative hypothesis (5) using elliptical windows.

	1%	2%	3%	4%	5%	6%	8%	Maximum reported cluster size (MRCS)												Default Setting
								10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall		
SCIC ₁	Freq	NA	NA	4	6	20	27	437	150	89	36	40	71	47	21	8	17	13	986	986
	Sen	NA	NA	0.175	0.183	0.395	0.441	0.742	0.848	0.812	0.731	0.785	0.939	0.966	0.957	0.738	0.935	0.769	0.778	0.793
	PPV	NA	NA	0.875	0.833	0.947	0.962	0.971	0.857	0.732	0.630	0.556	0.591	0.479	0.334	0.206	0.217	0.164	0.807	0.629
	Mis	NA	NA	0.123	0.126	0.091	0.084	0.041	0.043	0.071	0.105	0.131	0.116	0.180	0.325	0.440	0.519	0.585	0.090	0.198
SCIC ₂	Freq	NA	NA	2	6	19	22	405	151	84	45	37	48	43	27	23	39	35	986	986
	Sen	NA	NA	0.150	0.183	0.416	0.373	0.752	0.850	0.851	0.784	0.830	0.900	0.926	0.870	0.796	0.872	0.780	0.785	0.793
	PPV	NA	NA	0.750	0.917	0.936	0.959	0.966	0.842	0.734	0.576	0.482	0.485	0.398	0.278	0.207	0.208	0.176	0.747	0.629
	Mis	NA	NA	0.130	0.121	0.088	0.094	0.041	0.046	0.068	0.121	0.170	0.178	0.245	0.364	0.461	0.504	0.553	0.126	0.198
Elbow	Freq	NA	1	2	4	27	29	582	149	71	41	39	26	9	3	3	NA	NA	986	986
	Sen	NA	0.100	0.150	0.325	0.433	0.438	0.777	0.838	0.800	0.715	0.838	0.792	0.911	1.000	0.667	NA	NA	0.767	0.793
	PPV	NA	1.000	0.750	1.000	0.951	0.964	0.962	0.834	0.682	0.511	0.470	0.411	0.379	0.240	0.195	NA	NA	0.859	0.629
	Mis	NA	0.130	0.130	0.098	0.085	0.084	0.038	0.049	0.087	0.148	0.178	0.221	0.258	0.459	0.444	NA	NA	0.066	0.198
MCS-P	Freq	NA	NA	8	9	25	29	477	189	117	35	31	27	26	7	3	2	1	986	986
	Sen	NA	NA	0.063	0.233	0.308	0.431	0.762	0.862	0.846	0.814	0.803	0.919	0.954	0.986	0.667	1.000	1.000	0.774	0.793
	PPV	NA	NA	0.313	0.806	0.821	0.896	0.981	0.870	0.765	0.658	0.612	0.598	0.529	0.464	0.246	0.364	0.303	0.867	0.629
	Mis	NA	NA	0.156	0.121	0.108	0.088	0.037	0.038	0.059	0.089	0.106	0.109	0.134	0.168	0.343	0.254	0.333	0.056	0.198
MCHS-P	Freq	NA	NA	1	4	12	13	426	199	135	56	43	37	36	13	5	3	3	986	986
	Sen	NA	NA	0.100	0.200	0.308	0.385	0.758	0.863	0.862	0.816	0.770	0.881	0.922	0.923	0.580	0.833	0.600	0.795	0.793
	PPV	NA	NA	0.500	0.875	0.864	0.962	0.973	0.853	0.759	0.605	0.513	0.522	0.461	0.384	0.207	0.292	0.169	0.824	0.629
	Mis	NA	NA	0.145	0.123	0.109	0.091	0.039	0.042	0.060	0.107	0.150	0.155	0.186	0.241	0.394	0.333	0.507	0.070	0.198

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A14. Multinomial model: Simulation results for the true cluster model (D) and alternative hypothesis (6) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall	
SCIC ₁	Freq	NA	1	1	8	24	29	415	123	94	26	40	94	48	26	12	17	25	983	983
	Sen	NA	0.100	0.200	0.225	0.379	0.428	0.739	0.803	0.817	0.692	0.785	0.931	0.975	0.935	0.708	0.876	0.872	0.772	0.790
	PPV	NA	1.000	1.000	1.000	0.958	0.994	0.968	0.823	0.728	0.589	0.582	0.582	0.486	0.369	0.185	0.196	0.175	0.783	0.607
	Mis	NA	0.130	0.116	0.112	0.094	0.083	0.042	0.054	0.072	0.117	0.123	0.121	0.176	0.294	0.493	0.536	0.602	0.104	0.214
SCIC ₂	Freq	NA	1	1	5	20	25	382	132	96	34	37	58	52	30	20	33	57	983	983
	Sen	NA	0.100	0.200	0.200	0.355	0.456	0.741	0.827	0.833	0.797	0.859	0.893	0.929	0.877	0.725	0.882	0.844	0.782	0.790
	PPV	NA	1.000	1.000	1.000	0.958	0.988	0.967	0.819	0.723	0.563	0.473	0.453	0.377	0.295	0.194	0.205	0.181	0.722	0.607
	Mis	NA	0.130	0.116	0.116	0.098	0.081	0.042	0.053	0.072	0.127	0.179	0.199	0.264	0.362	0.475	0.511	0.571	0.143	0.214
Elbow	Freq	NA	1	1	3	25	40	538	147	89	35	47	38	11	3	NA	4	NA	982	983
	Sen	NA	0.100	0.200	0.200	0.364	0.518	0.767	0.829	0.783	0.803	0.800	0.824	0.891	1.000	NA	0.625	NA	0.761	0.790
	PPV	NA	1.000	1.000	1.000	0.938	0.966	0.966	0.804	0.677	0.518	0.432	0.411	0.342	0.320	NA	0.503	NA	0.841	0.607
	Mis	NA	0.130	0.116	0.116	0.099	0.074	0.039	0.056	0.090	0.147	0.195	0.223	0.306	0.372	NA	0.243	NA	0.073	0.214
MCS-P	Freq	NA	1	3	7	22	28	481	202	111	41	23	17	29	14	NA	1	3	983	983
	Sen	NA	0.100	0.000	0.100	0.323	0.375	0.755	0.852	0.849	0.832	0.774	0.906	0.986	0.986	NA	1.000	0.667	0.774	0.790
	PPV	NA	1.000	0.000	0.500	0.869	0.935	0.983	0.858	0.768	0.662	0.624	0.606	0.524	0.463	NA	0.357	0.208	0.870	0.607
	Mis	NA	0.130	0.174	0.149	0.106	0.094	0.038	0.042	0.059	0.087	0.103	0.101	0.137	0.169	NA	0.261	0.425	0.056	0.214
MCHS-P	Freq	NA	1	NA	4	16	21	416	202	132	57	40	27	35	17	5	3	7	983	983
	Sen	NA	0.100	NA	0.250	0.319	0.395	0.750	0.838	0.857	0.840	0.818	0.885	0.926	0.935	0.680	0.800	0.529	0.784	0.790
	PPV	NA	1.000	NA	0.781	0.833	0.976	0.978	0.835	0.754	0.624	0.503	0.507	0.478	0.452	0.194	0.246	0.151	0.823	0.607
	Mis	NA	0.130	NA	0.123	0.111	0.090	0.040	0.048	0.062	0.100	0.162	0.159	0.181	0.195	0.443	0.401	0.520	0.073	0.214

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A15. Multinomial model: Simulation results for the true cluster model (D) and alternative hypothesis (7) using elliptical windows.

	1%	2%	3%	4%	5%	6%	8%	Maximum reported cluster size (MRCS)												Default Setting
								10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall		
SCIC ₁	Freq	NA	1	3	7	18	30	457	154	94	52	45	49	32	14	7	9	19	991	991
	Sen	NA	0.100	0.200	0.243	0.361	0.500	0.759	0.846	0.823	0.842	0.851	0.914	0.941	0.829	0.786	0.822	0.811	0.783	0.798
	PPV	NA	1.000	1.000	1.000	0.981	0.980	0.969	0.852	0.738	0.612	0.551	0.564	0.508	0.319	0.222	0.201	0.172	0.821	0.631
	Mis	NA	0.130	0.116	0.110	0.093	0.074	0.039	0.044	0.069	0.105	0.137	0.137	0.158	0.306	0.445	0.515	0.588	0.084	0.195
SCIC ₂	Freq	NA	NA	2	5	13	30	412	140	86	54	48	47	32	33	19	32	38	991	991
	Sen	NA	NA	0.200	0.320	0.354	0.500	0.761	0.837	0.851	0.856	0.863	0.887	0.931	0.873	0.816	0.803	0.800	0.792	0.798
	PPV	NA	NA	1.000	1.000	0.974	0.969	0.964	0.848	0.731	0.583	0.484	0.421	0.370	0.285	0.220	0.194	0.178	0.747	0.631
	Mis	NA	NA	0.116	0.099	0.095	0.075	0.040	0.046	0.069	0.115	0.172	0.216	0.269	0.348	0.443	0.505	0.561	0.125	0.195
Elbow	Freq	NA	NA	1	10	26	36	583	143	77	41	35	28	2	4	2	3	NA	991	991
	Sen	NA	NA	0.200	0.430	0.369	0.536	0.790	0.838	0.821	0.788	0.829	0.825	0.850	0.875	0.750	0.933	NA	0.778	0.798
	PPV	NA	NA	1.000	1.000	0.981	0.951	0.962	0.825	0.664	0.529	0.424	0.384	0.472	0.229	0.172	0.404	NA	0.859	0.631
	Mis	NA	NA	0.116	0.083	0.093	0.073	0.036	0.051	0.092	0.138	0.205	0.242	0.159	0.438	0.543	0.275	NA	0.066	0.195
MCS-P	Freq	NA	NA	4	11	28	38	489	188	107	51	21	19	22	8	1	1	3	991	991
	Sen	NA	NA	0.050	0.264	0.321	0.526	0.775	0.839	0.818	0.884	0.867	0.905	0.936	1.000	1.000	1.000	0.900	0.777	0.798
	PPV	NA	NA	0.250	0.841	0.860	0.922	0.979	0.857	0.753	0.678	0.634	0.618	0.558	0.551	0.417	0.357	0.284	0.876	0.631
	Mis	NA	NA	0.159	0.115	0.106	0.074	0.035	0.043	0.064	0.077	0.094	0.101	0.127	0.134	0.203	0.261	0.343	0.054	0.195
MCHS-P	Freq	NA	NA	NA	5	9	16	443	199	129	77	37	29	23	16	5	1	2	991	991
	Sen	NA	NA	NA	0.240	0.322	0.500	0.772	0.844	0.848	0.892	0.835	0.862	0.878	0.869	0.680	1.000	0.900	0.804	0.798
	PPV	NA	NA	NA	0.900	0.963	0.930	0.972	0.845	0.747	0.641	0.511	0.478	0.443	0.382	0.236	0.357	0.281	0.831	0.631
	Mis	NA	NA	NA	0.116	0.100	0.080	0.037	0.045	0.064	0.091	0.155	0.181	0.202	0.258	0.388	0.261	0.348	0.067	0.195

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A16. Multinomial model: Simulation results for the true cluster model (E) and alternative hypothesis (1) using elliptical windows.

		Maximum reported cluster size (MRCS)																	Default Setting	
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%		
SCIC ₁	Freq	NA	NA	2	2	19	32	545	224	75	41	13	8	3	4	6	3	4	981	981
	Sen	NA	NA	0.200	0.200	0.363	0.388	0.751	0.837	0.843	0.844	0.777	0.750	0.567	0.800	0.833	0.833	0.950	0.762	0.781
	PPV	NA	NA	1.000	1.000	1.000	0.962	0.968	0.843	0.708	0.595	0.425	0.338	0.260	0.311	0.327	0.267	0.277	0.879	0.784
	Mis	NA	NA	0.116	0.116	0.092	0.092	0.040	0.046	0.077	0.113	0.187	0.245	0.300	0.286	0.275	0.362	0.388	0.060	0.093
SCIC ₂	Freq	NA	NA	2	3	20	32	511	202	76	51	32	20	14	6	5	2	5	981	981
	Sen	NA	NA	0.200	0.367	0.370	0.400	0.751	0.839	0.841	0.853	0.859	0.830	0.836	0.833	0.840	0.800	0.920	0.768	0.781
	PPV	NA	NA	1.000	1.000	0.982	0.958	0.969	0.838	0.698	0.582	0.407	0.336	0.300	0.306	0.306	0.286	0.298	0.848	0.784
	Mis	NA	NA	0.116	0.092	0.093	0.091	0.040	0.048	0.080	0.118	0.207	0.261	0.303	0.300	0.304	0.319	0.354	0.071	0.093
Elbow	Freq	NA	NA	1	5	22	36	591	188	58	32	32	5	4	1	2	3	NA	980	981
	Sen	NA	NA	0.200	0.620	0.436	0.444	0.766	0.821	0.819	0.831	0.834	0.660	0.825	0.900	0.900	0.833	NA	0.764	0.781
	PPV	NA	NA	1.000	1.000	0.935	0.946	0.963	0.815	0.656	0.543	0.393	0.300	0.302	0.391	0.354	0.294	NA	0.873	0.784
	Mis	NA	NA	0.116	0.055	0.088	0.086	0.039	0.055	0.095	0.135	0.213	0.270	0.308	0.217	0.254	0.314	NA	0.061	0.093
MCS-P	Freq	NA	NA	2	9	18	37	505	237	88	46	12	4	3	2	5	3	10	981	981
	Sen	NA	NA	0.100	0.444	0.367	0.405	0.751	0.849	0.865	0.878	0.900	0.750	0.733	0.850	0.840	0.900	0.950	0.772	0.781
	PPV	NA	NA	0.500	0.889	0.896	0.923	0.973	0.854	0.759	0.674	0.539	0.417	0.360	0.356	0.316	0.310	0.292	0.883	0.784
	Mis	NA	NA	0.145	0.085	0.099	0.092	0.040	0.043	0.059	0.079	0.129	0.181	0.227	0.246	0.287	0.304	0.342	0.055	0.093
MCHS-P	Freq	NA	NA	NA	6	16	33	473	242	102	67	13	6	3	2	5	4	9	981	981
	Sen	NA	NA	NA	0.583	0.419	0.418	0.744	0.845	0.864	0.864	0.815	0.717	0.733	0.850	0.840	0.800	0.956	0.776	0.781
	PPV	NA	NA	NA	1.000	0.900	0.868	0.967	0.852	0.751	0.646	0.508	0.389	0.360	0.356	0.316	0.310	0.293	0.865	0.784
	Mis	NA	NA	NA	0.060	0.095	0.099	0.042	0.043	0.061	0.090	0.143	0.203	0.227	0.246	0.287	0.286	0.341	0.059	0.093

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A17. Multinomial model: Simulation results for the true cluster model (E) and alternative hypothesis (2) using elliptical windows.

		Maximum reported cluster size (MRCS)																	Default Setting	
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%		
SCIC ₁	Freq	NA	NA	4	12	28	52	452	168	83	41	21	12	7	6	5	4	8	903	903
	Sen	NA	NA	0.125	0.175	0.314	0.356	0.665	0.729	0.716	0.722	0.624	0.608	0.714	0.717	0.800	0.950	0.963	0.650	0.679
	PPV	NA	NA	0.750	0.806	0.964	0.911	0.956	0.814	0.696	0.558	0.428	0.312	0.390	0.303	0.337	0.291	0.298	0.841	0.764
	Mis	NA	NA	0.134	0.128	0.102	0.100	0.054	0.064	0.087	0.121	0.184	0.255	0.246	0.290	0.287	0.351	0.335	0.081	0.106
SCIC ₂	Freq	NA	NA	4	12	27	51	423	167	85	51	31	23	9	7	3	2	8	903	903
	Sen	NA	NA	0.125	0.208	0.315	0.361	0.670	0.728	0.734	0.753	0.700	0.770	0.678	0.700	0.867	1.000	0.963	0.662	0.679
	PPV	NA	NA	0.750	0.889	0.963	0.921	0.957	0.808	0.698	0.552	0.434	0.355	0.263	0.278	0.305	0.328	0.298	0.824	0.764
	Mis	NA	NA	0.134	0.120	0.102	0.099	0.053	0.065	0.085	0.125	0.183	0.238	0.322	0.308	0.319	0.297	0.335	0.085	0.106
Elbow	Freq	NA	NA	4	9	31	51	477	172	70	36	30	11	2	1	2	3	NA	899	903
	Sen	NA	NA	0.225	0.211	0.352	0.375	0.673	0.728	0.686	0.686	0.707	0.764	0.750	0.800	0.800	1.000	NA	0.654	0.679
	PPV	NA	NA	1.000	0.852	0.924	0.904	0.948	0.804	0.677	0.500	0.405	0.335	0.317	0.320	0.322	0.338	NA	0.846	0.764
	Mis	NA	NA	0.112	0.121	0.101	0.099	0.054	0.066	0.095	0.147	0.202	0.242	0.268	0.275	0.275	0.285	NA	0.077	0.106
MCS-P	Freq	NA	NA	8	9	33	45	414	187	90	55	24	10	3	3	4	5	13	903	903
	Sen	NA	NA	0.138	0.122	0.318	0.371	0.659	0.739	0.750	0.745	0.675	0.760	0.933	0.800	0.925	0.960	0.962	0.663	0.679
	PPV	NA	NA	0.625	0.463	0.894	0.908	0.960	0.827	0.731	0.604	0.490	0.394	0.575	0.345	0.356	0.324	0.297	0.837	0.764
	Mis	NA	NA	0.136	0.150	0.105	0.098	0.053	0.060	0.075	0.104	0.147	0.196	0.140	0.251	0.254	0.296	0.337	0.077	0.106
MCHS-P	Freq	NA	NA	5	5	24	43	399	191	97	68	29	12	4	3	5	5	13	903	903
	Sen	NA	NA	0.220	0.160	0.338	0.360	0.652	0.742	0.747	0.738	0.672	0.717	0.950	0.800	0.820	0.960	0.962	0.671	0.679
	PPV	NA	NA	1.000	0.667	0.924	0.855	0.955	0.820	0.724	0.587	0.460	0.370	0.545	0.345	0.374	0.324	0.297	0.823	0.764
	Mis	NA	NA	0.113	0.136	0.102	0.105	0.055	0.061	0.077	0.111	0.165	0.210	0.149	0.251	0.235	0.296	0.337	0.080	0.106

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A18. Multinomial model: Simulation results for the true cluster model (E) and alternative hypothesis (3) using elliptical windows.

	1%	2%	3%	4%	5%	6%	8%	Maximum reported cluster size (MRCS)												Default Setting
								10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall		
SCIC ₁	Freq	NA	2	2	11	33	43	440	193	67	47	29	12	10	3	5	5	9	911	911
	Sen	NA	0.100	0.200	0.227	0.306	0.391	0.652	0.747	0.755	0.687	0.669	0.675	0.670	0.700	0.900	0.840	0.956	0.656	0.680
	PPV	NA	1.000	1.000	0.921	0.975	0.906	0.962	0.800	0.698	0.558	0.406	0.355	0.280	0.304	0.349	0.394	0.293	0.837	0.756
	Mis	NA	0.130	0.116	0.116	0.102	0.095	0.055	0.065	0.084	0.128	0.195	0.242	0.293	0.280	0.258	0.267	0.341	0.083	0.112
SCIC ₂	Freq	NA	2	2	12	33	40	423	185	67	53	33	23	15	4	4	6	9	911	911
	Sen	NA	0.100	0.200	0.233	0.309	0.388	0.650	0.739	0.769	0.692	0.715	0.783	0.740	0.750	0.825	0.833	0.956	0.660	0.680
	PPV	NA	1.000	1.000	0.956	0.970	0.903	0.963	0.793	0.694	0.526	0.395	0.353	0.289	0.298	0.271	0.365	0.293	0.817	0.756
	Mis	NA	0.130	0.116	0.114	0.102	0.096	0.055	0.067	0.085	0.141	0.202	0.250	0.299	0.297	0.359	0.295	0.341	0.090	0.112
Elbow	Freq	NA	1	1	9	37	42	486	174	57	52	27	14	3	2	1	3	NA	909	911
	Sen	NA	0.100	0.200	0.322	0.338	0.395	0.662	0.742	0.749	0.648	0.659	0.764	0.633	0.600	0.900	0.900	NA	0.654	0.680
	PPV	NA	1.000	1.000	0.936	0.981	0.903	0.955	0.791	0.647	0.489	0.337	0.322	0.298	0.265	0.360	0.306	NA	0.842	0.756
	Mis	NA	0.130	0.116	0.105	0.097	0.095	0.054	0.067	0.101	0.155	0.231	0.268	0.275	0.304	0.246	0.309	NA	0.081	0.112
MCS-P	Freq	NA	1	4	9	37	41	422	196	83	54	27	6	3	4	4	7	13	911	911
	Sen	NA	0.100	0.150	0.233	0.316	0.398	0.648	0.744	0.788	0.737	0.674	0.717	0.600	0.850	0.925	0.914	0.931	0.664	0.680
	PPV	NA	1.000	0.750	0.736	0.914	0.948	0.973	0.829	0.730	0.620	0.449	0.351	0.314	0.485	0.356	0.313	0.287	0.850	0.756
	Mis	NA	0.130	0.130	0.124	0.103	0.091	0.054	0.058	0.072	0.101	0.165	0.222	0.246	0.178	0.254	0.302	0.347	0.077	0.112
MCHS-P	Freq	NA	1	1	7	31	36	396	207	92	61	36	10	6	4	4	6	13	911	911
	Sen	NA	0.100	0.200	0.300	0.323	0.394	0.641	0.743	0.780	0.720	0.686	0.730	0.683	0.850	0.925	0.900	0.931	0.670	0.680
	PPV	NA	1.000	1.000	0.875	0.959	0.926	0.968	0.815	0.718	0.585	0.450	0.361	0.365	0.485	0.356	0.306	0.287	0.830	0.756
	Mis	NA	0.130	0.116	0.112	0.102	0.094	0.055	0.061	0.077	0.116	0.166	0.225	0.237	0.178	0.254	0.309	0.347	0.081	0.112

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A19. Multinomial model: Simulation results for the true cluster model (E) and alternative hypothesis (4) using elliptical windows.

		Maximum reported cluster size (MRCS)																	Default Setting	
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%		
SCIC ₁	Freq	NA	2	9	22	47	56	461	154	69	32	21	13	8	1	11	5	6	917	917
	Sen	NA	0.100	0.200	0.195	0.328	0.400	0.667	0.725	0.714	0.638	0.643	0.731	0.575	0.500	0.845	0.980	0.983	0.634	0.669
	PPV	NA	1.000	1.000	0.847	0.921	0.946	0.948	0.790	0.691	0.479	0.414	0.407	0.257	0.208	0.284	0.336	0.251	0.840	0.738
	Mis	NA	0.130	0.116	0.123	0.102	0.091	0.055	0.069	0.088	0.154	0.200	0.226	0.319	0.348	0.354	0.284	0.442	0.086	0.123
SCIC ₂	Freq	NA	2	8	21	43	57	428	149	65	41	37	23	19	2	10	5	7	917	917
	Sen	NA	0.100	0.213	0.248	0.316	0.411	0.665	0.724	0.717	0.673	0.762	0.739	0.674	0.750	0.830	0.940	0.957	0.644	0.669
	PPV	NA	1.000	1.000	0.925	0.933	0.949	0.949	0.778	0.689	0.481	0.403	0.358	0.257	0.260	0.253	0.336	0.248	0.812	0.738
	Mis	NA	0.130	0.114	0.112	0.102	0.089	0.055	0.072	0.089	0.153	0.208	0.257	0.328	0.333	0.399	0.278	0.439	0.096	0.123
Elbow	Freq	NA	2	8	21	45	64	484	131	68	35	32	9	9	NA	4	3	NA	915	917
	Sen	NA	0.100	0.213	0.248	0.342	0.427	0.671	0.724	0.712	0.657	0.781	0.533	0.578	NA	0.825	0.633	NA	0.635	0.669
	PPV	NA	1.000	1.000	0.878	0.932	0.939	0.941	0.776	0.650	0.429	0.402	0.243	0.221	NA	0.280	0.222	NA	0.837	0.738
	Mis	NA	0.130	0.114	0.115	0.099	0.088	0.055	0.073	0.102	0.174	0.205	0.319	0.353	NA	0.370	0.295	NA	0.085	0.123
MCS-P	Freq	NA	4	11	19	46	52	428	158	91	46	16	4	5	5	11	11	10	917	917
	Sen	NA	0.075	0.155	0.232	0.313	0.402	0.669	0.750	0.762	0.713	0.594	0.750	0.780	0.720	0.909	0.964	0.950	0.653	0.669
	PPV	NA	0.750	0.727	0.779	0.929	0.962	0.963	0.823	0.726	0.582	0.434	0.661	0.398	0.316	0.335	0.331	0.289	0.847	0.738
	Mis	NA	0.138	0.130	0.121	0.103	0.089	0.052	0.059	0.075	0.112	0.171	0.109	0.203	0.264	0.277	0.289	0.348	0.079	0.123
MCHS-P	Freq	NA	1	8	15	36	52	406	162	107	56	22	6	7	5	12	11	11	917	917
	Sen	NA	0.100	0.188	0.240	0.292	0.390	0.665	0.743	0.769	0.666	0.641	0.683	0.671	0.720	0.892	0.964	0.945	0.659	0.669
	PPV	NA	1.000	0.875	0.820	0.898	0.920	0.955	0.816	0.714	0.538	0.419	0.412	0.333	0.316	0.325	0.364	0.313	0.822	0.738
	Mis	NA	0.130	0.121	0.119	0.109	0.096	0.054	0.062	0.078	0.131	0.177	0.200	0.251	0.264	0.287	0.269	0.327	0.085	0.123

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A20. Multinomial model: Simulation results for the true cluster model (E) and alternative hypothesis (5) using elliptical windows.

	1%	2%	3%	4%	5%	6%	8%	Maximum reported cluster size (MRCS)												Default Setting
								10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall		
SCIC ₁	Freq	NA	1	1	7	22	34	521	213	65	44	13	10	9	2	NA	3	5	950	950
	Sen	NA	0.100	0.200	0.229	0.327	0.406	0.717	0.789	0.845	0.770	0.869	0.720	0.644	0.400	NA	0.900	0.880	0.722	0.738
	PPV	NA	1.000	1.000	1.000	0.964	0.937	0.962	0.825	0.717	0.573	0.495	0.343	0.286	0.200	NA	0.280	0.252	0.869	0.784
	Mis	NA	0.130	0.116	0.112	0.100	0.090	0.046	0.056	0.072	0.121	0.159	0.239	0.283	0.312	NA	0.362	0.400	0.066	0.093
SCIC ₂	Freq	NA	1	1	6	20	31	494	204	66	55	29	19	13	3	NA	3	5	950	950
	Sen	NA	0.100	0.200	0.200	0.330	0.406	0.718	0.788	0.836	0.756	0.866	0.779	0.654	0.567	NA	0.867	0.880	0.727	0.738
	PPV	NA	1.000	1.000	1.000	0.953	0.947	0.962	0.826	0.705	0.547	0.430	0.331	0.272	0.219	NA	0.290	0.252	0.844	0.784
	Mis	NA	0.130	0.116	0.116	0.101	0.090	0.046	0.056	0.077	0.132	0.194	0.256	0.302	0.338	NA	0.329	0.400	0.074	0.093
Elbow	Freq	NA	1	NA	7	29	36	561	180	55	34	25	13	3	2	NA	3	NA	949	950
	Sen	NA	0.100	NA	0.300	0.428	0.400	0.737	0.776	0.796	0.694	0.852	0.623	0.567	0.400	NA	0.933	NA	0.721	0.738
	PPV	NA	1.000	NA	1.000	0.941	0.929	0.956	0.811	0.664	0.491	0.394	0.290	0.269	0.182	NA	0.314	NA	0.864	0.784
	Mis	NA	0.130	NA	0.101	0.088	0.092	0.044	0.061	0.091	0.156	0.215	0.271	0.290	0.326	NA	0.304	NA	0.068	0.093
MCS-P	Freq	NA	3	2	13	27	34	480	218	81	48	14	6	5	1	2	5	11	950	950
	Sen	NA	0.033	0.000	0.169	0.344	0.421	0.720	0.790	0.840	0.800	0.814	0.817	0.700	0.000	0.800	0.880	0.918	0.722	0.738
	PPV	NA	0.333	0.000	0.692	0.900	0.897	0.974	0.839	0.741	0.640	0.529	0.421	0.364	0.000	0.326	0.383	0.278	0.867	0.784
	Mis	NA	0.150	0.174	0.132	0.102	0.092	0.044	0.052	0.066	0.091	0.133	0.184	0.223	0.420	0.268	0.258	0.364	0.064	0.093
MCHS-P	Freq	NA	1	NA	8	24	27	463	222	89	68	16	8	5	1	2	5	11	950	950
	Sen	NA	0.100	NA	0.250	0.350	0.415	0.714	0.786	0.848	0.791	0.806	0.863	0.660	0.000	0.800	0.880	0.918	0.732	0.738
	PPV	NA	1.000	NA	0.938	0.856	0.921	0.964	0.831	0.744	0.616	0.534	0.425	0.319	0.000	0.326	0.300	0.278	0.854	0.784
	Mis	NA	0.130	NA	0.112	0.105	0.091	0.046	0.054	0.064	0.101	0.130	0.185	0.252	0.420	0.268	0.316	0.364	0.067	0.093

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A21. Multinomial model: Simulation results for the true cluster model (E) and alternative hypothesis (6) using elliptical windows.

	1%	2%	3%	4%	5%	6%	8%	Maximum reported cluster size (MRCS)													Default Setting
								10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall			
SCIC ₁	Freq	NA	1	2	4	27	41	502	208	76	48	15	8	6	2	4	3	5	952	952	
	Sen	NA	0.100	0.200	0.225	0.326	0.388	0.707	0.782	0.749	0.708	0.687	0.750	0.717	0.750	0.775	0.933	0.960	0.701	0.723	
	PPV	NA	1.000	1.000	1.000	0.934	0.911	0.959	0.820	0.715	0.574	0.423	0.310	0.299	0.289	0.288	0.299	0.376	0.860	0.776	
	Mis	NA	0.130	0.116	0.112	0.101	0.095	0.048	0.057	0.080	0.122	0.188	0.279	0.283	0.297	0.315	0.338	0.371	0.070	0.097	
SCIC ₂	Freq	NA	1	2	2	28	31	477	201	77	56	32	20	14	3	3	1	4	952	952	
	Sen	NA	0.100	0.200	0.200	0.336	0.432	0.704	0.779	0.753	0.738	0.734	0.800	0.800	0.767	0.733	1.000	0.950	0.709	0.723	
	PPV	NA	1.000	1.000	1.000	0.932	0.943	0.958	0.815	0.705	0.565	0.385	0.329	0.304	0.300	0.264	0.345	0.421	0.834	0.776	
	Mis	NA	0.130	0.116	0.116	0.100	0.087	0.048	0.059	0.084	0.127	0.212	0.264	0.293	0.290	0.338	0.275	0.312	0.078	0.097	
Elbow	Freq	NA	1	2	6	33	37	537	191	59	35	31	12	5	NA	2	NA	NA	951	952	
	Sen	NA	0.100	0.200	0.383	0.394	0.435	0.724	0.774	0.700	0.651	0.719	0.783	0.680	NA	0.800	NA	NA	0.704	0.723	
	PPV	NA	1.000	1.000	1.000	0.926	0.923	0.956	0.798	0.694	0.480	0.371	0.327	0.282	NA	0.296	NA	NA	0.856	0.776	
	Mis	NA	0.130	0.116	0.089	0.093	0.089	0.046	0.063	0.091	0.161	0.220	0.262	0.296	NA	0.304	NA	NA	0.070	0.097	
MCS-P	Freq	NA	NA	4	6	37	40	464	207	101	51	11	4	7	3	6	5	6	952	952	
	Sen	NA	NA	0.150	0.233	0.330	0.390	0.706	0.800	0.784	0.735	0.727	1.000	0.700	0.900	0.850	0.920	0.967	0.709	0.723	
	PPV	NA	NA	0.750	0.833	0.922	0.879	0.970	0.844	0.746	0.622	0.513	0.478	0.330	0.355	0.403	0.317	0.395	0.868	0.776	
	Mis	NA	NA	0.130	0.118	0.101	0.097	0.046	0.050	0.069	0.101	0.137	0.159	0.246	0.251	0.244	0.299	0.292	0.065	0.097	
MCHS-P	Freq	NA	NA	2	6	28	33	434	213	108	70	22	6	10	4	6	4	6	952	952	
	Sen	NA	NA	0.200	0.317	0.371	0.406	0.695	0.793	0.792	0.744	0.732	0.967	0.650	0.825	0.850	0.975	0.967	0.715	0.723	
	PPV	NA	NA	1.000	0.944	0.950	0.859	0.966	0.828	0.739	0.598	0.473	0.476	0.300	0.322	0.403	0.334	0.395	0.844	0.776	
	Mis	NA	NA	0.116	0.104	0.095	0.097	0.048	0.054	0.070	0.110	0.158	0.159	0.271	0.279	0.244	0.286	0.292	0.070	0.097	

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A22. Multinomial model: Simulation results for the true cluster model (E) and alternative hypothesis (7) using elliptical windows.

		Maximum reported cluster size (MRCS)																	Default Setting	
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%		
SCIC ₁	Freq	NA	NA	3	5	28	35	518	199	81	55	21	11	8	2	3	1	2	972	972
	Sen	NA	NA	0.200	0.200	0.404	0.406	0.732	0.796	0.811	0.749	0.800	0.600	0.663	0.500	0.800	0.800	1.000	0.727	0.751
	PPV	NA	NA	1.000	1.000	0.972	0.933	0.960	0.816	0.718	0.566	0.474	0.316	0.270	0.214	0.269	0.286	0.303	0.859	0.763
	Mis	NA	NA	0.116	0.116	0.089	0.090	0.044	0.056	0.076	0.123	0.161	0.240	0.304	0.341	0.357	0.319	0.333	0.066	0.097
SCIC ₂	Freq	NA	NA	3	6	24	30	489	192	75	60	39	24	16	5	3	2	4	972	972
	Sen	NA	NA	0.233	0.300	0.396	0.423	0.740	0.799	0.813	0.777	0.826	0.708	0.613	0.740	0.833	0.850	0.875	0.738	0.751
	PPV	NA	NA	1.000	0.943	0.960	0.936	0.960	0.816	0.707	0.563	0.451	0.315	0.235	0.256	0.251	0.275	0.239	0.828	0.763
	Mis	NA	NA	0.111	0.106	0.091	0.087	0.043	0.056	0.080	0.126	0.178	0.261	0.342	0.345	0.386	0.348	0.431	0.077	0.097
Elbow	Freq	NA	NA	1	8	34	36	568	178	63	35	29	12	5	NA	1	2	NA	972	972
	Sen	NA	NA	0.200	0.325	0.465	0.450	0.758	0.779	0.776	0.717	0.755	0.558	0.500	NA	0.500	0.400	NA	0.731	0.751
	PPV	NA	NA	1.000	0.957	0.944	0.925	0.951	0.802	0.683	0.505	0.385	0.280	0.202	NA	0.147	0.143	NA	0.858	0.763
	Mis	NA	NA	0.116	0.101	0.083	0.086	0.042	0.061	0.087	0.149	0.211	0.269	0.374	NA	0.493	0.304	NA	0.067	0.097
MCS -P	Freq	NA	2	2	11	38	46	472	197	98	58	23	14	4	3	NA	1	3	972	972
	Sen	NA	0.050	0.100	0.236	0.426	0.465	0.727	0.806	0.792	0.791	0.783	0.707	0.775	0.767	NA	0.800	1.000	0.723	0.751
	PPV	NA	0.500	0.500	0.782	0.958	0.896	0.967	0.841	0.743	0.617	0.518	0.375	0.364	0.311	NA	0.286	0.319	0.864	0.763
	Mis	NA	0.145	0.145	0.120	0.087	0.085	0.044	0.049	0.068	0.098	0.137	0.205	0.225	0.275	NA	0.319	0.309	0.063	0.097
MCHS -P	Freq	NA	NA	1	9	21	34	440	207	114	74	33	16	7	8	1	3	4	972	972
	Sen	NA	NA	0.200	0.267	0.405	0.435	0.727	0.807	0.818	0.796	0.748	0.681	0.757	0.825	1.000	0.767	0.950	0.740	0.751
	PPV	NA	NA	1.000	0.867	0.884	0.823	0.964	0.832	0.742	0.610	0.492	0.353	0.338	0.324	0.294	0.443	0.297	0.834	0.763
	Mis	NA	NA	0.116	0.114	0.097	0.097	0.044	0.052	0.068	0.103	0.148	0.221	0.248	0.283	0.348	0.193	0.337	0.068	0.097

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A23. Ordinal model: Simulation results for true cluster model (A) with an alternative hypothesis (1) using elliptical windows.

	1%	2%	3%	4%	5%	6%	8%	Maximum reported cluster size (MRCS)												Default Setting
								10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall		
SCIC ₁	Freq	1	1	3	12	56	38	571	180	66	36	21	1	3	1	NA	1	NA	991	991
	Sen	0.200	0.200	0.333	0.433	0.582	0.684	0.942	0.976	0.982	0.972	0.971	0.800	1.000	1.000	NA	1.000	NA	0.913	0.945
	PPV	1.000	1.000	0.833	0.972	0.972	0.987	0.989	0.811	0.691	0.570	0.414	0.250	0.263	0.238	NA	0.161	NA	0.903	0.883
	Mis	0.058	0.058	0.053	0.042	0.032	0.024	0.005	0.018	0.034	0.057	0.104	0.188	0.203	0.232	NA	0.377	NA	0.018	0.020
SCIC ₂	Freq	NA	3	5	23	58	43	564	170	68	31	20	2	3	NA	NA	1	NA	991	991
	Sen	NA	0.267	0.440	0.539	0.614	0.698	0.943	0.979	0.982	0.968	0.970	0.900	1.000	NA	NA	1.000	NA	0.909	0.945
	PPV	NA	1.000	1.000	0.951	0.969	0.974	0.988	0.814	0.693	0.558	0.412	0.292	0.263	NA	NA	0.161	NA	0.906	0.883
	Mis	NA	0.053	0.041	0.036	0.030	0.024	0.005	0.018	0.033	0.059	0.104	0.167	0.203	NA	NA	0.377	NA	0.017	0.020
Gini	Freq	NA	NA	NA	13	29	29	564	192	71	44	33	3	5	3	NA	3	2	991	991
	Sen	NA	NA	NA	0.785	0.717	0.759	0.946	0.979	0.989	0.982	0.970	0.933	1.000	1.000	NA	1.000	1.000	0.944	0.945
	PPV	NA	NA	NA	0.969	0.956	0.931	0.989	0.814	0.697	0.558	0.411	0.292	0.256	0.234	NA	0.163	0.156	0.881	0.883
	Mis	NA	NA	NA	0.018	0.024	0.023	0.005	0.018	0.033	0.059	0.105	0.169	0.212	0.237	NA	0.372	0.391	0.020	0.020
Elbow	Freq	NA	2	10	74	75	53	548	148	56	18	6	1	NA	NA	NA	NA	NA	991	991
	Sen	NA	0.300	0.660	0.746	0.669	0.725	0.945	0.972	0.986	0.956	0.933	0.800	NA	NA	NA	NA	NA	0.900	0.945
	PPV	NA	1.000	1.000	0.949	0.942	0.954	0.984	0.807	0.694	0.547	0.380	0.250	NA	NA	NA	NA	NA	0.922	0.883
	Mis	NA	0.051	0.025	0.022	0.028	0.024	0.005	0.019	0.033	0.061	0.116	0.188	NA	NA	NA	NA	NA	0.015	0.020
MSC-P	Freq	NA	NA	4	25	42	32	548	186	71	39	29	3	5	3	NA	2	2	991	991
	Sen	NA	NA	0.650	0.736	0.719	0.744	0.943	0.975	0.983	0.979	0.972	0.933	1.000	1.000	NA	1.000	1.000	0.933	0.945
	PPV	NA	NA	1.000	0.953	0.942	0.960	0.987	0.811	0.691	0.558	0.409	0.292	0.256	0.234	NA	0.161	0.156	0.884	0.883
	Mis	NA	NA	0.025	0.022	0.025	0.022	0.005	0.018	0.034	0.059	0.105	0.169	0.212	0.237	NA	0.377	0.391	0.020	0.020
MSHC-P	Freq	NA	1	16	44	41	33	528	176	67	41	29	3	5	3	NA	2	2	991	991
	Sen	NA	0.400	0.625	0.764	0.717	0.752	0.941	0.976	0.985	0.980	0.972	0.933	1.000	1.000	NA	1.000	1.000	0.924	0.945
	PPV	NA	1.000	1.000	0.961	0.946	0.946	0.987	0.811	0.693	0.569	0.409	0.292	0.256	0.234	NA	0.161	0.156	0.886	0.883
	Mis	NA	0.043	0.027	0.020	0.025	0.023	0.005	0.018	0.033	0.057	0.105	0.169	0.212	0.237	NA	0.377	0.391	0.021	0.020

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A24. Ordinal model: Simulation results for true cluster model (A) with an alternative hypothesis (2) using elliptical windows.

	1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Maximum reported cluster size (MRCS)		Default Setting
																		Overall		
SCIC ₁	Freq	NA	NA	3	16	55	61	504	155	84	48	32	3	4	2	3	2	NA	972	972
	Sen	NA	NA	0.400	0.400	0.564	0.682	0.927	0.965	0.971	0.975	0.981	1.000	0.950	1.000	1.000	1.000	NA	0.896	0.939
	PPV	NA	NA	1.000	0.958	0.938	0.966	0.980	0.808	0.678	0.554	0.431	0.313	0.442	0.228	0.190	0.170	NA	0.874	0.835
	Mis	NA	NA	0.043	0.045	0.035	0.025	0.007	0.019	0.036	0.060	0.096	0.159	0.152	0.246	0.309	0.355	NA	0.023	0.030
SCIC ₂	Freq	NA	NA	4	23	52	67	501	154	81	47	30	3	4	3	2	1	NA	972	972
	Sen	NA	NA	0.350	0.487	0.585	0.681	0.928	0.966	0.973	0.974	0.973	1.000	0.900	0.933	1.000	1.000	NA	0.894	0.939
	PPV	NA	NA	1.000	0.962	0.942	0.956	0.980	0.807	0.678	0.552	0.423	0.313	0.259	0.198	0.193	0.179	NA	0.876	0.835
	Mis	NA	NA	0.047	0.039	0.033	0.026	0.007	0.019	0.036	0.060	0.101	0.159	0.196	0.290	0.304	0.333	NA	0.023	0.030
Gini	Freq	NA	NA	1	8	20	28	487	184	91	60	57	9	8	6	5	7	1	972	972
	Sen	NA	NA	0.800	0.675	0.650	0.743	0.937	0.962	0.982	0.983	0.972	1.000	0.950	0.933	1.000	0.971	1.000	0.938	0.939
	PPV	NA	NA	1.000	0.975	0.948	0.942	0.987	0.804	0.690	0.558	0.427	0.331	0.266	0.202	0.191	0.168	0.161	0.835	0.835
	Mis	NA	NA	0.014	0.025	0.028	0.023	0.006	0.020	0.034	0.059	0.098	0.148	0.194	0.278	0.307	0.352	0.377	0.030	0.030
Elbow	Freq	NA	NA	7	33	62	66	527	151	76	32	11	2	3	1	1	NA	NA	972	972
	Sen	NA	NA	0.543	0.667	0.635	0.703	0.931	0.954	0.958	0.963	0.964	1.000	0.933	0.800	1.000	NA	NA	0.892	0.939
	PPV	NA	NA	1.000	0.949	0.928	0.953	0.974	0.796	0.664	0.534	0.402	0.295	0.250	0.138	0.200	NA	NA	0.891	0.835
	Mis	NA	NA	0.033	0.027	0.031	0.025	0.007	0.021	0.039	0.065	0.112	0.174	0.208	0.377	0.290	NA	NA	0.020	0.030
MSC-P	Freq	NA	NA	1	16	34	41	475	179	87	59	50	8	6	6	4	4	2	972	972
	Sen	NA	NA	0.800	0.688	0.618	0.722	0.932	0.964	0.977	0.980	0.980	1.000	0.933	0.933	1.000	0.950	1.000	0.924	0.939
	PPV	NA	NA	1.000	0.950	0.927	0.965	0.983	0.804	0.687	0.555	0.422	0.330	0.265	0.202	0.191	0.168	0.154	0.842	0.835
	Mis	NA	NA	0.014	0.026	0.032	0.023	0.006	0.020	0.035	0.059	0.101	0.149	0.193	0.278	0.308	0.344	0.399	0.029	0.030
MSHC-P	Freq	NA	NA	7	26	37	42	466	173	84	57	50	8	6	6	4	4	2	972	972
	Sen	NA	NA	0.629	0.715	0.622	0.724	0.930	0.966	0.976	0.979	0.980	1.000	0.933	0.933	1.000	0.950	1.000	0.919	0.939
	PPV	NA	NA	0.964	0.946	0.894	0.955	0.980	0.807	0.687	0.558	0.422	0.330	0.265	0.202	0.191	0.168	0.154	0.842	0.835
	Mis	NA	NA	0.029	0.025	0.033	0.023	0.007	0.019	0.035	0.059	0.101	0.149	0.193	0.278	0.308	0.344	0.399	0.029	0.030

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A25. Ordinal model: Simulation results for true cluster model (A) with an alternative hypothesis (3) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall	
SCIC ₁	Freq	NA	2	8	11	57	56	464	175	90	42	21	6	2	3	2	2	1	942	942
	Sen	NA	0.200	0.375	0.345	0.575	0.682	0.924	0.968	0.973	0.986	0.981	1.000	1.000	0.933	1.000	0.800	1.000	0.893	0.933
	PPV	NA	1.000	1.000	0.833	0.955	0.949	0.980	0.802	0.682	0.568	0.427	0.318	0.295	0.226	0.204	0.138	0.156	0.871	0.839
	Mis	NA	0.058	0.045	0.054	0.034	0.026	0.007	0.020	0.035	0.056	0.098	0.157	0.174	0.237	0.283	0.377	0.391	0.024	0.029
SCIC ₂	Freq	NA	3	8	13	52	57	472	173	87	42	20	6	1	3	2	2	1	942	942
	Sen	NA	0.200	0.375	0.400	0.592	0.677	0.925	0.963	0.972	0.986	0.990	1.000	1.000	0.933	1.000	0.800	1.000	0.893	0.933
	PPV	NA	1.000	1.000	0.905	0.979	0.959	0.980	0.796	0.683	0.563	0.430	0.318	0.313	0.226	0.204	0.138	0.156	0.875	0.839
	Mis	NA	0.058	0.045	0.048	0.031	0.026	0.007	0.021	0.035	0.058	0.097	0.157	0.159	0.237	0.283	0.377	0.391	0.023	0.029
Gini	Freq	NA	NA	2	10	17	34	464	184	101	58	38	14	5	4	3	5	3	942	942
	Sen	NA	NA	0.500	0.680	0.624	0.706	0.931	0.967	0.968	0.990	0.984	1.000	0.960	1.000	1.000	0.920	1.000	0.932	0.933
	PPV	NA	NA	0.875	0.960	0.931	0.938	0.985	0.803	0.676	0.562	0.426	0.331	0.270	0.230	0.196	0.157	0.153	0.839	0.839
	Mis	NA	NA	0.043	0.026	0.032	0.025	0.006	0.020	0.036	0.058	0.098	0.148	0.191	0.243	0.300	0.365	0.401	0.029	0.029
Elbow	Freq	1	NA	16	31	56	57	487	168	73	31	11	6	NA	2	1	2	NA	941	942
	Sen	NA	NA	0.538	0.684	0.632	0.695	0.930	0.963	0.964	0.974	0.982	1.000	NA	1.000	1.000	0.800	NA	0.894	0.933
	PPV	NA	NA	0.969	0.937	0.959	0.941	0.977	0.794	0.671	0.535	0.413	0.325	NA	0.234	0.200	0.138	NA	0.886	0.839
	Mis	NA	NA	0.035	0.026	0.029	0.026	0.007	0.022	0.037	0.065	0.103	0.152	NA	0.239	0.290	0.377	NA	0.021	0.029
MSC-P	Freq	NA	NA	5	17	38	41	451	178	101	53	29	13	4	3	4	4	1	942	942
	Sen	NA	NA	0.520	0.612	0.653	0.688	0.931	0.967	0.972	0.989	0.986	1.000	1.000	0.933	1.000	0.900	1.000	0.919	0.933
	PPV	NA	NA	0.950	0.898	0.954	0.948	0.984	0.803	0.678	0.557	0.423	0.330	0.283	0.222	0.195	0.156	0.156	0.849	0.839
	Mis	NA	NA	0.038	0.033	0.028	0.026	0.006	0.020	0.036	0.060	0.099	0.148	0.185	0.242	0.301	0.362	0.391	0.027	0.029
MSHC-P	Freq	NA	NA	5	17	33	42	458	173	103	53	29	13	4	3	4	4	1	942	942
	Sen	NA	NA	0.440	0.647	0.618	0.681	0.929	0.968	0.967	0.989	0.986	1.000	1.000	0.933	1.000	0.900	1.000	0.917	0.933
	PPV	NA	NA	1.000	0.929	0.970	0.950	0.984	0.805	0.674	0.557	0.423	0.330	0.283	0.222	0.195	0.156	0.156	0.851	0.839
	Mis	NA	NA	0.041	0.029	0.029	0.026	0.006	0.020	0.037	0.060	0.099	0.148	0.185	0.242	0.301	0.362	0.391	0.027	0.029

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A26. Ordinal model: Simulation results for true cluster model (A) with an alternative hypothesis (4) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall	
SCIC ₁	Freq	NA	4	8	15	49	55	499	158	89	40	20	7	5	1	3	NA	2	955	955
	Sen	NA	0.200	0.375	0.400	0.563	0.658	0.921	0.946	0.969	0.980	0.970	1.000	1.000	1.000	1.000	NA	1.000	0.885	0.925
	PPV	NA	1.000	0.938	0.956	0.946	0.967	0.982	0.790	0.676	0.558	0.430	0.322	0.275	0.217	0.185	NA	0.154	0.875	0.844
	Mis	NA	0.058	0.047	0.045	0.034	0.027	0.007	0.022	0.036	0.059	0.096	0.155	0.191	0.261	0.319	NA	0.399	0.024	0.029
SCIC ₂	Freq	NA	4	8	16	50	59	499	160	86	36	19	7	4	2	3	NA	2	955	955
	Sen	NA	0.200	0.400	0.425	0.576	0.668	0.922	0.944	0.967	0.978	0.968	1.000	1.000	0.900	1.000	NA	1.000	0.884	0.925
	PPV	NA	1.000	1.000	0.946	0.967	0.962	0.981	0.787	0.674	0.564	0.431	0.322	0.275	0.180	0.185	NA	0.154	0.878	0.844
	Mis	NA	0.058	0.043	0.044	0.032	0.027	0.007	0.023	0.037	0.058	0.095	0.155	0.192	0.312	0.319	NA	0.399	0.024	0.029
Gini	Freq	NA	1	2	7	22	28	508	146	106	59	46	9	6	4	5	3	3	955	955
	Sen	NA	0.200	0.600	0.686	0.627	0.671	0.924	0.962	0.979	0.963	0.978	0.956	1.000	0.950	1.000	1.000	1.000	0.925	0.925
	PPV	NA	1.000	1.000	0.886	0.917	0.973	0.987	0.802	0.688	0.549	0.418	0.318	0.276	0.207	0.187	0.169	0.153	0.845	0.844
	Mis	NA	0.058	0.029	0.031	0.032	0.025	0.007	0.020	0.034	0.062	0.102	0.155	0.191	0.275	0.316	0.357	0.401	0.029	0.029
Elbow	Freq	NA	1	15	24	65	68	514	152	78	26	3	4	2	NA	2	1	NA	955	955
	Sen	NA	0.200	0.520	0.592	0.628	0.685	0.924	0.938	0.967	0.962	1.000	1.000	0.900	NA	1.000	1.000	NA	0.879	0.925
	PPV	NA	1.000	1.000	0.939	0.942	0.948	0.979	0.783	0.662	0.546	0.455	0.342	0.212	NA	0.185	0.161	NA	0.896	0.844
	Mis	NA	0.058	0.035	0.033	0.031	0.027	0.007	0.023	0.039	0.062	0.087	0.141	0.261	NA	0.319	0.377	NA	0.020	0.029
MSC-P	Freq	NA	2	5	9	35	54	491	156	98	42	36	8	8	2	3	3	3	955	955
	Sen	NA	0.200	0.520	0.622	0.651	0.685	0.922	0.947	0.971	0.981	0.983	1.000	1.000	0.900	1.000	1.000	1.000	0.908	0.925
	PPV	NA	1.000	1.000	0.896	0.943	0.953	0.984	0.791	0.680	0.556	0.429	0.333	0.271	0.180	0.185	0.169	0.152	0.857	0.844
	Mis	NA	0.058	0.035	0.034	0.029	0.027	0.007	0.022	0.036	0.059	0.097	0.147	0.196	0.312	0.319	0.357	0.406	0.027	0.029
MSHC-P	Freq	NA	6	21	22	41	50	467	149	95	41	36	9	7	2	3	3	3	955	955
	Sen	NA	0.367	0.571	0.736	0.688	0.692	0.921	0.948	0.977	0.980	0.983	0.956	1.000	0.900	1.000	1.000	1.000	0.900	0.925
	PPV	NA	1.000	1.000	0.955	0.930	0.961	0.983	0.790	0.683	0.551	0.429	0.317	0.272	0.180	0.185	0.169	0.152	0.859	0.844
	Mis	NA	0.046	0.031	0.022	0.028	0.025	0.007	0.022	0.035	0.061	0.097	0.155	0.195	0.312	0.319	0.357	0.406	0.028	0.029

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A27. Ordinal model: Simulation results for true cluster model (C) with an alternative hypothesis (1) using elliptical windows.

	1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Maximum reported cluster size (MRCS)		Default Setting
																		Overall		
SCIC ₁	Freq	NA	3	1	6	9	34	151	228	412	73	41	14	11	9	3	2	NA	997	997
	Sen	NA	0.100	0.200	0.333	0.311	0.476	0.603	0.729	0.783	0.747	0.720	0.807	0.755	0.922	0.967	0.700	NA	0.719	0.743
	PPV	NA	1.000	1.000	1.000	0.978	0.971	0.961	0.943	0.955	0.781	0.663	0.563	0.686	0.496	0.372	0.650	NA	0.915	0.825
	Mis	NA	0.130	0.116	0.097	0.101	0.078	0.062	0.046	0.038	0.069	0.099	0.123	0.119	0.169	0.242	0.196	NA	0.055	0.079
SCIC ₂	Freq	NA	3	3	10	33	75	178	222	385	54	17	7	3	5	1	1	NA	997	997
	Sen	NA	0.100	0.300	0.400	0.479	0.581	0.670	0.764	0.813	0.767	0.700	0.786	0.867	0.940	0.900	0.900	NA	0.737	0.743
	PPV	NA	1.000	1.000	0.980	0.977	0.964	0.947	0.935	0.953	0.778	0.626	0.578	0.433	0.496	0.346	0.300	NA	0.927	0.825
	Mis	NA	0.130	0.101	0.088	0.078	0.064	0.054	0.043	0.034	0.067	0.105	0.122	0.184	0.171	0.261	0.319	NA	0.050	0.079
Gini	Freq	NA	1	1	5	31	95	135	137	333	63	45	48	31	28	29	8	7	997	997
	Sen	NA	0.100	0.200	0.360	0.594	0.685	0.710	0.797	0.807	0.783	0.751	0.823	0.865	0.943	0.962	0.975	1.000	0.780	0.743
	PPV	NA	1.000	1.000	1.000	0.971	0.963	0.921	0.917	0.945	0.769	0.613	0.545	0.473	0.452	0.378	0.339	0.305	0.841	0.825
	Mis	NA	0.130	0.116	0.093	0.062	0.050	0.052	0.041	0.036	0.066	0.106	0.128	0.164	0.186	0.236	0.281	0.331	0.069	0.079
Elbow	Freq	NA	1	6	31	129	176	223	185	213	28	3	1	1	NA	NA	NA	NA	997	997
	Sen	NA	0.100	0.333	0.497	0.651	0.673	0.692	0.763	0.800	0.757	0.700	0.800	0.900	NA	NA	NA	NA	0.713	0.743
	PPV	NA	1.000	1.000	0.984	0.965	0.953	0.934	0.927	0.941	0.728	0.679	0.500	0.474	NA	NA	NA	NA	0.936	0.825
	Mis	NA	0.130	0.097	0.074	0.054	0.053	0.053	0.044	0.038	0.079	0.092	0.145	0.159	NA	NA	NA	NA	0.050	0.079
MSC -P	Freq	NA	1	3	15	75	143	173	166	311	47	23	14	12	8	3	2	1	997	997
	Sen	NA	0.100	0.300	0.440	0.599	0.683	0.708	0.793	0.824	0.768	0.717	0.836	0.858	0.963	0.933	0.950	1.000	0.751	0.743
	PPV	NA	1.000	1.000	0.977	0.966	0.953	0.935	0.930	0.941	0.757	0.617	0.559	0.490	0.541	0.346	0.329	0.294	0.909	0.825
	Mis	NA	0.130	0.101	0.083	0.061	0.051	0.050	0.040	0.034	0.071	0.106	0.121	0.158	0.150	0.266	0.290	0.348	0.052	0.079
MSHC -P	Freq	NA	1	7	33	92	116	179	168	283	53	24	15	13	8	2	2	1	997	997
	Sen	NA	0.100	0.329	0.455	0.596	0.662	0.703	0.789	0.809	0.783	0.758	0.847	0.862	0.975	0.950	0.950	1.000	0.735	0.743
	PPV	NA	1.000	1.000	0.990	0.974	0.946	0.933	0.937	0.939	0.765	0.611	0.554	0.482	0.489	0.346	0.329	0.294	0.908	0.825
	Mis	NA	0.130	0.097	0.080	0.061	0.055	0.051	0.039	0.037	0.068	0.106	0.122	0.163	0.172	0.268	0.290	0.348	0.055	0.079

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A28. Ordinal model: Simulation results for true cluster model (C) with an alternative hypothesis (2) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall	
SCIC ₁	Freq	NA	4	1	11	25	48	142	208	348	86	49	22	14	12	8	5	1	984	984
	Sen	NA	0.100	0.200	0.273	0.308	0.433	0.558	0.650	0.730	0.698	0.735	0.800	0.857	0.892	0.975	0.900	1.000	0.661	0.714
	PPV	NA	1.000	1.000	1.000	0.920	0.924	0.953	0.931	0.932	0.734	0.626	0.538	0.495	0.455	0.371	0.385	0.294	0.874	0.777
	Mis	NA	0.130	0.116	0.105	0.104	0.088	0.068	0.058	0.048	0.081	0.103	0.132	0.156	0.194	0.245	0.258	0.348	0.071	0.097
SCIC ₂	Freq	NA	3	2	17	43	71	172	199	319	76	39	14	11	8	5	4	1	984	984
	Sen	NA	0.100	0.250	0.329	0.435	0.530	0.610	0.671	0.755	0.707	0.736	0.821	0.873	0.913	0.960	0.950	1.000	0.673	0.714
	PPV	NA	1.000	1.000	0.985	0.944	0.926	0.936	0.924	0.933	0.735	0.619	0.515	0.463	0.401	0.359	0.322	0.294	0.883	0.777
	Mis	NA	0.130	0.109	0.098	0.085	0.075	0.063	0.057	0.045	0.080	0.104	0.140	0.166	0.212	0.255	0.297	0.348	0.067	0.097
Gini	Freq	NA	1	NA	8	22	62	112	136	290	77	78	57	46	29	36	21	9	984	984
	Sen	NA	0.100	NA	0.338	0.418	0.589	0.610	0.705	0.756	0.703	0.749	0.814	0.865	0.931	0.978	0.957	0.978	0.733	0.714
	PPV	NA	1.000	NA	1.000	0.966	0.944	0.925	0.910	0.942	0.720	0.606	0.520	0.466	0.428	0.381	0.329	0.307	0.791	0.777
	Mis	NA	0.130	NA	0.096	0.086	0.065	0.065	0.054	0.043	0.084	0.108	0.137	0.165	0.197	0.234	0.291	0.325	0.089	0.097
Elbow	Freq	NA	3	3	21	89	118	194	206	267	58	16	4	NA	2	1	2	NA	984	984
	Sen	NA	0.100	0.267	0.390	0.529	0.581	0.619	0.670	0.750	0.679	0.713	0.825	NA	0.950	1.000	1.000	NA	0.653	0.714
	PPV	NA	1.000	0.917	1.000	0.950	0.931	0.930	0.918	0.918	0.705	0.617	0.527	NA	0.432	0.385	0.345	NA	0.905	0.777
	Mis	NA	0.130	0.111	0.088	0.072	0.067	0.063	0.058	0.047	0.088	0.107	0.134	NA	0.188	0.232	0.275	NA	0.063	0.097
MSC-P	Freq	NA	1	1	15	52	87	159	177	275	75	59	25	19	9	13	11	6	984	984
	Sen	NA	0.100	0.300	0.367	0.510	0.598	0.627	0.699	0.765	0.725	0.742	0.808	0.858	0.944	0.977	0.945	0.950	0.702	0.714
	PPV	NA	1.000	1.000	0.983	0.962	0.932	0.928	0.919	0.932	0.744	0.612	0.519	0.456	0.414	0.371	0.329	0.286	0.856	0.777
	Mis	NA	0.130	0.101	0.093	0.074	0.065	0.062	0.054	0.043	0.077	0.107	0.137	0.169	0.203	0.244	0.289	0.353	0.073	0.097
MSHC-P	Freq	NA	1	5	23	68	97	153	172	260	66	54	25	20	10	13	11	6	984	984
	Sen	NA	0.100	0.320	0.413	0.521	0.608	0.626	0.701	0.766	0.708	0.735	0.812	0.865	0.950	0.977	0.945	0.950	0.695	0.714
	PPV	NA	1.000	0.950	0.991	0.964	0.944	0.931	0.920	0.929	0.734	0.607	0.528	0.472	0.413	0.371	0.329	0.286	0.860	0.777
	Mis	NA	0.130	0.101	0.086	0.072	0.062	0.062	0.054	0.044	0.081	0.109	0.134	0.163	0.204	0.244	0.289	0.353	0.073	0.097

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A29. Ordinal model: Simulation results for true cluster model (C) with an alternative hypothesis (3) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall	
SCIC ₁	Freq	NA	5	3	13	37	44	146	207	337	82	57	23	14	7	8	2	2	987	987
	Sen	NA	0.120	0.167	0.300	0.332	0.436	0.541	0.647	0.728	0.685	0.742	0.809	0.807	0.929	1.000	0.950	0.850	0.650	0.718
	PPV	NA	1.000	1.000	0.974	0.962	0.929	0.956	0.950	0.938	0.724	0.643	0.550	0.612	0.449	0.394	0.324	0.661	0.886	0.774
	Mis	NA	0.128	0.121	0.103	0.099	0.087	0.070	0.056	0.047	0.086	0.100	0.128	0.137	0.184	0.225	0.297	0.174	0.070	0.098
SCIC ₂	Freq	NA	3	4	19	40	75	165	222	315	61	43	17	9	6	5	2	1	987	987
	Sen	NA	0.100	0.250	0.358	0.383	0.516	0.596	0.676	0.752	0.707	0.747	0.806	0.844	0.917	1.000	0.950	1.000	0.666	0.718
	PPV	NA	1.000	1.000	0.932	0.956	0.939	0.938	0.939	0.931	0.723	0.627	0.520	0.513	0.460	0.392	0.324	0.323	0.892	0.774
	Mis	NA	0.130	0.109	0.097	0.092	0.076	0.065	0.054	0.046	0.084	0.103	0.136	0.159	0.176	0.226	0.297	0.304	0.066	0.098
Gini	Freq	NA	NA	NA	10	28	59	109	127	304	68	74	53	51	33	40	20	11	987	987
	Sen	NA	NA	NA	0.400	0.439	0.575	0.638	0.713	0.754	0.721	0.741	0.811	0.859	0.939	0.975	0.955	0.964	0.739	0.718
	PPV	NA	NA	NA	0.980	0.978	0.946	0.931	0.917	0.927	0.747	0.610	0.533	0.452	0.422	0.372	0.326	0.298	0.789	0.774
	Mis	NA	NA	NA	0.088	0.083	0.067	0.061	0.052	0.046	0.078	0.107	0.132	0.172	0.198	0.243	0.294	0.335	0.090	0.098
Elbow	Freq	NA	1	4	39	92	130	181	204	258	51	14	6	5	NA	2	NA	NA	987	987
	Sen	NA	0.100	0.250	0.418	0.505	0.578	0.599	0.680	0.745	0.694	0.771	0.867	0.880	NA	1.000	NA	NA	0.645	0.718
	PPV	NA	1.000	1.000	0.953	0.953	0.936	0.921	0.931	0.921	0.717	0.606	0.530	0.459	NA	0.408	NA	NA	0.909	0.774
	Mis	NA	0.130	0.109	0.087	0.076	0.067	0.067	0.055	0.048	0.086	0.108	0.130	0.168	NA	0.210	NA	NA	0.064	0.098
MSC-P	Freq	NA	NA	2	21	52	105	126	168	296	69	53	33	19	11	13	14	5	987	987
	Sen	NA	NA	0.350	0.405	0.504	0.594	0.636	0.708	0.760	0.716	0.747	0.782	0.868	0.927	0.969	0.950	0.980	0.703	0.718
	PPV	NA	NA	1.000	0.953	0.971	0.943	0.933	0.925	0.925	0.734	0.627	0.520	0.451	0.421	0.374	0.324	0.309	0.855	0.774
	Mis	NA	NA	0.094	0.089	0.074	0.065	0.061	0.052	0.045	0.080	0.103	0.137	0.172	0.196	0.241	0.295	0.322	0.073	0.098
MSHC-P	Freq	NA	NA	NA	23	61	86	121	168	294	69	57	37	25	12	15	14	5	987	987
	Sen	NA	NA	NA	0.370	0.495	0.553	0.616	0.693	0.737	0.703	0.751	0.795	0.868	0.933	0.967	0.950	0.980	0.689	0.718
	PPV	NA	NA	NA	0.968	0.953	0.925	0.941	0.927	0.926	0.734	0.633	0.521	0.474	0.420	0.417	0.324	0.309	0.848	0.774
	Mis	NA	NA	NA	0.093	0.077	0.072	0.063	0.054	0.048	0.081	0.102	0.136	0.167	0.197	0.224	0.295	0.322	0.077	0.098

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A30. Ordinal model: Simulation results for true cluster model (C) with an alternative hypothesis (4) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall	
SCIC ₁	Freq	NA	4	9	15	29	32	166	203	307	84	59	26	17	12	12	8	NA	983	983
	Sen	NA	0.100	0.200	0.247	0.310	0.413	0.523	0.648	0.716	0.679	0.734	0.800	0.865	0.875	0.950	0.950	NA	0.643	0.719
	PPV	NA	1.000	0.917	0.956	0.954	0.958	0.965	0.937	0.950	0.737	0.639	0.590	0.499	0.506	0.366	0.423	NA	0.879	0.746
	Mis	NA	0.130	0.119	0.111	0.103	0.088	0.072	0.057	0.047	0.083	0.101	0.119	0.157	0.181	0.246	0.250	NA	0.074	0.108
SCIC ₂	Freq	NA	3	9	19	38	52	187	211	300	67	44	20	11	7	10	5	NA	983	983
	Sen	NA	0.100	0.233	0.300	0.408	0.519	0.571	0.667	0.739	0.704	0.732	0.805	0.873	0.914	0.960	0.980	NA	0.657	0.719
	PPV	NA	1.000	0.917	0.947	0.957	0.938	0.956	0.934	0.951	0.741	0.626	0.537	0.465	0.410	0.368	0.343	NA	0.892	0.746
	Mis	NA	0.130	0.114	0.104	0.089	0.075	0.066	0.055	0.044	0.079	0.103	0.130	0.166	0.205	0.245	0.275	NA	0.069	0.108
Gini	Freq	NA	1	4	10	22	61	113	135	241	61	81	74	51	44	48	24	13	983	983
	Sen	NA	0.100	0.225	0.360	0.441	0.600	0.615	0.690	0.742	0.703	0.733	0.809	0.869	0.941	0.958	0.967	0.977	0.735	0.719
	PPV	NA	1.000	1.000	1.000	0.975	0.954	0.949	0.924	0.945	0.735	0.613	0.534	0.465	0.422	0.371	0.329	0.302	0.773	0.746
	Mis	NA	0.130	0.112	0.093	0.083	0.062	0.061	0.054	0.044	0.081	0.107	0.130	0.165	0.199	0.242	0.292	0.331	0.097	0.108
Elbow	Freq	NA	3	8	40	70	116	212	233	229	48	14	4	NA	2	3	1	NA	983	983
	Sen	NA	0.167	0.250	0.403	0.489	0.599	0.584	0.666	0.729	0.696	0.714	0.750	NA	0.900	0.967	1.000	NA	0.631	0.719
	PPV	NA	1.000	0.875	0.948	0.962	0.949	0.936	0.918	0.941	0.742	0.605	0.523	NA	0.400	0.364	0.370	NA	0.917	0.746
	Mis	NA	0.121	0.114	0.090	0.077	0.063	0.067	0.058	0.047	0.080	0.111	0.138	NA	0.210	0.251	0.246	NA	0.064	0.108
MSC-P	Freq	NA	1	4	23	42	91	151	177	254	68	56	39	20	13	26	12	6	983	983
	Sen	NA	0.100	0.225	0.391	0.495	0.614	0.613	0.694	0.757	0.706	0.732	0.790	0.875	0.908	0.965	0.967	0.983	0.698	0.719
	PPV	NA	1.000	1.000	0.972	0.970	0.957	0.942	0.930	0.946	0.749	0.611	0.521	0.473	0.402	0.372	0.333	0.289	0.852	0.746
	Mis	NA	0.130	0.112	0.090	0.076	0.060	0.062	0.053	0.042	0.078	0.107	0.137	0.162	0.210	0.242	0.286	0.353	0.077	0.108
MSHC-P	Freq	NA	3	8	35	53	96	153	169	238	67	51	36	19	14	25	12	4	983	983
	Sen	NA	0.200	0.275	0.406	0.506	0.604	0.614	0.695	0.772	0.706	0.729	0.789	0.884	0.914	0.964	0.967	0.975	0.691	0.719
	PPV	NA	1.000	0.969	0.971	0.965	0.950	0.934	0.925	0.946	0.745	0.609	0.521	0.473	0.433	0.371	0.333	0.287	0.855	0.746
	Mis	NA	0.116	0.107	0.088	0.075	0.062	0.063	0.053	0.040	0.079	0.109	0.137	0.162	0.197	0.242	0.286	0.355	0.076	0.108

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A31. Ordinal model: Simulation results for true cluster model (D) with an alternative hypothesis (1) using elliptical windows.

	1%	2%	3%	4%	5%	6%	8%	Maximum reported cluster size (MRCS)													Default Setting	
								10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall				
SCIC ₁	Freq	NA	NA	1	5	18	29	417	149	75	25	52	140	55	23	3	6	1	999	999		
	Sen	NA	NA	0.300	0.240	0.394	0.497	0.847	0.903	0.901	0.836	0.817	0.934	0.933	0.957	0.833	0.967	0.500	0.856	0.873		
	PPV	NA	NA	1.000	1.000	0.970	0.968	0.984	0.881	0.784	0.724	0.687	0.687	0.595	0.511	0.590	0.431	0.714	0.852	0.770		
	Mis	NA	NA	0.101	0.110	0.089	0.075	0.024	0.031	0.050	0.072	0.084	0.076	0.113	0.156	0.179	0.227	0.101	0.052	0.076		
SCIC ₂	Freq	NA	NA	1	5	13	29	520	220	108	32	22	32	12	3	NA	2	NA	999	999		
	Sen	NA	NA	0.300	0.340	0.523	0.559	0.867	0.923	0.929	0.966	0.923	0.963	0.992	1.000	NA	0.900	NA	0.879	0.873		
	PPV	NA	NA	1.000	0.900	0.972	0.968	0.983	0.883	0.784	0.711	0.742	0.670	0.621	0.486	NA	0.572	NA	0.908	0.770		
	Mis	NA	NA	0.101	0.101	0.071	0.066	0.022	0.029	0.047	0.063	0.063	0.078	0.097	0.155	NA	0.167	NA	0.034	0.076		
Gini	Freq	NA	NA	NA	NA	NA	1	6	369	171	121	56	41	62	100	46	14	6	6	999	999	
	Sen	NA	NA	NA	NA	NA	0.600	0.383	0.859	0.929	0.930	0.966	0.946	0.961	0.990	0.993	0.993	0.967	0.967	0.915	0.873	
	PPV	NA	NA	NA	NA	NA	1.000	0.913	0.991	0.881	0.783	0.708	0.708	0.668	0.570	0.507	0.392	0.429	0.302	0.819	0.770	
	Mis	NA	NA	NA	NA	NA	0.058	0.099	0.022	0.028	0.047	0.064	0.070	0.081	0.115	0.152	0.226	0.229	0.329	0.056	0.076	
Elbow	Freq	NA	NA	2	15	24	26	661	193	61	10	3	4	NA	NA	NA	NA	NA	999	999		
	Sen	NA	NA	0.450	0.560	0.592	0.546	0.881	0.917	0.921	0.970	0.800	0.950	NA	NA	NA	NA	NA	0.870	0.873		
	PPV	NA	NA	1.000	0.959	0.909	0.959	0.981	0.870	0.768	0.693	0.585	0.585	NA	NA	NA	NA	NA	0.938	0.770		
	Mis	NA	NA	0.080	0.067	0.068	0.070	0.020	0.032	0.052	0.068	0.111	0.112	NA	NA	NA	NA	NA	0.028	0.076		
MSC-P	Freq	NA	NA	NA	NA	NA	2	4	12	512	230	134	44	31	14	9	4	1	2	NA	999	999
	Sen	NA	NA	NA	NA	NA	0.450	0.525	0.483	0.871	0.931	0.934	0.973	0.952	0.936	0.944	1.000	1.000	0.900	NA	0.896	0.873
	PPV	NA	NA	NA	NA	NA	0.875	0.873	0.956	0.986	0.884	0.785	0.710	0.723	0.747	0.621	0.497	0.417	0.572	NA	0.904	0.770
	Mis	NA	NA	NA	NA	NA	0.087	0.080	0.080	0.021	0.028	0.047	0.063	0.068	0.068	0.103	0.149	0.203	0.167	NA	0.033	0.076
MSHC-P	Freq	NA	NA	1	9	18	31	503	190	116	48	38	18	14	9	1	2	1	999	999		
	Sen	NA	NA	0.300	0.456	0.583	0.584	0.880	0.925	0.932	0.965	0.932	0.950	0.900	0.989	1.000	0.900	0.900	0.900	0.884	0.873	
	PPV	NA	NA	1.000	0.950	0.985	0.970	0.986	0.873	0.787	0.709	0.717	0.686	0.763	0.601	0.417	0.572	0.281	0.903	0.770		
	Mis	NA	NA	0.101	0.082	0.062	0.064	0.019	0.030	0.046	0.064	0.071	0.084	0.068	0.116	0.203	0.167	0.348	0.035	0.076		

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A32. Ordinal model: Simulation results for true cluster model (D) with an alternative hypothesis (2) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall	
SCIC ₁	Freq	NA	1	3	12	39	48	380	123	64	39	41	131	73	25	8	5	NA	992	992
	Sen	NA	0.100	0.200	0.242	0.351	0.421	0.765	0.835	0.867	0.721	0.817	0.934	0.973	0.900	0.738	0.900	NA	0.781	0.834
	PPV	NA	1.000	1.000	1.000	0.956	0.962	0.980	0.860	0.771	0.680	0.665	0.662	0.576	0.531	0.630	0.344	NA	0.836	0.750
	Mis	NA	0.130	0.116	0.110	0.096	0.086	0.036	0.043	0.057	0.090	0.086	0.082	0.116	0.154	0.156	0.258	NA	0.066	0.088
SCIC ₂	Freq	NA	NA	2	8	33	47	449	191	109	49	19	45	31	8	NA	1	NA	992	992
	Sen	NA	NA	0.200	0.288	0.415	0.489	0.800	0.871	0.904	0.837	0.853	0.958	0.981	0.975	NA	1.000	NA	0.810	0.834
	PPV	NA	NA	1.000	1.000	0.973	0.968	0.975	0.861	0.767	0.677	0.670	0.666	0.584	0.496	NA	0.370	NA	0.879	0.750
	Mis	NA	NA	0.116	0.103	0.087	0.076	0.032	0.039	0.054	0.081	0.085	0.078	0.111	0.154	NA	0.246	NA	0.050	0.088
Gini	Freq	NA	NA	NA	1	4	15	322	141	116	76	42	77	106	56	21	9	6	992	992
	Sen	NA	NA	NA	0.200	0.350	0.407	0.771	0.864	0.915	0.886	0.893	0.943	0.988	0.982	0.971	0.978	0.983	0.863	0.834
	PPV	NA	NA	NA	1.000	1.000	0.990	0.983	0.863	0.774	0.667	0.660	0.643	0.553	0.473	0.384	0.348	0.308	0.780	0.750
	Mis	NA	NA	NA	0.116	0.094	0.087	0.035	0.039	0.051	0.080	0.088	0.088	0.121	0.165	0.231	0.271	0.324	0.073	0.088
Elbow	Freq	NA	NA	2	6	31	46	562	196	90	29	17	9	3	1	NA	NA	NA	992	992
	Sen	NA	NA	0.350	0.250	0.439	0.474	0.809	0.863	0.894	0.797	0.835	0.978	1.000	1.000	NA	NA	NA	0.798	0.8336 69
	PPV	NA	NA	1.000	1.000	0.937	0.965	0.973	0.851	0.754	0.650	0.651	0.641	0.559	0.435	NA	NA	NA	0.908	0.7504 61
	Mis	NA	NA	0.094	0.109	0.086	0.079	0.031	0.041	0.058	0.091	0.090	0.084	0.116	0.188	NA	NA	NA	0.044	0.0877 31
MSC-P	Freq	NA	NA	NA	2	10	22	462	193	134	68	22	33	31	9	3	2	1	992	992
	Sen	NA	NA	NA	0.350	0.430	0.459	0.805	0.890	0.915	0.865	0.882	0.936	0.965	0.956	1.000	0.950	1.000	0.841	0.834
	PPV	NA	NA	NA	1.000	1.000	0.987	0.978	0.868	0.768	0.666	0.632	0.646	0.540	0.473	0.376	0.334	0.323	0.866	0.750
	Mis	NA	NA	NA	0.094	0.083	0.080	0.031	0.035	0.053	0.082	0.093	0.087	0.131	0.167	0.242	0.283	0.304	0.049	0.088
MSHC-P	Freq	NA	NA	1	3	18	35	450	184	113	74	27	38	34	9	3	2	1	992	992
	Sen	NA	NA	0.400	0.367	0.483	0.517	0.810	0.880	0.909	0.874	0.893	0.934	0.968	0.944	1.000	0.950	1.000	0.835	0.834
	PPV	NA	NA	1.000	0.933	0.981	0.973	0.979	0.861	0.775	0.668	0.676	0.623	0.560	0.453	0.376	0.334	0.323	0.866	0.750
	Mis	NA	NA	0.087	0.097	0.076	0.072	0.030	0.038	0.051	0.081	0.083	0.096	0.124	0.176	0.242	0.283	0.304	0.050	0.088

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A33. Ordinal model: Simulation results for true cluster model (D) with an alternative hypothesis (3) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall	
SCIC ₁	Freq	NA	1	10	7	27	44	354	90	63	34	71	172	85	20	8	NA	2	988	988
	Sen	NA	0.100	0.180	0.271	0.333	0.459	0.729	0.778	0.776	0.709	0.782	0.940	0.944	0.910	0.913	NA	1.000	0.768	0.814
	PPV	NA	1.000	1.000	0.952	1.000	0.957	0.977	0.859	0.756	0.707	0.669	0.660	0.573	0.550	0.473	NA	0.323	0.817	0.737
	Mis	NA	0.130	0.119	0.108	0.097	0.082	0.042	0.050	0.067	0.087	0.090	0.082	0.118	0.147	0.199	NA	0.304	0.071	0.093
SCIC ₂	Freq	NA	1	8	6	23	50	465	150	107	42	32	57	34	8	5	NA	NA	988	988
	Sen	NA	0.100	0.200	0.417	0.426	0.512	0.779	0.863	0.861	0.826	0.800	0.947	0.968	0.938	0.960	NA	NA	0.792	0.814
	PPV	NA	1.000	1.000	0.943	0.986	0.953	0.979	0.861	0.771	0.683	0.676	0.647	0.551	0.535	0.384	NA	NA	0.874	0.737
	Mis	NA	0.130	0.116	0.089	0.084	0.074	0.035	0.040	0.056	0.080	0.089	0.085	0.126	0.147	0.229	NA	NA	0.054	0.093
Gini	Freq	NA	1	3	NA	5	17	325	140	128	56	43	80	108	44	27	7	4	988	988
	Sen	NA	0.100	0.267	NA	0.300	0.453	0.751	0.845	0.866	0.895	0.853	0.933	0.984	0.982	0.985	0.986	0.975	0.842	0.814
	PPV	NA	1.000	1.000	NA	1.000	0.992	0.983	0.861	0.773	0.687	0.655	0.633	0.556	0.473	0.415	0.386	0.385	0.788	0.737
	Mis	NA	0.130	0.106	NA	0.101	0.080	0.038	0.042	0.055	0.074	0.089	0.090	0.122	0.167	0.210	0.251	0.264	0.073	0.093
Elbow	Freq	NA	1	11	7	21	51	594	158	92	25	16	8	1	2	1	NA	NA	988	988
	Sen	NA	0.100	0.236	0.443	0.395	0.522	0.794	0.845	0.834	0.816	0.750	0.925	1.000	1.000	1.000	NA	NA	0.775	0.814
	PPV	NA	1.000	1.000	0.951	0.960	0.954	0.976	0.847	0.753	0.653	0.609	0.623	0.556	0.500	0.400	NA	NA	0.914	0.737
	Mis	NA	0.130	0.111	0.085	0.090	0.073	0.033	0.044	0.062	0.089	0.109	0.094	0.116	0.145	0.217	NA	NA	0.046	0.093
MSC-P	Freq	NA	1	7	3	9	24	450	185	145	58	29	21	28	18	9	1	NA	988	988
	Sen	NA	0.100	0.271	0.367	0.367	0.479	0.793	0.863	0.882	0.895	0.810	0.919	0.979	0.967	0.956	1.000	NA	0.821	0.814
	PPV	NA	1.000	1.000	1.000	1.000	0.966	0.982	0.864	0.774	0.692	0.643	0.653	0.557	0.499	0.379	0.333	NA	0.868	0.737
	Mis	NA	0.130	0.106	0.092	0.092	0.078	0.032	0.039	0.054	0.073	0.096	0.087	0.124	0.156	0.233	0.290	NA	0.051	0.093
MSHC-P	Freq	NA	1	8	8	33	52	399	155	112	55	36	49	43	24	11	2	NA	988	988
	Sen	NA	0.100	0.250	0.300	0.500	0.546	0.780	0.856	0.861	0.873	0.822	0.902	0.972	0.933	0.964	1.000	NA	0.798	0.814
	PPV	NA	1.000	1.000	1.000	0.984	0.973	0.979	0.859	0.774	0.682	0.636	0.669	0.560	0.525	0.382	0.500	NA	0.856	0.737
	Mis	NA	0.130	0.109	0.101	0.074	0.068	0.034	0.041	0.056	0.078	0.097	0.085	0.122	0.150	0.232	0.181	NA	0.058	0.093

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A34. Ordinal model: Simulation results for true cluster model (D) with an alternative hypothesis (4) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall	
SCIC ₁	Freq	NA	2	7	20	44	34	317	45	35	31	103	201	111	24	11	3	NA	988	988
	Sen	NA	0.100	0.200	0.240	0.350	0.412	0.701	0.764	0.737	0.703	0.779	0.941	0.962	0.950	0.900	0.833	NA	0.760	0.807
	PPV	NA	1.000	1.000	1.000	0.985	0.980	0.981	0.832	0.747	0.715	0.685	0.667	0.571	0.505	0.447	0.568	NA	0.799	0.706
	Mis	NA	0.130	0.116	0.110	0.095	0.087	0.045	0.055	0.074	0.084	0.087	0.080	0.117	0.156	0.206	0.203	NA	0.078	0.105
SCIC ₂	Freq	NA	2	3	21	39	48	478	100	73	35	39	91	41	12	5	1	NA	988	988
	Sen	NA	0.100	0.200	0.267	0.408	0.473	0.772	0.840	0.848	0.803	0.821	0.959	0.978	0.975	0.980	1.000	NA	0.774	0.807
	PPV	NA	1.000	1.000	1.000	0.980	0.971	0.980	0.853	0.771	0.714	0.683	0.671	0.566	0.523	0.390	0.370	NA	0.876	0.706
	Mis	NA	0.130	0.116	0.106	0.087	0.079	0.036	0.043	0.059	0.075	0.083	0.077	0.117	0.149	0.226	0.246	NA	0.057	0.105
Gini	Freq	NA	1	1	NA	12	11	349	118	107	57	40	73	123	52	30	10	4	988	988
	Sen	NA	0.100	0.200	NA	0.317	0.436	0.726	0.855	0.888	0.904	0.850	0.922	0.979	0.990	0.983	0.990	0.900	0.836	0.807
	PPV	NA	1.000	1.000	NA	1.000	0.955	0.988	0.860	0.777	0.688	0.679	0.623	0.558	0.489	0.404	0.353	0.285	0.786	0.706
	Mis	NA	0.130	0.116	NA	0.099	0.084	0.041	0.041	0.053	0.074	0.084	0.094	0.121	0.159	0.217	0.265	0.341	0.076	0.105
Elbow	Freq	NA	2	10	12	43	44	631	123	67	14	17	15	7	2	1	NA	NA	988	988
	Sen	NA	0.100	0.260	0.317	0.409	0.500	0.788	0.830	0.815	0.750	0.782	0.960	0.971	0.950	1.000	NA	NA	0.757	0.807
	PPV	NA	1.000	0.980	1.000	0.969	0.971	0.975	0.841	0.749	0.693	0.634	0.660	0.523	0.434	0.385	NA	NA	0.923	0.706
	Mis	NA	0.130	0.109	0.099	0.088	0.075	0.034	0.047	0.066	0.085	0.097	0.078	0.137	0.188	0.232	NA	NA	0.047	0.105
MSC-P	Freq	NA	1	3	2	20	23	481	154	118	59	27	41	41	8	6	2	2	988	988
	Sen	NA	0.100	0.233	0.200	0.390	0.491	0.770	0.872	0.893	0.880	0.830	0.924	0.980	0.963	0.967	1.000	1.000	0.810	0.807
	PPV	NA	1.000	1.000	1.000	0.994	0.971	0.982	0.858	0.776	0.688	0.673	0.643	0.561	0.450	0.385	0.364	0.308	0.870	0.706
	Mis	NA	0.130	0.111	0.116	0.089	0.076	0.036	0.039	0.053	0.076	0.086	0.088	0.122	0.178	0.229	0.254	0.326	0.054	0.105
MSHC-P	Freq	NA	1	7	4	18	14	452	148	131	70	35	40	47	10	7	2	2	988	988
	Sen	NA	0.100	0.271	0.300	0.372	0.457	0.767	0.861	0.903	0.890	0.863	0.928	0.974	0.960	0.971	1.000	1.000	0.814	0.807
	PPV	NA	1.000	1.000	1.000	0.993	0.964	0.985	0.856	0.782	0.695	0.679	0.637	0.585	0.525	0.425	0.364	0.308	0.863	0.706
	Mis	NA	0.130	0.106	0.101	0.092	0.081	0.036	0.041	0.051	0.073	0.084	0.089	0.115	0.151	0.207	0.254	0.326	0.055	0.105

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A35. Ordinal model: Simulation results for true cluster model (D) with an alternative hypothesis (5) using elliptical windows.

	1%	2%	3%	4%	5%	6%	8%	Maximum reported cluster size (MRCS)												Default Setting
								10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall		
SCIC ₁	Freq	1	2	3	4	30	40	419	163	88	29	44	92	53	20	3	6	NA	997	997
	Sen	0.100	0.100	0.167	0.175	0.363	0.473	0.798	0.909	0.889	0.793	0.795	0.928	0.915	0.895	0.833	0.833	NA	0.812	0.847
	PPV	1.000	1.000	0.833	0.875	0.956	0.982	0.983	0.883	0.779	0.689	0.648	0.676	0.584	0.568	0.600	0.526	NA	0.862	0.794
	Mis	0.130	0.130	0.126	0.123	0.095	0.078	0.031	0.030	0.052	0.081	0.096	0.080	0.121	0.147	0.169	0.200	NA	0.055	0.073
SCIC ₂	Freq	NA	2	2	5	21	30	508	221	113	40	18	22	11	4	NA	NA	NA	997	997
	Sen	NA	0.100	0.150	0.220	0.452	0.527	0.828	0.917	0.904	0.888	0.878	0.964	0.955	0.975	NA	NA	NA	0.842	0.847
	PPV	NA	1.000	0.750	0.900	0.936	0.975	0.980	0.877	0.785	0.683	0.646	0.684	0.549	0.508	NA	NA	NA	0.902	0.794
	Mis	NA	0.130	0.130	0.116	0.084	0.071	0.028	0.030	0.050	0.076	0.092	0.072	0.121	0.149	NA	NA	NA	0.040	0.073
Gini	Freq	NA	NA	2	2	4	5	375	169	144	58	25	54	90	43	15	8	3	997	997
	Sen	NA	NA	0.150	0.300	0.325	0.400	0.807	0.910	0.908	0.941	0.864	0.948	0.991	0.986	0.980	0.963	0.967	0.878	0.847
	PPV	NA	NA	0.750	1.000	0.875	0.971	0.984	0.877	0.786	0.699	0.620	0.652	0.556	0.494	0.423	0.489	0.306	0.818	0.794
	Mis	NA	NA	0.130	0.101	0.105	0.090	0.030	0.031	0.049	0.068	0.099	0.085	0.120	0.157	0.210	0.197	0.324	0.059	0.073
Elbow	Freq	NA	1	5	9	32	32	623	191	73	25	5	1	NA	NA	NA	NA	NA	997	997
	Sen	NA	0.100	0.280	0.533	0.463	0.556	0.846	0.918	0.895	0.848	0.820	0.900	NA	NA	NA	NA	NA	0.835	0.847
	PPV	NA	1.000	0.867	0.986	0.899	0.969	0.978	0.869	0.766	0.657	0.589	0.600	NA	NA	NA	NA	NA	0.928	0.794
	Mis	NA	0.130	0.110	0.069	0.086	0.067	0.025	0.032	0.055	0.085	0.110	0.101	NA	NA	NA	NA	NA	0.035	0.073
MSC-P	Freq	NA	NA	3	5	9	17	482	241	133	56	24	11	12	3	NA	NA	1	997	997
	Sen	NA	NA	0.267	0.460	0.456	0.541	0.829	0.919	0.910	0.921	0.875	0.909	0.950	1.000	NA	NA	1.000	0.859	0.847
	PPV	NA	NA	0.778	1.000	0.913	0.983	0.981	0.877	0.786	0.682	0.617	0.640	0.578	0.464	NA	NA	0.303	0.892	0.794
	Mis	NA	NA	0.116	0.078	0.085	0.068	0.027	0.030	0.049	0.074	0.100	0.092	0.118	0.169	NA	NA	0.333	0.040	0.073
MSHC-P	Freq	NA	NA	6	9	21	37	486	189	119	56	23	20	15	10	1	4	1	997	997
	Sen	NA	NA	0.233	0.411	0.495	0.586	0.842	0.906	0.908	0.932	0.861	0.960	0.940	0.960	0.500	0.800	1.000	0.848	0.847
	PPV	NA	NA	0.917	0.986	0.951	0.957	0.981	0.872	0.788	0.690	0.641	0.638	0.631	0.609	1.000	0.786	0.303	0.894	0.794
	Mis	NA	NA	0.114	0.087	0.077	0.064	0.025	0.033	0.048	0.071	0.095	0.095	0.104	0.114	0.072	0.080	0.333	0.041	0.073

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A36. Ordinal model: Simulation results for true cluster model (D) with an alternative hypothesis (6) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall	
SCIC ₁	Freq	NA	3	4	12	32	35	419	128	85	24	49	110	65	25	3	2	3	999	999
	Sen	NA	0.100	0.200	0.258	0.359	0.437	0.785	0.870	0.871	0.713	0.765	0.943	0.945	0.876	0.767	0.750	0.833	0.794	0.828
	PPV	NA	1.000	1.000	0.979	0.983	0.978	0.981	0.879	0.774	0.701	0.672	0.663	0.578	0.588	0.587	0.672	0.542	0.854	0.778
	Mis	NA	0.130	0.116	0.109	0.094	0.084	0.033	0.036	0.055	0.088	0.092	0.081	0.118	0.137	0.184	0.174	0.237	0.060	0.080
SCIC ₂	Freq	NA	1	1	11	23	32	513	199	112	36	23	27	13	6	1	NA	1	999	999
	Sen	NA	0.100	0.200	0.327	0.400	0.494	0.820	0.894	0.900	0.858	0.857	0.944	0.946	1.000	0.900	NA	1.000	0.826	0.828
	PPV	NA	1.000	1.000	0.977	0.984	0.977	0.980	0.876	0.780	0.688	0.670	0.677	0.607	0.549	0.360	NA	0.313	0.902	0.778
	Mis	NA	0.130	0.116	0.099	0.088	0.076	0.029	0.033	0.051	0.077	0.088	0.077	0.116	0.135	0.246	NA	0.319	0.043	0.080
Gini	Freq	NA	1	NA	NA	10	9	365	174	134	57	38	50	88	48	17	2	6	999	999
	Sen	NA	0.100	NA	NA	0.390	0.356	0.801	0.885	0.910	0.902	0.879	0.938	0.988	0.979	0.982	1.000	1.000	0.867	0.828
	PPV	NA	1.000	NA	NA	0.967	1.000	0.990	0.874	0.787	0.687	0.648	0.634	0.564	0.488	0.412	0.352	0.310	0.817	0.778
	Mis	NA	0.130	NA	NA	0.091	0.093	0.030	0.035	0.049	0.074	0.089	0.089	0.118	0.158	0.216	0.268	0.324	0.062	0.080
Elbow	Freq	NA	1	1	8	32	38	624	182	84	22	6	1	NA	NA	NA	NA	NA	999	999
	Sen	NA	0.100	0.300	0.350	0.459	0.539	0.835	0.888	0.865	0.841	0.783	1.000	NA	NA	NA	NA	NA	0.819	0.828
	PPV	NA	1.000	1.000	0.969	0.957	0.965	0.979	0.862	0.766	0.672	0.579	0.714	NA	NA	NA	NA	NA	0.929	0.778
	Mis	NA	0.130	0.101	0.096	0.083	0.070	0.027	0.037	0.057	0.082	0.114	0.058	NA	NA	NA	NA	NA	0.037	0.080
MSC-P	Freq	NA	1	NA	4	17	17	480	227	140	52	25	11	12	10	1	NA	2	999	999
	Sen	NA	0.100	NA	0.400	0.453	0.429	0.825	0.904	0.904	0.894	0.844	0.900	0.942	0.980	1.000	NA	1.000	0.847	0.828
	PPV	NA	1.000	NA	0.938	0.948	0.990	0.986	0.874	0.782	0.685	0.622	0.657	0.661	0.466	0.400	NA	0.313	0.892	0.778
	Mis	NA	0.130	NA	0.091	0.085	0.084	0.027	0.033	0.050	0.075	0.099	0.087	0.098	0.167	0.217	NA	0.319	0.042	0.080
MSHC-P	Freq	NA	2	2	9	33	47	446	198	117	53	35	19	22	13	1	NA	2	999	999
	Sen	NA	0.150	0.350	0.400	0.500	0.547	0.828	0.893	0.897	0.883	0.866	0.889	0.968	0.969	1.000	NA	1.000	0.830	0.828
	PPV	NA	1.000	1.000	1.000	0.980	0.979	0.986	0.867	0.784	0.690	0.688	0.628	0.623	0.517	0.400	NA	0.313	0.889	0.778
	Mis	NA	0.123	0.094	0.087	0.074	0.068	0.027	0.035	0.050	0.075	0.082	0.097	0.104	0.147	0.217	NA	0.319	0.045	0.080

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A37. Ordinal model: Simulation results for true cluster model (D) with an alternative hypothesis (7) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall	
SCIC ₁	Freq	NA	3	5	7	28	47	459	150	90	38	36	68	45	9	8	3	1	997	997
	Sen	NA	0.100	0.200	0.243	0.364	0.500	0.808	0.883	0.877	0.839	0.817	0.915	0.920	0.944	0.625	0.833	1.000	0.804	0.836
	PPV	NA	1.000	1.000	0.917	0.988	0.970	0.983	0.876	0.775	0.693	0.667	0.675	0.586	0.538	0.814	0.567	0.303	0.878	0.810
	Mis	NA	0.130	0.116	0.114	0.093	0.075	0.030	0.035	0.054	0.079	0.089	0.082	0.120	0.148	0.112	0.203	0.333	0.052	0.068
SCIC ₂	Freq	NA	2	3	8	19	47	534	177	108	51	17	16	11	2	NA	1	1	997	997
	Sen	NA	0.100	0.200	0.288	0.495	0.543	0.830	0.894	0.902	0.894	0.906	0.944	0.918	1.000	NA	1.000	1.000	0.829	0.836
	PPV	NA	1.000	1.000	0.969	0.992	0.970	0.981	0.877	0.778	0.686	0.685	0.713	0.625	0.488	NA	0.345	0.303	0.909	0.810
	Mis	NA	0.130	0.116	0.105	0.074	0.069	0.027	0.033	0.052	0.076	0.078	0.071	0.104	0.152	NA	0.275	0.333	0.041	0.068
Gini	Freq	NA	1	1	1	7	6	403	173	132	69	32	49	70	39	12	2	NA	997	997
	Sen	NA	0.200	0.200	0.200	0.429	0.433	0.816	0.894	0.889	0.926	0.938	0.941	0.969	0.997	0.975	1.000	NA	0.870	0.836
	PPV	NA	1.000	1.000	1.000	0.913	1.000	0.987	0.874	0.781	0.682	0.625	0.659	0.558	0.486	0.450	0.358	NA	0.834	0.810
	Mis	NA	0.116	0.116	0.116	0.089	0.082	0.028	0.034	0.052	0.074	0.095	0.085	0.121	0.159	0.197	0.261	NA	0.056	0.068
Elbow	Freq	NA	1	5	11	28	44	659	162	66	18	1	NA	2	NA	NA	NA	NA	997	997
	Sen	NA	0.200	0.300	0.509	0.507	0.555	0.845	0.892	0.891	0.861	0.800	NA	0.750	NA	NA	NA	NA	0.826	0.836
	PPV	NA	1.000	0.967	0.989	0.937	0.951	0.979	0.864	0.754	0.663	0.667	NA	0.417	NA	NA	NA	NA	0.936	0.810
	Mis	NA	0.116	0.104	0.072	0.077	0.069	0.025	0.036	0.059	0.085	0.087	NA	0.188	NA	NA	NA	NA	0.035	0.068
MSC-P	Freq	NA	1	2	8	20	29	503	185	127	61	25	14	15	5	NA	1	1	997	997
	Sen	NA	0.200	0.350	0.463	0.505	0.559	0.832	0.882	0.887	0.907	0.900	0.943	0.900	1.000	NA	1.000	1.000	0.839	0.836
	PPV	NA	1.000	0.917	0.984	0.961	0.976	0.981	0.872	0.776	0.687	0.654	0.701	0.571	0.512	NA	0.345	0.303	0.894	0.810
	Mis	NA	0.116	0.101	0.080	0.075	0.066	0.027	0.036	0.053	0.074	0.092	0.077	0.129	0.142	NA	0.275	0.333	0.043	0.068
MSHC-P	Freq	NA	1	6	7	20	16	489	188	139	63	29	16	14	6	1	1	1	997	997
	Sen	NA	0.200	0.283	0.443	0.490	0.513	0.845	0.884	0.894	0.913	0.955	0.944	0.879	0.950	1.000	1.000	1.000	0.850	0.836
	PPV	NA	1.000	1.000	0.982	0.940	0.982	0.987	0.869	0.778	0.675	0.646	0.700	0.616	0.627	0.556	0.345	0.303	0.891	0.810
	Mis	NA	0.116	0.104	0.083	0.078	0.072	0.024	0.036	0.052	0.077	0.087	0.080	0.118	0.106	0.116	0.275	0.333	0.042	0.068

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A38. Ordinal model: Simulation results for true cluster model (E) with an alternative hypothesis (1) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall	
SCIC ₁	Freq	NA	1	2	4	17	29	584	237	70	20	9	3	3	2	3	6	5	995	995
	Sen	NA	0.100	0.200	0.200	0.394	0.431	0.824	0.897	0.896	0.915	0.867	0.733	0.767	0.700	0.900	0.617	0.880	0.824	0.850
	PPV	NA	1.000	1.000	0.917	0.980	0.967	0.979	0.866	0.767	0.690	0.574	0.586	0.601	0.612	0.531	0.772	0.400	0.919	0.844
	Mis	NA	0.130	0.116	0.120	0.089	0.085	0.028	0.035	0.054	0.072	0.111	0.135	0.150	0.152	0.184	0.116	0.296	0.040	0.060
SCIC ₂	Freq	NA	NA	2	3	19	28	585	239	73	29	9	2	1	NA	2	2	1	995	995
	Sen	NA	NA	0.200	0.433	0.447	0.511	0.841	0.900	0.908	0.914	0.933	0.800	0.800	NA	0.900	0.700	0.700	0.843	0.850
	PPV	NA	NA	1.000	1.000	0.957	0.949	0.980	0.867	0.768	0.668	0.601	0.692	0.381	NA	0.617	0.739	0.875	0.921	0.844
	Mis	NA	NA	0.116	0.082	0.084	0.077	0.026	0.034	0.053	0.079	0.101	0.101	0.217	NA	0.152	0.080	0.058	0.036	0.060
Gini	Freq	NA	NA	NA	3	12	13	398	284	127	81	24	8	5	6	8	12	14	995	995
	Sen	NA	NA	NA	0.500	0.533	0.454	0.821	0.909	0.912	0.912	0.921	0.988	0.880	0.900	0.925	0.917	0.964	0.864	0.850
	PPV	NA	NA	NA	1.000	0.925	0.958	0.981	0.872	0.774	0.679	0.579	0.511	0.676	0.409	0.345	0.339	0.296	0.857	0.844
	Mis	NA	NA	NA	0.072	0.075	0.084	0.028	0.032	0.051	0.076	0.110	0.141	0.099	0.210	0.266	0.284	0.339	0.051	0.060
Elbow	Freq	NA	NA	4	22	45	40	647	179	45	11	2	NA	NA	NA	NA	NA	NA	995	995
	Sen	NA	NA	0.375	0.595	0.627	0.565	0.854	0.892	0.873	0.882	0.850	NA	NA	NA	NA	NA	NA	0.832	0.850
	PPV	NA	NA	1.000	0.979	0.931	0.920	0.978	0.861	0.723	0.647	0.572	NA	NA	NA	NA	NA	NA	0.937	0.844
	Mis	NA	NA	0.091	0.061	0.062	0.073	0.024	0.036	0.068	0.088	0.116	NA	NA	NA	NA	NA	NA	0.034	0.060
MSC-P	Freq	NA	NA	NA	8	19	23	499	272	101	48	14	4	1	NA	3	2	1	995	995
	Sen	NA	NA	NA	0.525	0.558	0.522	0.834	0.910	0.913	0.925	0.957	1.000	1.000	NA	0.933	0.750	1.000	0.855	0.850
	PPV	NA	NA	NA	0.979	0.933	0.932	0.979	0.869	0.772	0.678	0.593	0.521	0.667	NA	0.335	0.532	0.323	0.901	0.844
	Mis	NA	NA	NA	0.071	0.071	0.078	0.027	0.033	0.051	0.076	0.105	0.138	0.072	NA	0.280	0.196	0.304	0.039	0.060
MSHC-P	Freq	NA	NA	2	14	38	39	453	236	108	66	21	7	4	1	4	1	1	995	995
	Sen	NA	NA	0.250	0.507	0.545	0.567	0.833	0.909	0.908	0.918	0.938	0.986	0.875	1.000	0.950	0.800	1.000	0.842	0.850
	PPV	NA	NA	1.000	0.986	0.961	0.959	0.977	0.871	0.772	0.681	0.589	0.509	0.841	0.400	0.430	0.286	0.323	0.893	0.844
	Mis	NA	NA	0.109	0.072	0.070	0.068	0.027	0.033	0.052	0.075	0.108	0.143	0.043	0.217	0.225	0.319	0.304	0.043	0.060

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A39. Ordinal model: Simulation results for true cluster model (E) with an alternative hypothesis (2) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall	
SCIC ₁	Freq	NA	2	4	16	40	53	509	191	80	38	12	4	3	4	7	3	8	974	974
	Sen	NA	0.100	0.200	0.244	0.328	0.396	0.737	0.805	0.818	0.816	0.675	0.650	0.567	0.725	0.757	0.933	0.950	0.714	0.770
	PPV	NA	1.000	1.000	0.906	0.973	0.936	0.965	0.837	0.755	0.633	0.496	0.473	0.508	0.405	0.509	0.315	0.348	0.886	0.805
	Mis	NA	0.130	0.116	0.113	0.099	0.092	0.042	0.050	0.064	0.094	0.144	0.156	0.179	0.207	0.209	0.304	0.301	0.061	0.080
SCIC ₂	Freq	NA	2	2	11	24	44	508	213	93	44	19	4	1	3	1	NA	5	974	974
	Sen	NA	0.100	0.200	0.255	0.375	0.423	0.759	0.831	0.835	0.832	0.789	0.775	0.500	0.833	0.800	NA	0.940	0.754	0.770
	PPV	NA	1.000	1.000	0.939	0.950	0.931	0.965	0.839	0.754	0.631	0.532	0.465	0.278	0.373	0.320	NA	0.381	0.884	0.805
	Mis	NA	0.130	0.116	0.111	0.094	0.089	0.039	0.047	0.063	0.094	0.128	0.163	0.261	0.227	0.275	NA	0.275	0.055	0.080
Gini	Freq	NA	NA	1	3	17	26	367	239	122	80	38	14	10	5	10	10	32	974	974
	Sen	NA	NA	0.200	0.267	0.365	0.404	0.731	0.838	0.840	0.853	0.818	0.836	0.790	0.840	0.900	0.920	0.922	0.780	0.770
	PPV	NA	NA	1.000	0.889	0.941	0.936	0.965	0.851	0.757	0.650	0.541	0.457	0.388	0.353	0.347	0.437	0.303	0.817	0.805
	Mis	NA	NA	0.116	0.111	0.095	0.091	0.043	0.044	0.062	0.087	0.125	0.166	0.216	0.246	0.261	0.220	0.331	0.073	0.080
Elbow	Freq	NA	1	4	9	34	48	566	206	65	27	8	2	NA	2	1	NA	NA	973	974
	Sen	NA	0.100	0.350	0.378	0.447	0.442	0.767	0.829	0.808	0.733	0.788	0.800	NA	0.850	0.800	NA	NA	0.749	0.770
	PPV	NA	1.000	1.000	0.963	0.930	0.919	0.961	0.831	0.731	0.588	0.509	0.475	NA	0.369	0.320	NA	NA	0.898	0.805
	Mis	NA	0.130	0.094	0.092	0.086	0.088	0.038	0.049	0.070	0.111	0.138	0.159	NA	0.232	0.275	NA	NA	0.051	0.080
MSC-P	Freq	NA	NA	2	5	19	32	444	231	117	68	29	10	2	2	3	1	9	974	974
	Sen	NA	NA	0.350	0.240	0.363	0.416	0.749	0.842	0.843	0.838	0.828	0.780	0.600	0.800	0.867	0.800	0.933	0.771	0.770
	PPV	NA	NA	1.000	0.867	0.943	0.931	0.965	0.847	0.755	0.637	0.542	0.447	0.314	0.372	0.347	0.667	0.335	0.858	0.805
	Mis	NA	NA	0.094	0.116	0.095	0.090	0.040	0.044	0.062	0.091	0.124	0.171	0.246	0.225	0.256	0.087	0.312	0.058	0.080
MSHC-P	Freq	NA	NA	2	12	29	38	432	220	107	68	26	16	7	3	3	2	9	974	974
	Sen	NA	NA	0.350	0.375	0.445	0.447	0.750	0.837	0.837	0.849	0.796	0.856	0.829	0.833	0.867	0.850	0.933	0.766	0.770
	PPV	NA	NA	1.000	0.911	0.942	0.944	0.963	0.849	0.752	0.648	0.535	0.522	0.390	0.373	0.347	0.742	0.335	0.858	0.805
	Mis	NA	NA	0.094	0.095	0.084	0.084	0.040	0.045	0.063	0.087	0.129	0.146	0.219	0.227	0.256	0.065	0.312	0.060	0.080

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A40. Ordinal model: Simulation results for true cluster model (E) with an alternative hypothesis (3) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall	
SCIC ₁	Freq	NA	5	7	13	42	50	506	204	69	30	13	5	6	3	7	7	8	975	975
	Sen	NA	0.100	0.171	0.215	0.307	0.402	0.725	0.795	0.784	0.780	0.631	0.680	0.617	0.467	0.843	0.829	0.788	0.696	0.747
	PPV	NA	1.000	0.929	0.929	0.944	0.902	0.973	0.842	0.726	0.640	0.490	0.580	0.658	0.944	0.390	0.507	0.476	0.890	0.812
	Mis	NA	0.130	0.122	0.117	0.103	0.093	0.043	0.050	0.072	0.093	0.148	0.133	0.140	0.082	0.236	0.205	0.254	0.062	0.080
SCIC ₂	Freq	NA	4	2	13	28	45	516	221	82	33	18	3	2	NA	2	4	2	975	975
	Sen	NA	0.100	0.200	0.238	0.350	0.456	0.745	0.806	0.829	0.785	0.806	0.867	0.850	NA	0.900	0.875	1.000	0.735	0.747
	PPV	NA	1.000	1.000	0.955	0.944	0.896	0.970	0.843	0.734	0.638	0.522	0.583	0.609	NA	0.360	0.554	0.294	0.892	0.812
	Mis	NA	0.130	0.116	0.113	0.098	0.087	0.040	0.049	0.066	0.093	0.133	0.116	0.123	NA	0.246	0.181	0.348	0.055	0.080
Gini	Freq	NA	NA	1	3	21	26	361	247	132	73	41	11	12	5	13	14	15	975	975
	Sen	NA	NA	0.200	0.200	0.371	0.435	0.712	0.807	0.830	0.807	0.837	0.882	0.867	0.900	0.915	0.929	0.933	0.762	0.747
	PPV	NA	NA	1.000	1.000	0.953	0.896	0.974	0.850	0.743	0.648	0.537	0.476	0.505	0.434	0.352	0.374	0.284	0.824	0.812
	Mis	NA	NA	0.116	0.116	0.095	0.090	0.044	0.047	0.065	0.089	0.127	0.162	0.159	0.191	0.258	0.264	0.352	0.072	0.080
Elbow	Freq	NA	1	4	15	40	50	580	196	61	17	9	1	NA	NA	1	NA	NA	975	975
	Sen	NA	0.100	0.275	0.313	0.413	0.476	0.762	0.793	0.808	0.612	0.856	0.700	NA	NA	0.900	NA	NA	0.731	0.747
	PPV	NA	1.000	1.000	0.961	0.929	0.891	0.971	0.832	0.717	0.568	0.517	0.500	NA	NA	0.360	NA	NA	0.909	0.812
	Mis	NA	0.130	0.105	0.101	0.091	0.084	0.038	0.052	0.072	0.122	0.135	0.145	NA	NA	0.246	NA	NA	0.051	0.080
MSC-P	Freq	NA	1	2	6	29	32	448	231	103	62	36	5	3	3	7	3	4	975	975
	Sen	NA	0.100	0.300	0.300	0.379	0.444	0.738	0.796	0.856	0.803	0.872	0.960	0.867	0.800	0.900	0.900	0.925	0.753	0.747
	PPV	NA	1.000	1.000	1.000	0.938	0.893	0.975	0.844	0.744	0.642	0.559	0.538	0.542	0.391	0.360	0.472	0.270	0.865	0.812
	Mis	NA	0.130	0.101	0.101	0.095	0.089	0.041	0.050	0.063	0.091	0.119	0.136	0.150	0.213	0.248	0.222	0.373	0.060	0.080
MSHC-P	Freq	NA	1	6	14	48	63	394	214	109	55	32	9	7	6	8	4	5	975	975
	Sen	NA	0.100	0.250	0.350	0.408	0.484	0.722	0.796	0.835	0.800	0.838	0.978	0.786	0.733	0.888	0.775	0.900	0.725	0.747
	PPV	NA	1.000	1.000	0.982	0.960	0.918	0.974	0.843	0.739	0.644	0.546	0.521	0.562	0.568	0.353	0.429	0.264	0.862	0.812
	Mis	NA	0.130	0.109	0.095	0.088	0.081	0.043	0.050	0.065	0.090	0.125	0.140	0.143	0.169	0.254	0.261	0.377	0.065	0.080

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A41. Ordinal model: Simulation results for true cluster model (E) with an alternative hypothesis (4) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall	
SCIC ₁	Freq	NA	6	10	29	49	58	523	155	59	17	16	12	12	5	8	5	7	971	971
	Sen	NA	0.117	0.180	0.234	0.343	0.412	0.700	0.763	0.753	0.735	0.625	0.592	0.650	0.680	0.900	0.760	0.829	0.655	0.717
	PPV	NA	1.000	0.950	0.933	0.939	0.968	0.968	0.837	0.724	0.595	0.515	0.660	0.558	0.699	0.347	0.448	0.493	0.894	0.803
	Mis	NA	0.128	0.120	0.114	0.098	0.087	0.047	0.055	0.076	0.107	0.141	0.132	0.159	0.128	0.261	0.217	0.269	0.068	0.090
SCIC ₂	Freq	NA	3	7	23	38	56	533	176	66	34	13	6	5	1	4	2	4	971	971
	Sen	NA	0.133	0.200	0.257	0.342	0.455	0.720	0.791	0.773	0.838	0.623	0.683	0.680	0.900	0.850	1.000	0.975	0.695	0.717
	PPV	NA	1.000	0.929	0.941	0.928	0.963	0.969	0.838	0.726	0.650	0.494	0.505	0.415	0.900	0.472	0.345	0.296	0.896	0.803
	Mis	NA	0.126	0.118	0.111	0.100	0.082	0.044	0.052	0.074	0.087	0.152	0.169	0.203	0.029	0.217	0.275	0.341	0.061	0.090
Gini	Freq	NA	1	2	7	22	27	399	203	119	62	38	10	13	8	13	22	25	971	971
	Sen	NA	0.100	0.200	0.257	0.336	0.441	0.695	0.779	0.809	0.835	0.761	0.810	0.746	0.763	0.877	0.941	0.940	0.735	0.717
	PPV	NA	1.000	1.000	1.000	0.909	0.969	0.970	0.840	0.747	0.651	0.520	0.517	0.411	0.349	0.384	0.350	0.316	0.821	0.803
	Mis	NA	0.130	0.116	0.108	0.102	0.083	0.047	0.053	0.066	0.087	0.135	0.146	0.198	0.243	0.249	0.275	0.326	0.079	0.090
Elbow	Freq	NA	3	5	28	47	61	588	157	50	17	10	2	1	NA	1	NA	NA	970	971
	Sen	NA	0.167	0.220	0.325	0.394	0.495	0.733	0.778	0.702	0.782	0.620	0.750	0.800	NA	0.800	NA	NA	0.691	0.717
	PPV	NA	1.000	1.000	0.935	0.935	0.948	0.967	0.824	0.690	0.623	0.412	0.417	0.421	NA	0.308	NA	NA	0.912	0.803
	Mis	NA	0.121	0.113	0.101	0.092	0.078	0.042	0.055	0.088	0.096	0.183	0.181	0.188	NA	0.290	NA	NA	0.057	0.090
MSC-P	Freq	NA	2	3	14	29	40	467	184	105	56	28	8	6	4	10	7	8	971	971
	Sen	NA	0.200	0.233	0.257	0.338	0.458	0.719	0.799	0.810	0.834	0.739	0.638	0.767	0.675	0.880	1.000	0.925	0.725	0.717
	PPV	NA	1.000	1.000	0.940	0.931	0.957	0.973	0.841	0.745	0.657	0.514	0.524	0.444	0.494	0.397	0.339	0.285	0.864	0.803
	Mis	NA	0.116	0.111	0.111	0.100	0.082	0.044	0.050	0.067	0.085	0.140	0.161	0.186	0.181	0.243	0.284	0.348	0.066	0.090
MSHC-P	Freq	NA	3	5	21	26	35	426	190	118	63	36	8	7	5	10	8	10	971	971
	Sen	NA	0.200	0.220	0.290	0.315	0.446	0.706	0.806	0.825	0.848	0.761	0.713	0.771	0.740	0.880	1.000	0.930	0.726	0.717
	PPV	NA	1.000	1.000	0.937	0.927	0.957	0.971	0.842	0.746	0.663	0.522	0.423	0.485	0.495	0.397	0.362	0.356	0.852	0.803
	Mis	NA	0.116	0.113	0.106	0.103	0.084	0.046	0.049	0.065	0.083	0.136	0.183	0.170	0.174	0.243	0.264	0.296	0.068	0.090

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A42. Ordinal model: Simulation results for true cluster model (E) with an alternative hypothesis (5) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting	
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall		
SCIC ₁	Freq	NA	1	1	10	30	44	540	236	65	31	12	10	1	1	NA	4	6	992	992	
	Sen	NA	0.100	0.200	0.220	0.347	0.386	0.787	0.847	0.874	0.855	0.733	0.550	0.800	0.800	NA	0.750	0.817	0.768	0.808	
	PPV	NA	1.000	1.000	1.000	0.986	0.902	0.975	0.850	0.760	0.656	0.564	0.594	0.889	0.400	NA	0.595	0.455	0.905	0.832	
	Mis	NA	0.130	0.116	0.113	0.096	0.095	0.034	0.043	0.058	0.084	0.127	0.146	0.043	0.203	NA	0.196	0.266	0.049	0.065	
SCIC ₂	Freq	NA	1	NA	4	24	33	547	254	75	34	14	3	NA	NA	NA	2	1	992	992	
	Sen	NA	0.100	NA	0.250	0.392	0.448	0.805	0.862	0.881	0.865	0.821	0.700	NA	NA	NA	0.800	0.800	0.802	0.808	
	PPV	NA	1.000	NA	1.000	0.934	0.894	0.974	0.851	0.765	0.661	0.548	0.454	NA	NA	NA	0.530	0.800	0.904	0.832	
	Mis	NA	0.130	NA	0.109	0.094	0.088	0.032	0.041	0.056	0.082	0.121	0.179	NA	NA	NA	0.210	0.058	0.044	0.065	
Gini	Freq	NA	NA	1	6	18	20	362	290	134	83	31	9	4	3	5	11	15	992	992	
	Sen	NA	NA	0.300	0.333	0.406	0.495	0.798	0.855	0.870	0.870	0.845	0.878	0.825	0.867	0.900	0.900	0.900	0.940	0.820	0.808
	PPV	NA	NA	1.000	1.000	0.909	0.900	0.975	0.855	0.768	0.662	0.558	0.479	0.448	0.481	0.352	0.352	0.302	0.842	0.832	
	Mis	NA	NA	0.101	0.097	0.094	0.082	0.032	0.041	0.057	0.082	0.117	0.159	0.178	0.184	0.255	0.271	0.343	0.058	0.065	
Elbow	Freq	NA	1	2	23	53	46	583	210	52	15	5	1	NA	NA	NA	1	NA	992	992	
	Sen	NA	0.100	0.300	0.487	0.555	0.530	0.822	0.862	0.846	0.727	0.860	0.500	NA	NA	NA	0.900	NA	0.793	0.808	
	PPV	NA	1.000	0.875	0.968	0.914	0.888	0.973	0.845	0.738	0.595	0.553	0.294	NA	NA	NA	0.281	NA	0.917	0.832	
	Mis	NA	0.130	0.109	0.078	0.073	0.078	0.029	0.042	0.065	0.107	0.119	0.246	NA	NA	NA	0.348	NA	0.042	0.065	
MSC-P	Freq	NA	NA	2	12	25	28	455	277	100	53	23	4	3	NA	3	4	3	992	992	
	Sen	NA	NA	0.300	0.350	0.448	0.507	0.811	0.866	0.876	0.874	0.813	0.800	0.767	NA	0.900	0.825	0.933	0.812	0.808	
	PPV	NA	NA	0.875	1.000	0.918	0.908	0.977	0.853	0.762	0.654	0.544	0.411	0.433	NA	0.442	0.504	0.470	0.881	0.832	
	Mis	NA	NA	0.109	0.094	0.087	0.080	0.030	0.041	0.057	0.084	0.123	0.192	0.184	NA	0.222	0.192	0.242	0.048	0.065	
MSHC-P	Freq	NA	NA	5	18	39	50	409	244	108	65	31	7	3	3	3	4	3	992	992	
	Sen	NA	NA	0.260	0.428	0.490	0.548	0.807	0.862	0.884	0.872	0.858	0.886	0.767	0.900	0.967	0.875	0.933	0.802	0.808	
	PPV	NA	NA	0.950	0.983	0.932	0.922	0.974	0.857	0.765	0.655	0.560	0.464	0.389	0.697	0.416	0.563	0.470	0.874	0.832	
	Mis	NA	NA	0.110	0.085	0.080	0.073	0.031	0.040	0.056	0.084	0.116	0.170	0.208	0.106	0.227	0.174	0.242	0.051	0.065	

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A43. Ordinal model: Simulation results for true cluster model (E) with an alternative hypothesis (6) using elliptical windows.

	1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Maximum reported cluster size (MRCS)		Default Setting
																		Overall		
SCIC ₁	Freq	1	1	4	12	38	42	542	222	63	38	10	4	2	2	3	6	5	995	995
	Sen	0.100	0.100	0.200	0.225	0.321	0.419	0.771	0.839	0.822	0.816	0.870	0.525	0.600	0.500	0.933	0.733	0.880	0.749	0.790
	PPV	1.000	1.000	1.000	1.000	0.949	0.930	0.975	0.851	0.752	0.653	0.571	0.613	0.700	0.729	0.350	0.536	0.560	0.905	0.830
	Mis	0.130	0.130	0.116	0.112	0.101	0.089	0.036	0.044	0.064	0.088	0.113	0.127	0.145	0.101	0.261	0.227	0.223	0.052	0.068
SCIC ₂	Freq	NA	NA	NA	9	29	29	547	231	81	44	16	5	1	NA	1	1	1	995	995
	Sen	NA	NA	NA	0.233	0.376	0.507	0.788	0.845	0.831	0.855	0.888	0.820	0.800	NA	1.000	0.900	0.900	0.785	0.790
	PPV	NA	NA	NA	0.972	0.920	0.928	0.974	0.854	0.747	0.657	0.567	0.426	0.400	NA	0.357	0.321	0.900	0.900	0.830
	Mis	NA	NA	NA	0.113	0.096	0.078	0.034	0.043	0.065	0.085	0.115	0.183	0.203	NA	0.261	0.290	0.029	0.047	0.068
Gini	Freq	NA	NA	NA	4	21	19	375	255	143	88	32	12	10	3	8	14	11	995	995
	Sen	NA	NA	NA	0.475	0.414	0.421	0.767	0.842	0.841	0.868	0.906	0.875	0.860	0.867	0.888	0.921	0.964	0.803	0.790
	PPV	NA	NA	NA	1.000	0.913	0.939	0.972	0.855	0.755	0.666	0.559	0.463	0.517	0.356	0.381	0.431	0.350	0.837	0.830
	Mis	NA	NA	NA	0.076	0.092	0.089	0.037	0.043	0.062	0.082	0.117	0.167	0.158	0.246	0.241	0.233	0.314	0.062	0.068
Elbow	Freq	NA	NA	1	16	47	43	598	201	63	17	7	1	NA	NA	1	NA	NA	995	995
	Sen	NA	NA	0.500	0.369	0.481	0.542	0.805	0.847	0.798	0.782	0.814	0.600	NA	NA	1.000	NA	NA	0.779	0.790
	PPV	NA	NA	1.000	0.984	0.924	0.915	0.973	0.845	0.734	0.613	0.486	0.316	NA	NA	0.357	NA	NA	0.917	0.830
	Mis	NA	NA	0.072	0.092	0.082	0.075	0.031	0.044	0.071	0.101	0.151	0.246	NA	NA	0.261	NA	NA	0.044	0.068
MSC-P	Freq	NA	NA	1	10	35	27	466	239	117	57	25	5	3	1	3	1	5	995	995
	Sen	NA	NA	0.500	0.410	0.443	0.481	0.781	0.861	0.852	0.874	0.888	0.820	0.833	0.900	0.867	0.900	0.880	0.794	0.790
	PPV	NA	NA	1.000	0.975	0.908	0.919	0.974	0.853	0.753	0.672	0.557	0.426	0.435	0.750	0.297	0.321	0.629	0.879	0.830
	Mis	NA	NA	0.072	0.087	0.088	0.083	0.035	0.041	0.062	0.079	0.118	0.183	0.188	0.058	0.314	0.290	0.171	0.051	0.068
MSHC-P	Freq	NA	NA	2	19	52	50	423	211	113	69	29	8	8	1	2	3	5	995	995
	Sen	NA	NA	0.300	0.395	0.448	0.532	0.774	0.854	0.854	0.884	0.900	0.875	0.850	0.800	0.800	0.933	0.940	0.776	0.790
	PPV	NA	NA	1.000	0.954	0.937	0.939	0.971	0.851	0.761	0.662	0.549	0.594	0.495	0.800	0.290	0.359	0.406	0.871	0.830
	Mis	NA	NA	0.101	0.091	0.085	0.074	0.036	0.042	0.060	0.082	0.121	0.136	0.185	0.058	0.312	0.261	0.296	0.056	0.068

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification

Table A44. Ordinal model: Simulation results for true cluster model (E) with an alternative hypothesis (7) using elliptical windows.

		Maximum reported cluster size (MRCS)																		Default Setting
		1%	2%	3%	4%	5%	6%	8%	10%	12%	15%	20%	25%	30%	35%	40%	45%	50%	Overall	
SCIC ₁	Freq	NA	2	5	7	33	51	566	207	71	29	12	3	3	NA	2	3	3	997	997
	Sen	NA	0.100	0.180	0.214	0.397	0.425	0.783	0.857	0.835	0.841	0.783	0.700	0.633	NA	0.700	0.500	0.900	0.763	0.812
	PPV	NA	1.000	1.000	1.000	0.981	0.952	0.976	0.854	0.753	0.660	0.550	0.469	0.557	NA	0.656	0.833	0.634	0.915	0.832
	Mis	NA	0.130	0.119	0.114	0.089	0.087	0.034	0.042	0.063	0.085	0.123	0.164	0.184	NA	0.203	0.097	0.150	0.048	0.063
SCIC ₂	Freq	NA	NA	2	NA	26	36	569	217	78	41	18	4	3	NA	1	NA	2	997	997
	Sen	NA	NA	0.300	NA	0.427	0.486	0.804	0.865	0.853	0.849	0.850	0.825	0.767	NA	1.000	NA	0.850	0.802	0.812
	PPV	NA	NA	1.000	NA	0.937	0.959	0.976	0.850	0.757	0.662	0.561	0.482	0.367	NA	0.313	NA	0.805	0.904	0.832
	Mis	NA	NA	0.101	NA	0.088	0.078	0.031	0.041	0.060	0.084	0.118	0.163	0.222	NA	0.319	NA	0.051	0.044	0.063
Gini	Freq	NA	NA	NA	7	13	18	362	269	156	88	38	11	10	6	2	7	10	997	997
	Sen	NA	NA	NA	0.429	0.431	0.489	0.795	0.862	0.854	0.877	0.861	0.827	0.820	0.883	0.900	0.929	0.980	0.824	0.812
	PPV	NA	NA	NA	0.952	0.895	0.946	0.975	0.861	0.758	0.660	0.564	0.449	0.429	0.382	0.353	0.437	0.355	0.839	0.832
	Mis	NA	NA	NA	0.085	0.091	0.079	0.033	0.040	0.060	0.083	0.116	0.174	0.194	0.234	0.254	0.228	0.313	0.057	0.063
Elbow	Freq	NA	NA	1	24	50	48	615	174	53	21	7	2	2	NA	NA	NA	NA	997	997
	Sen	NA	NA	0.400	0.450	0.576	0.542	0.824	0.847	0.830	0.800	0.857	0.900	0.700	NA	NA	NA	NA	0.793	0.812
	PPV	NA	NA	1.000	0.975	0.914	0.947	0.973	0.842	0.747	0.621	0.516	0.460	0.346	NA	NA	NA	NA	0.921	0.832
	Mis	NA	NA	0.087	0.082	0.070	0.072	0.029	0.045	0.064	0.097	0.135	0.174	0.232	NA	NA	NA	NA	0.042	0.063
MSC-P	Freq	NA	NA	NA	17	39	29	454	225	115	66	28	7	7	2	2	2	4	997	997
	Sen	NA	NA	NA	0.447	0.492	0.534	0.806	0.856	0.851	0.879	0.850	0.914	0.857	0.700	0.850	0.900	0.925	0.804	0.812
	PPV	NA	NA	NA	0.952	0.932	0.946	0.975	0.855	0.753	0.657	0.557	0.454	0.473	0.280	0.523	0.527	0.564	0.874	0.832
	Mis	NA	NA	NA	0.084	0.080	0.072	0.031	0.042	0.061	0.084	0.119	0.172	0.174	0.297	0.181	0.181	0.178	0.051	0.063
MSHC-P	Freq	NA	NA	3	19	30	30	427	223	130	70	35	8	9	5	1	3	4	997	997
	Sen	NA	NA	0.300	0.447	0.493	0.533	0.811	0.873	0.852	0.874	0.869	0.925	0.822	0.860	0.800	0.933	0.925	0.812	0.812
	PPV	NA	NA	1.000	0.969	0.910	0.958	0.972	0.864	0.751	0.654	0.563	0.490	0.494	0.374	0.296	0.590	0.564	0.866	0.832
	Mis	NA	NA	0.101	0.082	0.080	0.071	0.031	0.038	0.062	0.085	0.116	0.156	0.172	0.241	0.304	0.140	0.178	0.052	0.063

Freq, frequency; Sen, sensitivity; PPV, positive predictive value; Mis, misclassification



국문요약

다항분포를 기반으로 하는 공간 검색 통계량에서

최대 도출 군집 크기의 최적화

공간 질병 군집을 정확하게 탐색하는 것은 역학 및 공중보건 분야에서 중요한 문제이다. 공간 검색 통계량은 공간 역학 및 질병 감시에서 공간 클러스터를 탐색하기 위한 방법으로 널리 사용되는 방법 중 하나이다. 공간 검색 통계량은 베르누이, 포아송, 지수, 순서형, 정규 및 다항 모형과 같은 다양한 확률 모형에 대하여 개발되었다. 공간 검색 통계량을 사용할 때 연구자들은 최대 도출 군집 크기를 지정해 주어야 한다. 많은 연구에서 최대 도출 군집 크기를 기본 설정인 전체 인구의 50%로 설정하여 분석한다. 그러나 최대 도출 군집 크기를 기본 설정인 50%로 지정하면 덜 의미 있는 지역을 포함하여 실제 클러스터보다 큰 클러스터가 보고될 수 있다. 최근 연구에서 최적의 최대 도출 군집 크기를 선택하기 위하여 지니계수를 사용한 최적화 기준이 개발되었다. 그러나 지니계수 방법은 명목형인 범주형 자료에 적용하기 어렵기 때문에 다항 모형 기반의 공간 검색 통계량에 대한 지니계수를 명확히 정의하기는 어렵다. 따라서 본 연구에서는 다항 모형 및 순서형 모형 기반의 공간 검색 통계량에서 최적의 최대 도출 군집 크기를 선택하기 위하여 공간



클러스터 정보 기준(SCIC) 방법을 제안하고자 한다. 제안한 SCIC 방법은 우도 기반의 방법으로 모든 확률 모형에 대하여 적용이 가능한 방법이다. 본 연구에서는 순서형 모형 기반의 공간 검색 통계량과 비교하여 다향 모형 기반의 공간 검색 통계량에서 제안한 SCIC 방법의 성능을 평가하기 위한 시뮬레이션 연구를 수행하였다. 시뮬레이션 연구에서 사용한 성능 평가 지표는 민감도, 양성예측도, 오분류율이다. 또한 본 연구에서 제안한 SCIC 방법의 유용성을 설명하기 위하여 지역 사회건강조사 자료를 사용한 실제 자료 분석을 수행하였다.

핵심 되는 말: 공간 군집 탐색, 공간 검색 통계량, 다향 모형, 최대 도출 군집 크기