STATISTICS IN MEDICINE
Statist. Med. 2006; **25**:841–852
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.2419

An extended power of cluster detection tests

Kunihiko Takahashi*,† and Toshiro Tango

Department of Technology Assessment and Biostatistics, National Institute of Public Health, 2-3-6 Minami, Wako, Saitama 351-0197, Japan

SUMMARY

Several tests have been proposed to detect spatial disease clustering without prior information on their locations. In order to compare the performance of these tests, most authors employ the usual power, i.e. the rejection probability of the null hypothesis of no clustering due to various reasons. However, the usual power is not always appropriate for evaluating the cluster detection tests since their purpose is to both reject the null hypothesis and identify the cluster areas accurately. In this paper, we propose an extended power of the cluster detection tests, which includes the usual power as a special case. Further, we define the profile of the extended power, which can be expected to play an important role in the evaluation and comparison of several cluster detection tests. The proposed extended power and its profile are demonstrated by two tests—Kulldorff's circular spatial scan statistic and a flexible spatial scan statistic proposed by Tango and Takahashi. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: cluster detection; hot-spot clusters; hypothesis testing; power; spatial epidemiology

1. INTRODUCTION

Several test statistics have been proposed to detect for spatial clustering, which have been applied to a wide variety of epidemiological studies for spatial disease cluster detection [1, 2]. In particular, Kulldorff's circular spatial scan statistic [3, 4] has been used extensively along with his software SaTScan [5]. Tests for spatial randomness can be classified according to their purpose. Focused tests have been developed to detect the existence of a local cluster around a predetermined point source, while general tests search for clusters without any preconceived assumptions about their location [6]. There are two types of general tests. *Cluster detection tests* (CDTs) such as that developed by Turnbull *et al.* [7], Besag and Newell [6], Kulldorff and Nagarwalla [3], Tango [8], Duczmal and Assunção [9], Patil and Taillie [10] and Tango and Takahashi [11] are used to both detect local clusters, without any prior information on their

*Correspondence to: Kunihiko Takahashi, Department of Technology Assessment and Biostatistics, National Institute of Public Health, 2-3-6 Minami, Wako, Saitama 351-0197, Japan.
†E-mail: kunihiko@niph.go.jp

Received 23 August 2005
Accepted 1 September 2005

location, and to determine their statistical significance. On the other hand, *global clustering tests*, such as those developed by Moran [12], Whitemore *et al.* [13], Oden [14], Tango [15], Rogerson [16] and Bonetti and Pagano [17], are used to detect the presence of clusters in a study area without determining the statistical significance of individual clusters.

In order to compare the performance of these tests, the usual power has been treated in the same manner as in the usual hypothesis tests [4, 18]. In recent power comparisons of the disease clustering tests, Kulldorff's circular scan statistic was demonstrated to have the best power for detecting localized clusters [19]. However, it should be noted that the power estimates reflected the 'power to reject the null hypothesis for whatever reasons,' while the probability of both rejecting the null hypothesis and accurately identifying the true cluster is a different matter altogether. In order to compare the performance of the CDTs, Tango and Takahashi [11] proposed a bivariate power distribution classified according to the number of regions detected as the *most likely cluster* (MLC) and the number of true hot-spot regions included in the MLC. Since the bivariate power distribution contains all the information on the performance of any procedure for detecting the hot-spot clusters, any summary measures including the usual power for evaluating the CDT can be based on the power distribution.

In this paper, we propose an extended power of the CDTs, which includes the usual power as a special case, and a profile of the extended power for evaluating and comparing several CDTs. The proposed extended power and its profile will be demonstrated by two tests, namely, Kulldorff's circular spatial scan statistic [3, 4] and a flexible spatial scan statistic proposed by Tango and Takahashi [11] along with their software FleXScan [20].

2. MOTIVATING EXAMPLE

As a motivating example of our study, we present the results of the application of three CDTs—the scan statistics of Kulldorff's, Duczmal and Assunção's, and Tango and Takahashi—to a simulated disease map in the areas of Tokyo Metropolis and Kanagawa prefecture in Japan wherein there are $m=113$ regions that comprise wards, cities, and villages (Figure 1). In this map, we simulated a random sample of $n=235$ cases by assuming the hot-spot cluster regions $\mathbf{C}=\{14,15,26,27\}$ whose relative risk was set to 3.0 and the cases to be Poisson distributed (see Reference [11] for details). In this application, we selected a maximum length of $K=15$ for the MLC in order to achieve a reasonably accurate comparison and a number $B=999$ for the Monte Carlo replications. The results are summarized as follows.

- Kulldorff's scan statistic detected $l=2$ regions $\{14,15\}$ as the MLC with a log likelihood ratio of 20.1 and $p=1/(999+1)=0.001$, and the estimated relative risk was $\hat{\theta}=3.47$.
- Tango and Takahashi's scan statistic detected $l=5$ regions $\{14,15,26,27,33\}$ as the MLC with a log likelihood ratio of 29.7 and $p=0.001$, and the estimated relative risk was $\hat{\theta}=3.41$.
- Duczmal and Assunção's scan statistic detected $l=15$ connected regions $\{14,15,24,26,27,31,32,33,48,54,69,77,78,90,110\}$ as the MLC with a log likelihood ratio of 31.8 and $p=0.001$, and the estimated relative risk was $\hat{\theta}=2.40$.

Figure 1. The 113 regions that comprise wards, cities, and villages in the areas of Tokyo Metropolis and Kanagawa prefecture in Japan. The region number used in the text is also indicated. In particular, the region numbers of two hot-spot clusters considered in this paper are $\mathbf{A} = \{14, 15, 20\}$ (circular) and $\mathbf{C} = \{14, 15, 26, 27\}$ (non-circular) (see Reference [11] for details).

All these tests rejected the null hypothesis of 'H_0: there are no clusters,' but detected different regions as MLC. In particular, Duczmal and Assunção's scan statistic detected a cluster of a peculiar shape that was considerably larger than the true cluster. This example is sufficient to show that the usual power is an inappropriate measure of the performance of a CDT. Therefore, a different measure that can reflect the extent of misclassification is required.

We consider two types of misclassifications when applying the CDT. One is a *false negative test result* (FN) in which the CDT misses a region included in the true cluster. The other is a *false positive test result* (FP) in which the CDT incorrectly detects a region that is not present in the true cluster. In the above example, Kulldorff's scan statistic has the regions $\{26, 27\}$ as the FNs, but it does not have any FPs. Tango and Takahashi's scan statistic has the region $\{33\}$ as the FP, and Duczmal and Assunção's scan statistic has 11 regions $\{24, 31, 32, 33, 48, 54, 69, 77, 78, 90, 110\}$ as the FPs. However, both these tests do not have any FNs.

In the next section, we propose the extended power of the CDTs, which is based on the bivariate power distribution proposed by Tango and Takahashi [11] and the newly introduced penalties for the FPs and FNs.

3. EXTENDED POWER

In order to compare the performance of the CDTs, Tango and Takahashi [11] proposed a bivariate power distribution based upon Monte Carlo simulation. They introduced a bivariate power distribution $P(l,s)$ classified according to the length l of the significant MLC and the number s of the assumed hot-spot regions included therein

$$P(l,s) = \frac{\#\{\text{significant MLC has length } l \text{ and includes } s \text{ true regions}\}}{\#\{\text{trials for each simulation}\}} \tag{1}$$

where $l \geqslant 1$ and $s \geqslant 0$. Based on $P(l,s)$, we examined the following powers:

1. The usual power, i.e. $P(+,+) = \sum_{l\geqslant 1}\sum_{s\geqslant 0} P(l,s)$.
2. The joint power $P(l,s)$, especially $P(s^*,s^*)$ where s^* is the length of the hot-spot cluster assumed in the simulation.
3. The marginal power distribution of s ($\geqslant 0$), $P(+,s) = \sum_{l\geqslant 1} P(l,s)$ and its conditional power $P(+,s)/P(+,+)$.
4. The marginal power distribution of l ($\geqslant 1$), $P(l.+) = \sum_{s\geqslant 0} P(l,s)$.

They prepared tables of $P(l,s)$ for the following two hot-spot cluster models:

- $\mathbf{A} = \{14, 15, 20\}$ (circular cluster; $s^* = 3$);
- $\mathbf{C} = \{14, 15, 26, 27\}$ (non-circular cluster; $s^* = 4$).

The powers are calculated for tests of nominal α levels of 0.05 and for the expected total number of cases 200 under the null hypothesis, which are based on Monte Carlo simulation using Poisson random numbers. For each simulation, 1000 trials were carried out. The resultant power distributions $P(l,s) \times 1000$ are reproduced in Tables I and II for each of the cluster models, respectively, in the form of cross table classified by l ('*length*' in tables) and s ('*include*' in tables).

Both tests have high usual powers for the circular cluster **A** (Table I), while Tango and Takahashi's scan statistic has higher power for the non-circular cluster **C** (Table II). Table I shows that Kulldorff's scan statistic detects the circular cluster **A** considerably accurately with power $P(3,3) = 738/1000$, while Table II shows that it exhibits zero power $P(4,4) = 0/1000$ for detecting the non-circular cluster **C** accurately. On the other hand, Tango and Takahashi's scan statistic does not exhibit such high power for identifying the clusters accurately, $P(3,3) = 142/1000$ for the cluster **A** and $P(4,4) = 138/1000$ for the cluster **C**. However, the power distribution appears to be concentrated in a relatively narrow range of the length l on the line $s = s^*$, thereby indicating that the observed significant MLC contains the true cluster with a considerably high probability. In particular, for the cluster **C** with length $s^* = 4$, the marginal power of Tango and Takahashi's scan statistic $P(+,s^*) = 850/1000$ and its conditional marginal power $P(+,s^*)/P(+,+) = 850/890$ are much higher than that of Kulldorff's scan statistic, $P(+,s^*) = 254/1000$ and $P(+,s^*)/P(+,+) = 254/801$, respectively. Furthermore, Kulldorff's scan statistic exhibits a greater tendency to detect a larger cluster than the true cluster as compared with that of Tango and Takahashi. For example, the probability that the length of MLC for the cluster **C** ($s^* = 4$) is greater than or equal to 12 is 213/1000 compared with 2/1000 for Tango and Takahashi's scan statistic. This tendency is shown even in the circular cluster **A** where the same probabilities are 35/1000 *versus* 2/1000. Therefore, the bivariate power distribution can be considered to provide very

Table I. Comparison between the bivariate power distribution $P(l,s)$ of Kulldorff's circular scan statistic and Tango and Takahashi's flexible scan statistic for the hot-spot cluster $\mathbf{A}=\{14,15,20\}$. The nominal α-level was set to 0.05 and 1000 trials were carried out [11].

Kulldorff						Tango and Takahashi					
	Include s hot-spot regions						Include s hot-spot regions				
Length l	0	1	2	3	Total	Length l	0	1	2	3	Total
1	0	0			0	1	0	0			0
2	1	0	0		1	2	0	0	0		0
3	0	0	0	738	738	3	0	0	0	142	142
4	0	0	0	134	134	4	0	0	0	116	116
5	0	0	0	39	39	5	0	0	0	137	137
6	0	0	0	12	12	6	0	0	0	149	149
7	0	0	0	9	9	7	0	0	0	165	165
8	0	0	0	1	1	8	0	0	0	131	131
9	0	0	2	3	5	9	0	0	0	84	84
10	0	0	0	2	2	10	0	0	0	27	27
11	0	0	0	4	4	11	0	0	0	11	11
12	0	0	0	12	12	12	0	0	0	2	2
13	0	0	0	14	14	13	0	0	0	0	0
14	0	0	0	3	3	14	0	0	0	0	0
15	0	0	0	6	6	15	0	0	0	0	0
Total	1	0	2	977	980	Total	0	0	0	964	964
	Usual power = 0.980						Usual power = 0.964				

useful information for evaluating the CDT performance. An observation of the bivariate power distribution suggests that it can be summarized into a single index that takes a value in the interval $[0,1]$.

We now define the extended power of the CDTs based on the bivariate distribution $P(l,s)$ by introducing penalties for the FPs and FNs as

$$I(w^-,w^+)=\sum_{l\geqslant 1}\sum_{s\geqslant 0} W(l,s;w^-,w^+)P(l,s) \tag{2}$$

where $W(l,s;w^-,w^+)$ is a weight function such that

$$W(l,s;w^-,w^+)=\begin{cases}\sqrt{(1-\min\{w^-(s^*-s),1\})\,(1-\min\{w^+(l-s),1\})}\\ \quad (s\leqslant l; 0\leqslant s\leqslant s^*, 1\leqslant l)\\ 0 \quad (\text{otherwise})\end{cases} \tag{3}$$

and w^- and w^+ are the predefined penalties for the FNs and FPs (per region), respectively, with the following constraint:

$$0\leqslant w^+\leqslant w^-\leqslant 1 \tag{4}$$

Table II. Comparison between the bivariate power distribution $P(l,s)$ of Kulldorff's circular scan statistic and Tango and Takahashi's flexible scan statistic for the non-circular hot-spot cluster $\mathbf{C}=\{14,15,26,27\}$. The nominal α-level was set to 0.05 and 1000 trials were carried out [11].

Kulldorff							Tango and Takahashi						
	Include s hot-spot regions							Include s hot-spot regions					
Length l	0	1	2	3	4	Total	Length l	0	1	2	3	4	Total
1	1	0				1	1	0	0				0
2	0	0	351			351	2	0	0	0			0
3	2	0	4	0		6	3	0	0	0	0		0
4	0	0	3	0	0	3	4	0	0	0	0	138	138
5	2	0	2	0	0	4	5	0	0	0	3	147	150
6	1	0	0	0	0	1	6	1	0	0	2	200	203
7	0	0	0	81	0	81	7	0	1	0	4	147	152
8	0	0	10	18	38	66	8	0	0	2	9	107	118
9	0	0	2	0	26	28	9	0	0	0	10	71	81
10	0	0	0	29	3	32	10	1	0	2	5	28	36
11	0	0	1	13	1	15	11	0	0	0	0	10	10
12	0	0	2	4	60	66	12	0	0	0	0	2	2
13	0	0	0	5	62	67	13	0	0	0	0	0	0
14	0	0	0	10	27	37	14	0	0	0	0	0	0
15	0	0	0	6	37	43	15	0	0	0	0	0	0
Total	6	0	375	166	254	801	Total	2	1	4	33	850	890
	Usual power = 0.801							Usual power = 0.890					

The inequality between w^- and w^+ is set because it is natural to consider that an FN is more important for a CDT than an FP. It should be noted that $W(l,s;w^-,w^+)$ implies the geometric mean of '1—penalty for FNs' and '1—penalty for FPs,' hence we can easily obtain $0 \leqslant I \leqslant 1$. Next we consider the following three special cases:

1. $w^- = w^+ = 0$, i.e. $I(0,0) = \sum_{l\geqslant 1}\sum_{s\geqslant 0} P(l,s)$ equals the usual power.
2. $w^- = w^+ = 1$, i.e. $I(1,1) = P(s^*,s^*)$ denotes the power to detect the true cluster accurately.
3. $w^- = 1,\ w^+ = 0$, i.e. $I(1,0) = \sum_{l\geqslant 1} P(l,s^*)$ denotes the power for which the MLC includes all the regions within the true cluster.

The weight function $W(l,s;w^-,w^+)$ and the estimated extended power $I(w^-,w^+)$ for these three special cases are shown in Table III for Kulldorff's and Tango and Takahashi's scan statistics using the simulated results of Table I (circular cluster model **A** with $s^*=3$). All three results show that Kulldorff's scan statistic is better than that of Tango and Takahashi's for the circular cluster **A**. However, the difference varies with the values of w^- and w^+.

A natural procedure of determining these two penalties will be derived as follows. First, we consider the penalty for the FNs given by

$$\min\{w^-(s^*-s),1\}$$

Table III. Weight function $W(l, s; w^-, w^+)$ and estimated extended power $I(w^-, w^+)$ for Kulldorff's circular and Tango and Takahashi's flexible scan statistic tests for the circular hot-spot cluster **A**.

	Include s hot-spot regions					Include s hot-spot regions			
Length l	0	1	2	3	Length l	0	1	2	3
		$W(l,s;0,0)$					$W(l,s;1,1)$		
1	1	1			1	0	0		
2	1	1	1		2	0	0	0	
3	1	1	1	1	3	0	0	0	1
4	1	1	1	1	4	0	0	0	0
5	1	1	1	1	5	0	0	0	0
6	1	1	1	1	6	0	0	0	0
7	1	1	1	1	7	0	0	0	0
8	1	1	1	1	8	0	0	0	0
9	1	1	1	1	9	0	0	0	0
10	1	1	1	1	10	0	0	0	0
11	1	1	1	1	11	0	0	0	0
12	1	1	1	1	12	0	0	0	0
13	1	1	1	1	13	0	0	0	0
14	1	1	1	1	14	0	0	0	0
15	1	1	1	1	15	0	0	0	0
Kulldorff's $I(0,0)=0.980$					Kulldorff's $I(1,1)=0.738$				
Tango and Takahashi's $I(0,0)=0.964$					Tango and Takahashi's $I(1,1)=0.142$				
		$W(l,s;1,0)$					$W(l,s;1/3,1/3)$		
1	0	0			1	0	0.577		
2	0	0	0		2	0	0.471	0.816	
3	0	0	0	1	3	0	0.333	0.667	1
4	0	0	0	1	4	0	0	0.471	0.816
5	0	0	0	1	5	0	0	0	0.577
6	0	0	0	1	6	0	0	0	0
7	0	0	0	1	7	0	0	0	0
8	0	0	0	1	8	0	0	0	0
9	0	0	0	1	9	0	0	0	0
10	0	0	0	1	10	0	0	0	0
11	0	0	0	1	11	0	0	0	0
12	0	0	0	1	12	0	0	0	0
13	0	0	0	1	13	0	0	0	0
14	0	0	0	1	14	0	0	0	0
15	0	0	0	1	15	0	0	0	0
Kulldorff's $I(1,0)=0.977$					Kulldorff's $I(1/3,1/3)=0.870$				
Tango and Takahashi's $I(1,0)=0.964$					Tango and Takahashi's $I(1/3,1/3)=0.316$				

where $s=1,\ldots,s^*$. If $s=s^*$ the FNs are eliminated and the penalty is set to zero. If $s=0$, the penalty for the FNs attains the maximum value of 1, i.e.

$$w^-(s^*-0)=1 \quad \text{or} \quad w^-=1/s^* \tag{5}$$

On the other hand, the penalty for the FPs is given by

$$\min\{w^+(l-s),1\}$$

where $l = s, \ldots, m$. If $l = s$, the FPs are eliminated and the penalty is set to zero. We assume that the *maximum allowable number* of regions detected as the FPs is $l_0(\geqslant s^*)$. Then, when $l = s + l_0$, the penalty for the FPs attains the maximum value of 1, i.e.

$$w^+ l_0 = 1 \quad \text{or} \quad w^+ = 1/l_0 \tag{6}$$

Generally, l_0 should be determined according to the ratio of the relative weight of the FP to that of the FN, i.e. w^+/w^-, which might depend on the cost of misclassifications.

For the previous examples on the circular cluster model **A**, we consider the following two cases for $l_0 = s^*$ and $l_0 = 2s^*$:

1. $l_0 = s^* = 3$: $I(1/3, 1/3) = 0.870$ (Kulldorff) and 0.316 (Tango and Takahashi). The weight function $W(l, s; 1/3, 1/3)$ for this case is shown in Table III.
2. $l_0 = 2s^* = 6$: $I(1/3, 1/6) = 0.906$ (Kulldorff) and 0.614 (Tango and Takahashi).

For the non-circular cluster model **C** shown in Table II, we estimated the following values:

1. $I(0, 0) = 0.801$ (Kulldorff) and 0.890 (Tango and Takahashi);
2. $I(1, 1) = 0.000$ (Kulldorff) and 0.138 (Tango and Takahashi);
3. $I(1, 0) = 0.254$ (Kulldorff) and 0.850 (Tango and Takahashi);
4. $I(1/4, 1/4) = 0.253$ (Kulldorff) and 0.483 (Tango and Takahashi);
5. $I(1/4, 1/8) = 0.371$ (Kulldorff) and 0.719 (Tango and Takahashi).

These results indicate that Tango and Takahashi's scan statistic can detect the non-circular clusters such as **C** more accurately as compared with that of Kulldorff's.

4. PROFILE OF THE EXTENDED POWER

In several cases, it is difficult to set the value of l_0 in advance. For dealing with such situations, we consider a new measure $I(w^-, w^+)$ to be a function of the ratio $r = w^+/w^-$ $(0 \leqslant r \leqslant 1)$ with $w^- = 1/s^*$. Then, we define the *profile of the extended power* as

$$Q(r \mid s^*) = I(1/s^*, r/s^*), \quad (0 \leqslant r \leqslant 1) \tag{7}$$

which represents the extended power continuously for all the values of r where $l_0 = s^*/r$. Figure 2 shows the plots of the profile $Q(r \mid s^*)$ against r $(0 \leqslant r \leqslant 1)$ for the two types of scan statistics applied to the circular cluster model **A** (Figure 2(a)) and the non-circular cluster model **C** (Figure 2(b)). The profile $Q(r \mid s^*)$, where $r = 1$ implies the extended power when $w^- = w^+$, i.e. $l_0 = s^*$. On the other hand, $r = 0.5$ implies $l_0 = 2s^*$, and $r = 0$ implies that $l_0 = \infty$ and the penalties for the FPs is 0. In Figure 2(a), Kulldorff's scan statistic is shown to be uniformly more powerful than that of Tango and Takahashi. The difference in the values of I is very small when r is small, but it increases considerably with r. On the other hand, Figure 2(b) shows that Tango and Takahashi's scan statistic is uniformly more powerful than that of Kulldorff, and the difference in the values of I appears almost constant irrespective of the values of r. In certain cases, depending on the shape of the true cluster, there will be situations that two values of I cross each other at some point r'. Therefore, the profile of the extended power is expected to play an important role in the simultaneous evaluation and comparison of several CDTs.

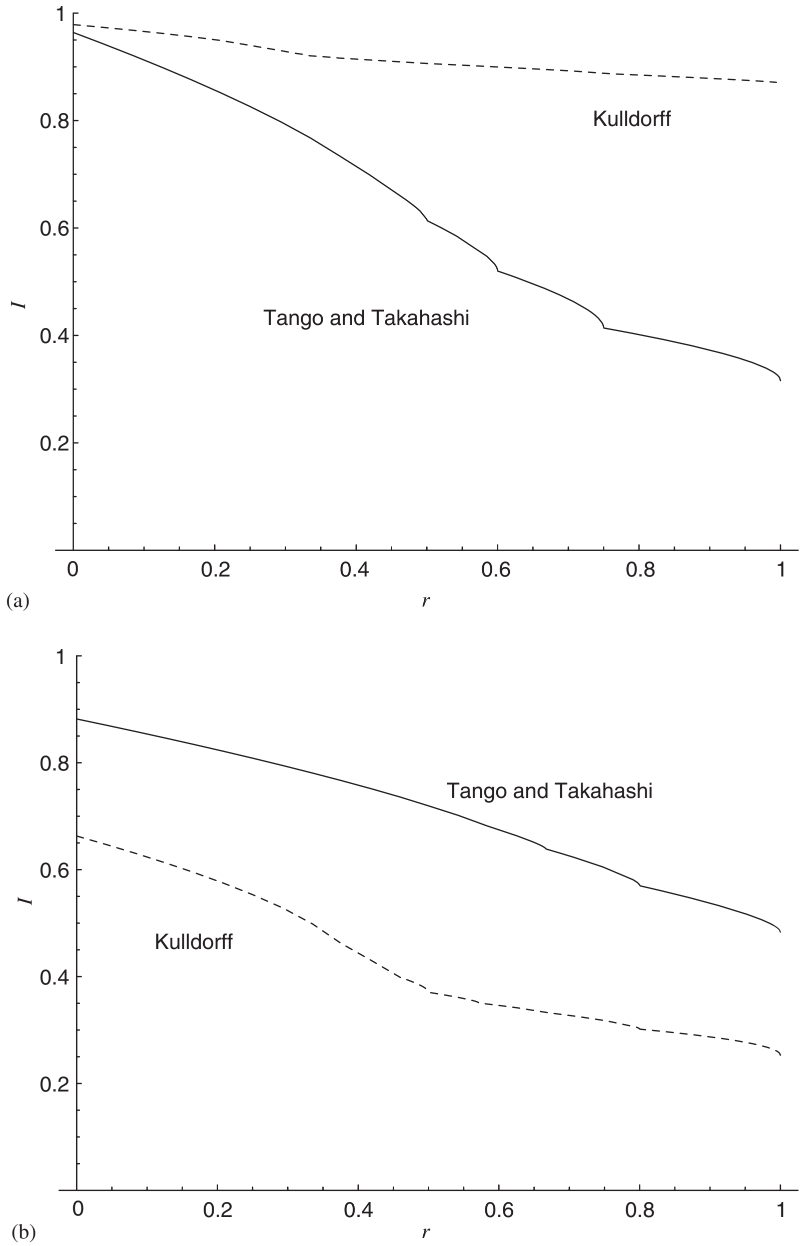

Figure 2. Profile of the extended power $Q(r \mid s^*) = I(1/s^*, r/s^*)$ for Kulldorff's circular and Tango and Takahashi's flexible scan statistics: (a) $\mathbf{A} = \{14, 15, 20\}$ ($s^* = 3$); and (b) $\mathbf{C} = \{14, 15, 26, 27\}$ ($s^* = 4$).

5. GUIDANCE FOR THE APPLICATION

In this section, we will give some guidance for the application of the proposed procedures in order to investigate the performance of some CDT. At first, we consider a situation where the number of cluster is assumed to be one.

Step 0: As an entire study region, use a real region such as Tokyo Metropolitan area or area of upstate New York where real regional populations and number of cases of some disease under study are available.

Step 1: Assume a true (practically feasible) cluster model where the location, the length s^* and its relative risk should be determined. Set $P(l,s)=0$.

Step 2: Under the above alternative hypothesis, generate a disease map based upon a random sample of n total number of cases. If possible, n should be set equal to the actual total number of cases of disease under study.

Step 3: Apply the CDT to the random sample. If we could identify the significant MLC with l and s, then set $P(l,s) \leftarrow P(l,s)+1$.

Step 4: Repeat steps 2–3 on a large number of random replications (e.g. 1000 times) and estimate the bivariate power distribution (1).

Step 5: If we can easily set the value of l_0, then calculate the extended power (2) where $w^- = 1/s^*$ and $w^+ = 1/l_0$. Otherwise, plot the profile of the extended power (7).

Of course, we can repeat the above procedure for a single different cluster model.

Next, we consider a situation when multiple non-overlapping clusters, say k kinds, of different size and shape exist in the study area. In this case, the proposed extended power and its profile can be applied in many ways. One simple method will be to estimate a single integrated bivariate power distribution where multiple true clusters are combined into an integrated disconnected cluster area wherein the length of cluster area is defined as the total number of lengths of individual cluster, i.e. $s^* = s_1^* + s_2^* + \cdots, s_k^*$. To perform this simple method for multiple clusters, steps 1 and 3 of the procedure above for a single cluster model should be modified as follows:

Step 1: Assume a cluster model consisting of k non-overlapping true clusters where the locations, the length $s_1^*, s_2^*, \ldots, s_k^*$ and their relative risks $\theta_1, \theta_2, \ldots, \theta_k$ should be determined. Set $P(l,s)=0$.

Step 3: Apply the CDT to the random sample. If we could identify m significant clusters, i.e. the significant MLC with length l_1, the significant secondary cluster with length $l_2, \ldots,$ and the mth cluster with length l_m, then set $l \leftarrow l_1 + l_2 + \cdots + l_m$ and count s true regions included in the identified l regions. Then, set $P(l,s) \leftarrow P(l,s)+1$.

6. DISCUSSION

Though the usual power has been widely used for evaluating the CDT, the usual power is not always appropriate since the purpose of CDT is to both reject the null hypothesis and identify the cluster areas accurately. In order to evaluate the CDT, the key problem is two types of misclassifications. One is FN, and the other is FP. Then, Tango and Takahashi [11] proposed a bivariate power distribution classified according to the number of regions detected as the MLC and the number of true hot-spot regions included in the MLC. The bivariate power distribution shows us very useful information for evaluating the CDT performance. However, an observation of the bivariate power distribution suggests that it can be summarized into a single index.

In this paper, we proposed a new measure termed extended power, which takes a value in the interval $[0, 1]$, to evaluate the performance of the CDTs. This extended power is defined as the weighted sum of the bivariate power distribution $P(l, s)$ wherein the weight is given by the geometric mean of (1—penalty for the FNs) and (1—penalty for the FPs). This measure includes the usual power as a special case.

The extended power could be calculated by using different values of w^- and w^+, which means that the researcher could set different values under different situation. Any values satisfying constraint (4) could be used depending on the situations. This expandability might be desirable for some researchers. But for the researchers who are not familiar with this, it may be unhelpful. Then, we proposed the quantities $w^- = 1/s^*$ and $w^+ = 1/l_0$ as the values of the penalties for FN and FP, respectively, where s^* is the length of the assumed true cluster, and l_0 is the maximum allowable number of regions detected as the FPs. Theoretically, l_0 should be determined according to the ratio of the relative weight of the FP to that of the FN. In general, however, we think that $l_0 = s^* \sim 3s^*$ are reasonable choices.

If we can estimate the average cost per region of the FNs and FPs, we can consider the cost comparison proposed by Tango and Takahashi [11]. Let C^- and C^+ denote the average costs of the FNs and FPs per region, and the random variables L and S denote the length l of a detected cluster and the number of true regions s included in the detected cluster, respectively. Then, $s^* - S$ represents the number of FNs, and $L - S$ represents the number of FPs. The expected total cost C is then given by

$$C = C^-\{E(s^* - S) + r_c E(L - S)\}$$

where $r_c = C^+/C^-$ is the ratio of the average cost of FNs to the average cost of FPs. In this case, we also assume $r_c \leqslant 1$, similar to the case of the ratio r of the two penalties. Therefore, it might be possible to set $r = r_c$ for the proposed extended power.

However, in many situations, it might be difficult to set the value of l_0 in advance. Then, we proposed the profile of the extended power $Q(r \mid s^*)$, which can conveniently be used to evaluate and compare different CDTs without fixing the value of the ratio $r = w^+/w^-$ with $w^- = 1/s^*$. The profile could be plotted automatically based upon the bivariate power distribution $P(l, s)$. This profile is expected to play an important role in the simultaneous evaluation and comparison of several CDTs. The performance of CDT could be illustrated by the plots of the profile clearly. Furthermore, if we can assume some prior distribution of r, we can also consider the expected value of extended power as

$$E[Q(r \mid s^*)] = \int_0^1 r\,p(r)\,\mathrm{d}r \tag{8}$$

where $p(r)$ is a probability density function of r. For example, when r is distributed according to the uniform distribution $\mathrm{U}(0, 1)$, an area under the profile in Figure 2 implies the expected power, and using this index, we can compare the performance of Kulldorff's scan statistic and that of Tango and Takahashi.

In the Monte Carlo simulation performed in this study, the number of clusters was assumed to be one. However, the proposed extended power of the CDTs and its profile can be applied for multiple clusters of different shapes. In Section 5, we described a simple method in which we apply the extended power and its profile to an integrated cluster area as an overall performance measure. However, this simple method ignores the variability of the FP and the FN for individual true cluster and so, the method which can take such variability into account might be better depending on the situation. We would like to leave such a method in our future work.

Finally, it should be noted that the proposed extended power and its profile are solely based upon a Monte Carlo simulation study, not on some theoretical basis due primarily to the fact that it is quite difficult to derive theoretical bivariate power distribution for alternative cluster models.

REFERENCES

1. Lawson A, Biggeri A, Böhning D, Lesaffre E, Viel JF, Bertollini R (eds). *Disease Mapping and Risk Assessment for Public Health*. Wiley: New York, 1999.
2. Waller LA, Gotway CA. *Applied Spatial Statistics for Public Health Data*. Wiley: New York, 2004.
3. Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. *Statistics in Medicine* 1995; **14**:799–810.
4. Kulldorff M. A spatial scan statistic. *Communications in Statistics* 1997; **26**:1481–1496.
5. Kulldorff M. *SaTScan v4.0*: *Software for the Spatial and Space-time Scan Statistics*. Information Management Services Inc., http://www.satscan.org/, 2004.
6. Besag J, Newell J. The detection of clusters in rare diseases. *Journal of the Royal Statistical Society*, *Series A* 1991; **154**:143–155.
7. Turnbull B, Iwano E, Burnett W, Howe H, Clark L. Monitoring for clusters of disease: application to leukemia incidence in upstate New York. *American Journal of Epidemiology* 1990; **132**(suppl.):136–143.
8. Tango T. A test for spatial disease clustering adjusted for multiple testing. *Statistics in Medicine* 2000; **19**:191–204.
9. Duczmal L, Assunção R. A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters. *Computational Statistics and Data Analysis* 2004; **45**:269–286.
10. Patil GP, Taillie C. Upper level set scan statistic for detecting arbitrarily shaped hotspots. *Environmental and Ecological Statistics* 2004; **11**:183–197.
11. Tango T, Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics* 2005; **4**:11.
12. Moran PAP. Notes on continuous stochastic phenomena. *Biometrika* 1950; **37**:17–23.
13. Whitemore AS, Friend N, Brown BW, Holly EA. A test to detect clusters of disease. *Biometrika* 1987; **74**:631–635.
14. Oden N. Adjusting Moran's I for population density. *Statistics in Medicine* 1995; **14**:17–26.
15. Tango T. A class of tests for detecting 'general' and 'focused' clustering of rare diseases. *Statistics in Medicine* 1995; **14**:2323–2334.
16. Rogerson PA. The detection of clusters using a spatial version of the chi-square goodness-of-fit statistic. *Geographical Analysis* 1999; **31**(1):130–147.
17. Bonetti M, Pagano M. The interpoint distance distribution as a descriptor of point patterns, with an application to spatial disease clustering. *Statistics in Medicine* 2005; **24**:753–773.
18. Glaz J, Naus J, Wallenstein S. *Scan Statistics*. Springer: Berlin, 2001.
19. Kulldorff M, Tango T, Park PJ. Power comparisons for disease clustering tests. *Computational Statistics and Data Analysis* 2003; **42**:665–684.
20. Takahashi K, Yokoyama T, Tango T. *FleXScan*: *Software for the Flexible Spatial Scan Statistic*. National Institute of Public Health: Japan, 2005.

