表3　われわれが解析に使用するコード表およびコードの変遷(続)

Site of Cancer	ICD-O-2 Codes	ICD-9 Codes	ICD-10 Codes	本研究班で使用するコード
Multiple myeloma	morphology 9731-9732	203.0, 238.6	C90.0, C90.2	C90.0, C90.2
Leukemias	morphology 9800-9949	204-208, 202.4, 203.1	C91-C95	C91-C95
Lymphocytic	morphology 9820-9826, 9828	204	C91.0-C91.3, C91.7-C91.9	C91.0-C91.3, C91.7-C91.9
Acute lymphocytic	morphology 9821, 9828	204	C91.0	C91.0
Chronic lymphocytic	C42.0, C42.1, C42.4 morphology 9823	204.1	C91.1	C91.1
Other lymphocytic	morphology 9820, 9822, 9824, 9825, 982	204.2-204.9	C91.2-C91.3, C91.7, C91.9	C91.2-C91.3, C91.7, C91.9
Granulocytic (myeloid)	morphology 9860-9868, 9871-9874	205	C92	C92
Acute granulocytic	morphology 9840, 9861, 9866, 9867, 9871-9874	205	C92.0, C92.4-C92.5	C92.0, C92.4-C92.5
Chronic granulocytic	morphology 9863, 9868	205.1	C92.1	C92.1
Other granulocytic	morphology 9860, 9862, 9864, 9865	205.2-205.9	C92.2-C92.3, C92.7, C92.9	C92.2-C92.3, C92.7, C92.9
Monocytic	morphology 9890-9894	206	C93	C93
Acute monocytic	morphology 9891	206	C93.0	C93.0
Chronic and other monocytic	morphology 9890, 9892-9894	206.1-206.9	C93.1-C93.9	C93.1-C93.9
Other leukemia	morphology 9800-9804, 9827, 9830, 9840-9842, 9850, 9870, 9880, 9900, 9910, 9930-9941	207, 208, 202.4, 203.1	C94, C95, C90.1, C91.4-C91.5	C94, C95, C90.1, C91.4-C91.5
Other acute	morphology 9801, 9841	207.0, 208.0	C94.0, C95.0	C94.0, C95.0
Other chronic	morphology 9803, 9842	207.1, 208.1	C94.1, C95.1	C94.1, C95.1
Aleukemic, subacute and leukemia, NOS	morphology 9800, 9802, 9804, 9827, 9830, 9850, 9870, 9880, 9900, 9910,	202.4, 203.1, 207.2, 207.8, 208.2-208.9	C90.1, C91.4-C91.5, C94.2-C94.7, C95.2-C95.9	C90.1, C91.4-C91.5, C94.2-C94.7, C95.2-C95.9

表3　われわれが解析に使用するコード表およびコードの変遷

Site of Cancer	ICD-O-2 Codes	ICD-9 Codes	ICD-10 Codes	本研究班で使用するコード
All invasive malignant cancers	C00-C80	140-208, 238.6 (except 202.4, 202.6, 202.9)	C00-C97	C00-C97
Stomach	C16	151	C16	C16
Colorectal	C18-C20, C26.0	153-154.1, 159.0	C18-C20, C26.0	C18-C20, C26.0
Liver	C22.0	155	C22	C22
Nasal cavity, middle ear and sinuses	C30-C31	160	C30-C31	C30-C31
Nasopharynx	C11	147	C11	C11
Larynx	C32	161	C32	C32
Lung and bronchus	C34	162.2-162.5, 162.8-162.9	C34	C34
bladder		188	C67	C67
Soft tissue (connective)	C49	171	C49	C49
fibrosarcoma, malignant fibrous histiocytoma				
loposarcoma, leiomyosarcoma, Rhabdomyosarcoma				
angiosarcoma, synovial sarcoma, other				
excluding malignant schwannoma, myxoid neurogenic sarcoma				
Lymphomas	morphology 9590-9717	200-202.2, 202.8-202.9	C81-C85, C96.3	C81-C85, C96.3
Hodgkin lymphoma	morphology 9650-9667	201	C81	C81
Non-Hodgkin lymphomas	morphology 9590-9595, 9670-9717, 9720, 9722, 9723, 9761,	200, 202.0-202.2, 202.8-202.9	C82-C85, C96.3	C82-C85, C88.0, C96.1, C96.3

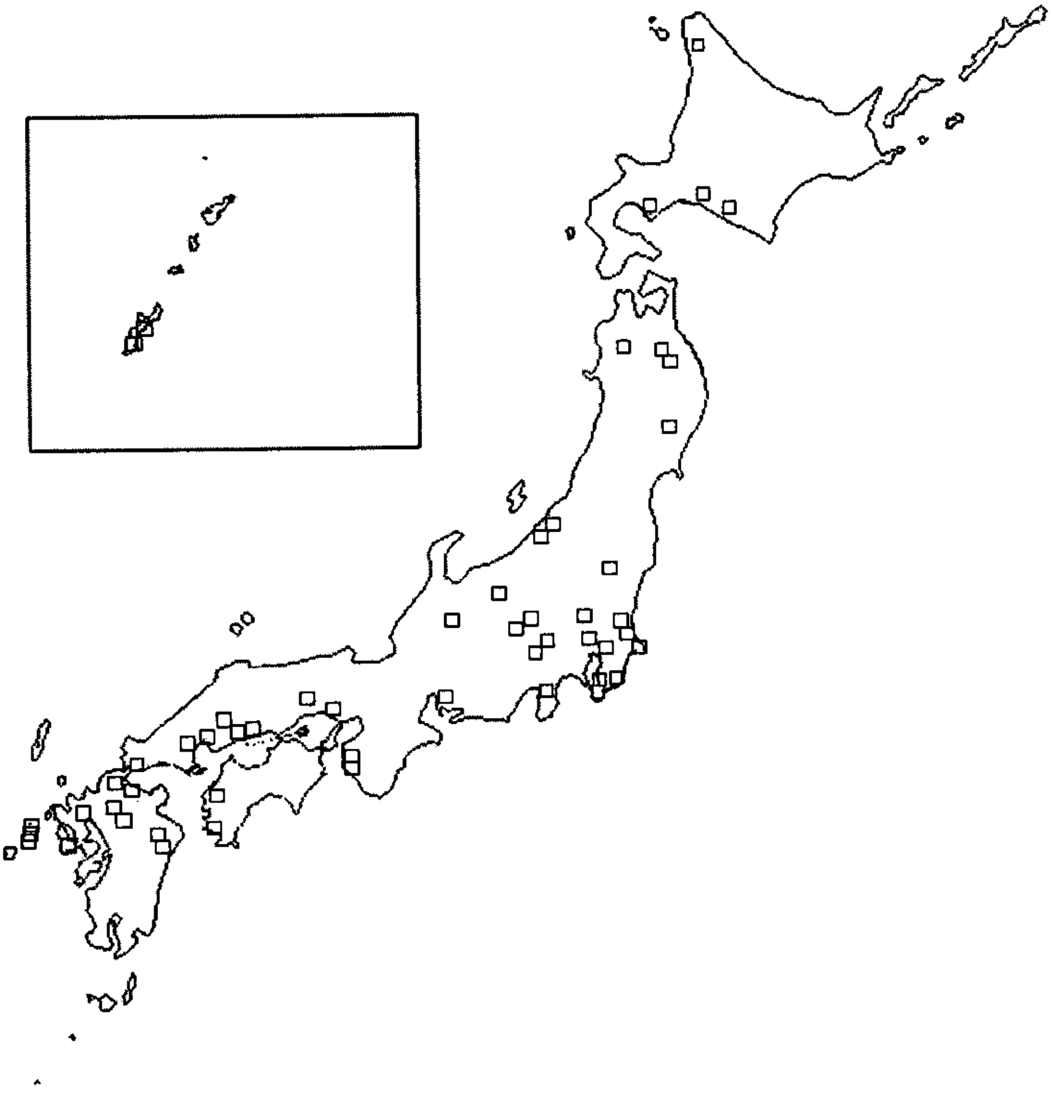
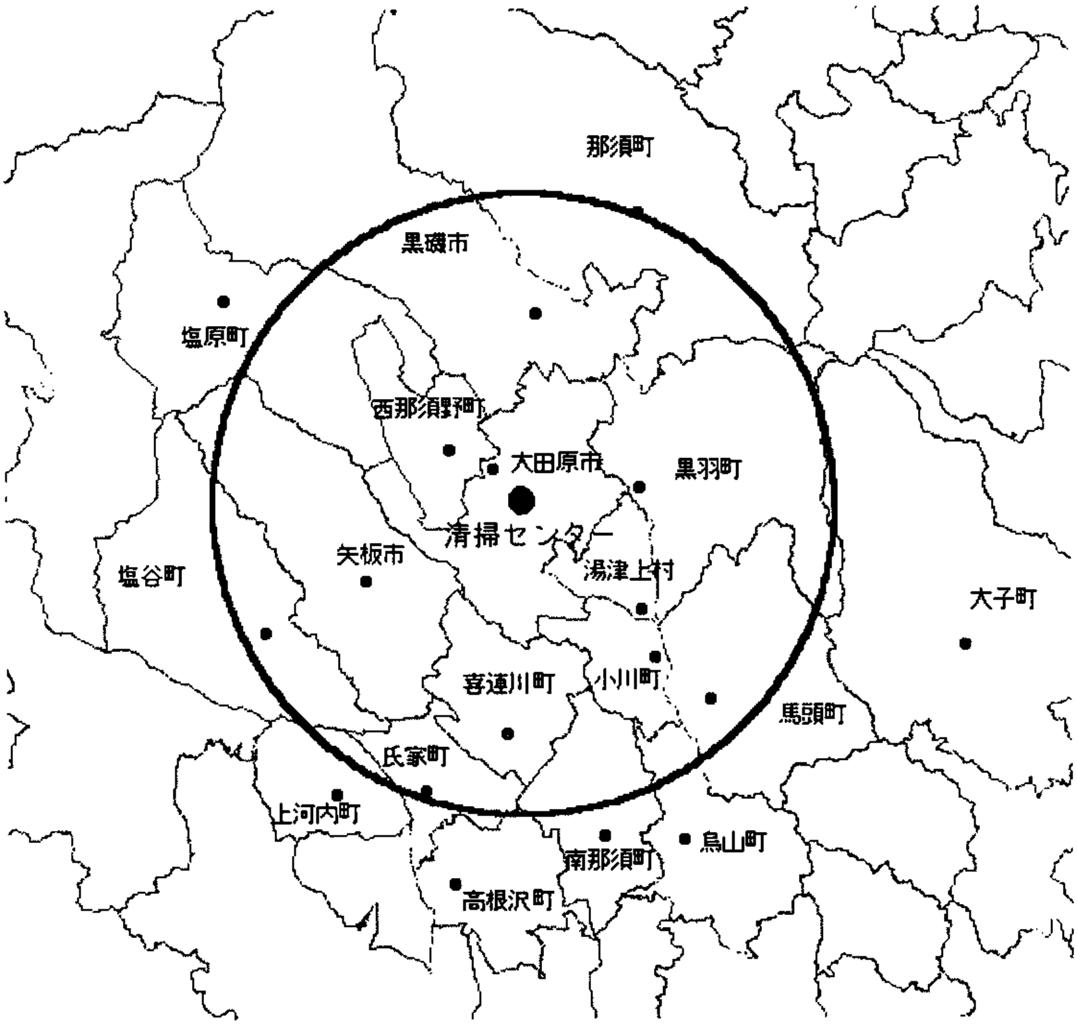


図１．全国５１ごみ焼却施設

6　栃木県大田原市　清掃センター



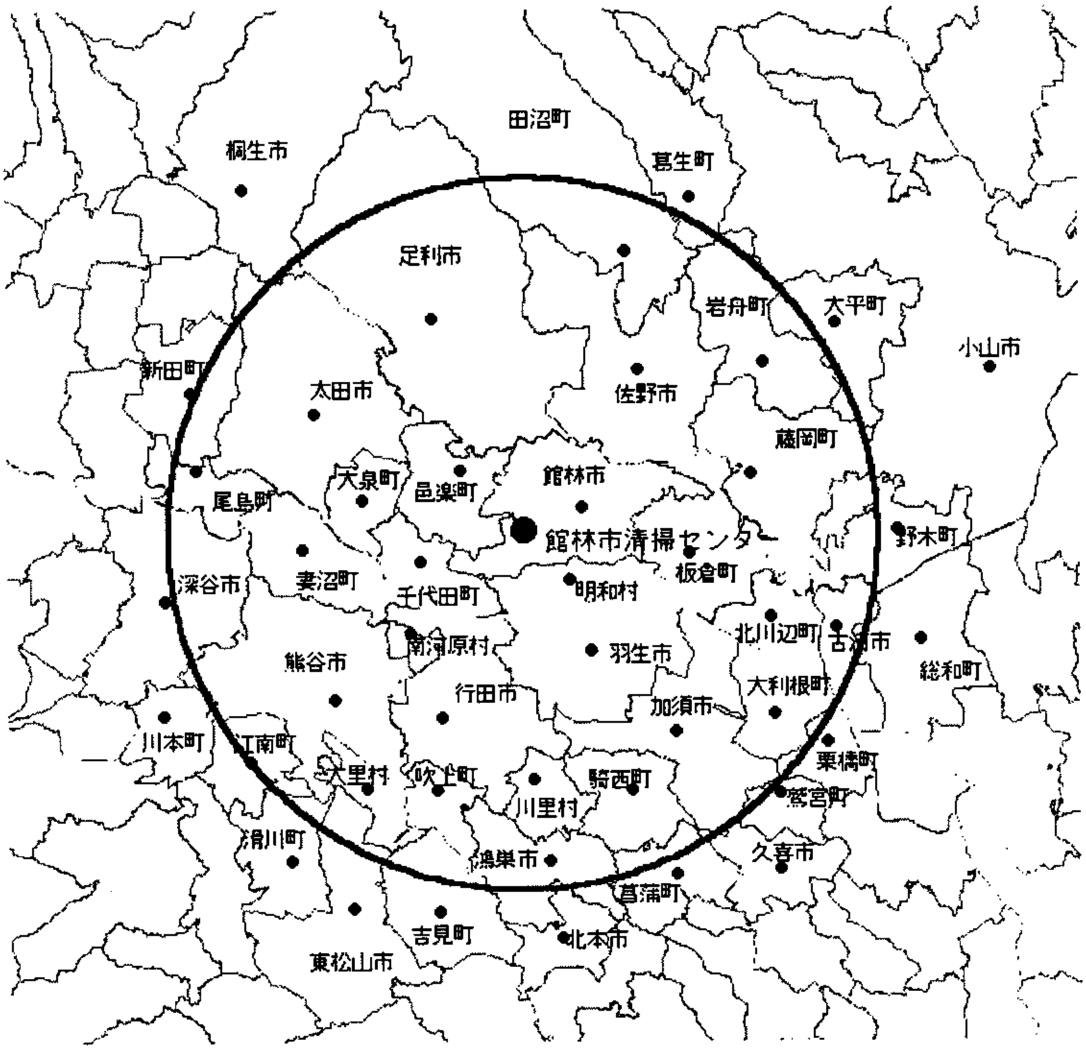
順位	施設コード	県	郡	市区町村	行政コード
1	12	栃木県	大田原市	大田原市	09-210
2	12	栃木県	那須郡	西那須野町	09-409
3	12	栃木県	那須郡	黒羽町	09-406
4	12	栃木県	那須郡	湯津上村	09-405
5	12	栃木県	矢板市	矢板市	09-211
6	12	栃木県	黒磯市	黒磯市	09-212
7	12	栃木県	那須郡	小川町	09-404
8	12	栃木県	塩谷郡	喜連川町	09-387
9	12	栃木県	那須郡	馬頭町	09-403
10	12	栃木県	塩谷郡	塩谷町	09-384
11	12	栃木県	塩谷郡	氏家町	09-385
12	12	栃木県	那須郡	那須町	09-407
13	12	栃木県	那須郡	南那須町	09-401
14	12	栃木県	河内郡	上河内町	09-303
15	12	栃木県	那須郡	塩原町	09-410
16	12	栃木県	那須郡	烏山町	09-402
17	12	栃木県	塩谷郡	高根沢町	09-386
18	12	茨城県	久慈郡	大子町	08-364

図２(a)．ごみ焼却施設とその周辺の半径20kmの同心円内とその境界に位置する市区町村－栃木県大田原市清掃センター周辺

6 栃木県大田原市 清掃センター

順位	施設コード	県	郡	市区町村	行政コード	十進経度	十進緯度	相対座標X	相対座標Y	距離
1	12	栃木県	大田原市	大田原市	09-210	140.01888889	36.86777778	1861.67	-2061.13	2777.43
2	12	栃木県	那須郡	西那須野町	09-409	139.98694444	36.87916667	4711.43	-3319.70	5763.50
3	12	栃木県	那須郡	黒羽町	09-406	140.12416667	36.85694444	-7527.10	-882.56	7578.67
4	12	栃木県	那須郡	湯津上村	09-405	140.12555556	36.78666667	-7674.77	6914.51	10330.18
5	12	栃木県	矢板市	矢板市	09-211	139.92750000	36.80361111	10001.62	5070.20	11213.35
6	12	栃木県	黒磯市	黒磯市	09-212	140.04944444	36.95861111	-839.41	-12145.12	12174.09
7	12	栃木県	那須郡	小川町	09-404	140.13416667	36.75888889	-8452.92	9994.10	13089.45
8	12	栃木県	塩谷郡	喜連川町	09-387	140.02805556	36.71583333	1010.19	14795.63	14830.07
9	12	栃木県	那須郡	馬頭町	09-403	140.17472222	36.73500000	-12082.87	12632.44	17480.68
10	12	栃木県	塩谷郡	塩谷町	09-384	139.85388889	36.77444444	16568.54	8310.20	18535.80
11	12	栃木県	塩谷郡	氏家町	09-385	139.96972222	36.68222222	6215.35	18533.75	19548.16
12	12	栃木県	那須郡	那須町	09-407	140.12416667	37.01666667	-7473.04	-18604.36	20049.16
13	12	栃木県	那須郡	南那須町	09-401	140.09722222	36.65527778	-5186.43	21499.43	22116.16
14	12	栃木県	河内郡	上河内町	09-303	139.90444444	36.68000000	12048.46	18786.62	22318.21
15	12	栃木県	那須郡	塩原町	09-410	139.82555556	36.96666667	19095.68	-13017.13	23110.40
16	12	栃木県	那須郡	烏山町	09-402	140.15500000	36.65361111	-10351.96	21668.59	24014.39
17	12	栃木県	塩谷郡	高根沢町	09-386	139.99000000	36.62777778	4393.44	24571.51	24961.20
18	12	茨城県	久慈郡	大子町	08-364	140.35861111	36.76500000	-28486.03	9229.61	29943.94

7 群馬県館林市 館林市清掃センター



順位	施設コード	県	郡	市区町村	行政コード	順位	施設コード	県	郡	市区町村	行政コード
1	13	群馬県	館林市	館林市	10-207	24	13	埼玉県	北埼玉郡	大利根町	11-425
2	13	群馬県	邑楽郡	明和村	10-522	25	13	茨城県	古河市	古河市	08-204
3	13	群馬県	邑楽郡	邑楽町	10-525	26	13	埼玉県	鴻巣市	鴻巣市	11-217
4	13	群馬県	邑楽郡	千代田町	10-523	27	13	群馬県	新田郡	尾島町	10-481
5	13	埼玉県	羽生市	羽生市	11-216	28	13	埼玉県	大里郡	江南町	11-402
6	13	埼玉県	北埼玉郡	南河原村	11-422	29	13	群馬県	新田郡	新田町	10-482
7	13	群馬県	邑楽郡	大泉町	10-524	30	13	埼玉県	深谷市	深谷市	11-218
8	13	群馬県	邑楽郡	板倉町	10-521	31	13	埼玉県	北葛飾郡	鷲宮町	11-462
9	13	栃木県	佐野市	佐野市	09-204	32	13	埼玉県	北葛飾郡	栗橋町	11-461
10	13	埼玉県	行田市	行田市	11-206	33	13	栃木県	安蘇郡	葛生町	09-422
11	13	埼玉県	大里郡	妻沼町	11-403	34	13	埼玉県	南埼玉郡	菖蒲町	11-446
12	13	栃木県	足利市	足利市	09-202	35	13	栃木県	下都賀郡	大平町	09-365
13	13	栃木県	下都賀郡	藤岡町	09-366	36	13	栃木県	下都賀郡	野木町	09-364
14	13	群馬県	太田市	太田市	10-205	37	13	埼玉県	比企郡	吉見町	11-347
15	13	埼玉県	北埼玉郡	川里村	11-423	38	13	埼玉県	比企郡	滑川町	11-341
16	13	埼玉県	加須市	加須市	11-210	39	13	埼玉県	大里郡	川本町	11-406
17	13	埼玉県	熊谷市	熊谷市	11-202	40	13	埼玉県	北本市	北本市	11-233
18	13	埼玉県	北埼玉郡	北川辺町	11-424	41	13	埼玉県	東松山市	東松山市	11-212
19	13	埼玉県	北足立郡	吹上町	11-304	42	13	茨城県	猿島郡	総和町	08-541
20	13	埼玉県	北埼玉郡	騎西町	11-421	43	13	埼玉県	久喜市	久喜市	11-232
21	13	栃木県	下都賀郡	岩舟町	09-367	44	13	群馬県	桐生市	桐生市	10-203
22	13	栃木県	安蘇郡	田沼町	09-421	45	13	栃木県	小山市	小山市	09-208
23	13	埼玉県	大里郡	大里村	11-401						

図2(b). ごみ焼却施設とその周辺の半径20kmの同心円内とその境界に位置する市区町村－群馬県館林市清掃センター周辺

7 群馬県館林市　館林市清掃センター

順位	施設コード	県	郡	市区町村	行政コード	十進経度	十進緯度	相対座標X	相対座標Y	距離
1	13	群馬県	館林市	館林市	10-207	139.54555556	36.24166667	-3274.75	-1345.62	3540.43
2	13	群馬県	邑楽郡	明和村	10-522	139.53833333	36.20527778	-2613.41	2689.44	3750.07
3	13	群馬県	邑楽郡	邑楽町	10-525	139.47083333	36.26055556	3432.59	-3463.71	4876.47
4	13	群馬県	邑楽郡	千代田町	10-523	139.44583333	36.21472222	5699.09	1612.40	5922.79
5	13	埼玉県	羽生市	羽生市	11-216	139.55194444	36.16944444	-3825.66	6668.42	7687.88
6	13	埼玉県	北埼玉郡	南河原村	11-422	139.43888889	36.17805556	6339.93	5677.71	8510.64
7	13	群馬県	邑楽郡	大泉町	10-524	139.40833333	36.24472222	9055.73	-1729.96	9219.50
8	13	群馬県	邑楽郡	板倉町	10-521	139.61361111	36.21972222	-9385.27	1104.92	9450.09
9	13	栃木県	佐野市	佐野市	09-204	139.58166667	36.31138889	-6540.10	-9071.68	11183.39
10	13	埼玉県	行田市	行田市	11-206	139.45888889	36.13555556	4559.28	10399.77	11355.27
11	13	埼玉県	大里郡	妻沼町	11-403	139.37166667	36.22055556	12363.57	936.05	12398.95
12	13	栃木県	足利市	足利市	09-202	139.45305556	36.33722222	4996.45	-11975.42	12975.94
13	13	栃木県	下都賀郡	藤岡町	09-366	139.65250000	36.25916667	-12889.29	-3263.85	13296.11
14	13	群馬県	太田市	太田市	10-205	139.37888889	36.28861111	11678.92	-6611.09	13420.28
15	13	埼玉県	北埼玉郡	川里村	11-423	139.51611111	36.10444444	-578.87	13869.53	13881.61
16	13	埼玉県	加須市	加須市	11-210	139.60527778	36.12833333	-8612.34	11241.84	14161.62
17	13	埼玉県	熊谷市	熊谷市	11-202	139.39194444	36.14416667	10579.28	9419.16	14164.81
18	13	埼玉県	北埼玉郡	北川辺町	11-424	139.66500000	36.18722222	-13998.53	4719.72	14772.77
19	13	埼玉県	北足立郡	吹上町	11-304	139.45583333	36.09833333	4850.29	14528.10	15316.37
20	13	埼玉県	北埼玉郡	騎西町	11-421	139.57805556	36.09888889	-6153.82	14502.27	15753.90
21	13	栃木県	下都賀郡	岩舟町	09-367	139.65944444	36.31555556	-13524.47	-9518.58	16538.28
22	13	栃木県	安蘇郡	田沼町	09-421	139.57305556	36.37194444	-5785.00	-15791.93	16818.18
23	13	埼玉県	大里郡	大里村	11-401	139.41222222	36.09916667	8776.31	14419.53	16880.36
24	13	埼玉県	北埼玉郡	大利根町	11-425	139.66694444	36.13750000	-14163.94	10236.17	17475.59
25	13	茨城県	古河市	古河市	08-204	139.70611111	36.18166667	-17694.86	5341.68	18483.55
26	13	埼玉県	鴻巣市	鴻巣市	11-217	139.52555556	36.06277778	-1414.49	18494.69	18548.70
27	13	群馬県	新田郡	尾島町	10-481	139.30472222	36.25916667	18358.06	-3378.27	18666.31
28	13	埼玉県	大里郡	江南町	11-402	139.33861111	36.11222222	15396.28	12939.95	20111.88
29	13	群馬県	新田郡	新田町	10-482	139.30138889	36.29861111	18633.51	-7755.98	20183.24
30	13	埼玉県	深谷市	深谷市	11-218	139.28472222	36.19444444	20195.68	3792.14	20548.62
31	13	埼玉県	北葛飾郡	鷲宮町	11-462	139.67055556	36.09722222	-14481.41	14705.07	20638.57
32	13	埼玉県	北葛飾郡	栗橋町	11-461	139.70000000	36.12305556	-17136.32	11843.20	20830.63
33	13	栃木県	安蘇郡	葛生町	09-422	139.61444444	36.39888889	-9505.15	-18772.03	21041.32
34	13	埼玉県	南埼玉郡	菖蒲町	11-446	139.60472222	36.05611111	-8543.51	19253.92	21064.30
35	13	栃木県	下都賀郡	大平町	09-365	139.70500000	36.33472222	-17617.43	-11638.59	21114.70
36	13	栃木県	下都賀郡	野木町	09-364	139.74416667	36.23000000	-21122.39	-16.60	21122.39
37	13	埼玉県	比企郡	吉見町	11-347	139.45694444	36.03666667	4776.71	21369.67	21897.03
38	13	埼玉県	比企郡	滑川町	11-341	139.36416667	36.06277778	13122.30	18436.66	22629.75
39	13	埼玉県	大里郡	川本町	11-406	139.28472222	36.13583333	20232.42	10294.50	22700.83
40	13	埼玉県	北本市	北本市	11-233	139.53333333	36.02388889	-2101.76	22811.10	22907.73
41	13	埼玉県	東松山市	東松山市	11-212	139.40333333	36.03888889	9606.14	21103.15	23186.65
42	13	茨城県	猿島郡	総和町	08-541	139.75916667	36.17583333	-22466.01	5993.78	23251.82
43	13	埼玉県	久喜市	久喜市	11-232	139.67027778	36.05888889	-14449.28	18957.64	23836.39
44	13	群馬県	桐生市	桐生市	10-203	139.33388889	36.40194444	15655.65	-19204.63	24777.35
45	13	栃木県	小山市	小山市	09-208	139.80333333	36.31138889	-26443.17	-9042.71	27946.59

固定発生源周辺における超過リスク検出のための統計モデルに関する研究
（食品・化学物質安全総合研究事業）分担研究報告書

研究者　丹後俊郎　国立保健医療科学院技術評価部長

研究要旨：ごみ焼却施設などの固定発生源周辺の問題している疾病の発生（死亡）状況の経年的推移の大きさと固定発生源からの距離をモデル化して、環境汚染により超過リスクを鋭敏に検出する方法論、つまり、空間的変動と時間的変動を同時に考慮する方法論を検討した。その際、市区町村別の死亡数がゼロとなる場合を考慮した方法を新しく検討し、かつ、施設からの距離以外の要因も考慮に入れた柔軟な統計モデルを2002年の英国王立統計学会主催の国際統計学会で発表（招待講演）した。

A. 研究目的

本研究班では欧米で示唆されているごみ焼却施設周辺における悪性新生物死亡の超過リスクに焦点を当てて、世界で始めての大規模な後ろ向きコホート調査により悪性新生物死因別死亡の経年変化を追跡することにより、ダイオキシン類の健康影響を検討することをねらいとしている。したがって、本研究では、ごみ焼却施設などの固定発生源周辺の問題している疾病の発生（死亡）状況の経年的推移が固定発生源の周辺に大きいか否かを鋭敏に検出する方法論、つまり、空間的変動と時間的変動を同時に考慮する方法論を検討する。

B. 研究方法

これまでの固定発生源周辺の環境汚染による超過リスクを検出する方法の多くは、周辺地域あるいは周辺住民に関する曝露情報がほとんどないため、1）曝露量は固定発生源からの距離に反比例する、2）ある一定期間に発生した疾病の発生頻度は距離に反比例する、という基本的仮定をおいて、疾病の空間的分布を検討している。これに対して、本研究では、同様な距離減衰の仮定をおくが、ごみ焼却施設周辺の悪性新生物死亡の経年的推移の変化とごみ焼却施設からの距離との関連性を検討して、環境汚染による超過リスクを検出する空間-時間モデル(space-time model)を新しく検討する。その際、周辺の市区町村の年単位の死亡数は死因によってはゼロの場合も想定できることから、その可能性にも対処できるモデルを検討する。

C. 研究結果

本研究班においては、ごみ焼却施設周辺の市区町村別死亡データは来年度に収集予定である。したがって、本年度の検討においては、固定発生源という点では類似している原子力発電所周辺の周産期死亡を例にして、疾病発生の経年的推移の変化の大きさと固定発生源からの距離との関連性を検討し、固定発生源周辺に超過リスクが大きいか否かを鋭敏に検出する方法論、つまり、空間的変動と時間的変動を同時に考慮する方法論を検討した。

その詳細は2002年の英国王立統計学会主催の国際統計学会で発表（招待講演）した内容を次ページに掲載したのでそれを参照されたい。

D. 考察・結論

本研究で提案した新しい統計モデルは、次年度において、simulationによる性能の検討が必要であるが、本研究班の統計解析に利用できる準備がとのったと考えている。

F. 研究発表

学会発表

Tango, T. A space-time model for excess environmental risks around putative sources based on small area data with many zero counts. (Invited paper). The 2002 International Conference of the Royal Statistical Society, 3-6 September 2002, University of Plymouth,UK, Abstracts p76.

A space-time model for excess environmental risks around putative sources based on small area data with many zero counts

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

SUMMARY

Many statistical spatial models have been proposed to detect an excess environmental risk around putative sources. However, spatial models are not always successful since it ignores a subtle but significant differences in temporal trends. In this paper, a simple and easy-to-interprete space-time model is proposed for detecting such a subtle excess risk based on small area data even when we have many zero frequencies. The proposed model is illustrated with small area data regarding an excess risk of perinatal deaths associated with maternal residential proximity to nuclear power plants in Japan.

Key words: Monte Carlo permutation; nuclear power plant, Poisson regression, relative risk.

1 INTRODUCTION

Since the 1980s there has been growing interest in the analysis of small area data to investigate the relation between the risk of a disease and proximity of residence to a prespecified putative source of hazard. It is well known that the apparent excess of cases of childhood leukaemia near the nuclear reprocessing plant such as that in the village of the Seascale at Sellafield has been extensively investigated(Bethell *et al*, 1994). More recently, there are great public concerns on the health effects of so called *dioxin*, organic compounds such as polychlorinated dibenzodioxins (PCDDs) and dibenzofurans (PCDFs) emitted from municipal solid waste incinerators(Elliott *et al*, 1996) and of radiofrequency radiation emitted from radio and television transmitters(Dolk *et al*, 1997).

Many statistical procedures, sometimes called *focused tests*, to detect such an execss risk or a cluster of cases around a putative source of hazard have been proposed. Among others, Stone's test(1988) is very popular since it is a nonparametric test based on traditional epidemiological estimates SMR or SIR (standardized mortality or incidence ratio). It has, however, been shown to be not so powerful. As a locally most powerful test, score tests have been proposed as an alternative test(Waller *et al*, 1992; Lawson 1993; Tango 1995). Bithell(1995) considered a *linear risk* score test based on the reciprocal of the rank of the distance from a point source in relation to the most powerful test against any given alternative hypothesis. Tango(2002) proposed a more flexible extension of score test that is powerful to detect a peak-decline trend in risk with distance. Diggle(1990), Diggle and

Rowlingson(1994) and Diggle *et al*(1997) have proposed point process models based on exact locations of cases.

In general, to investigate the long-term health effects of environmental pollutions, spatial modelling is not always successful since it ignores a subtle but significant differences arising in temporal trends. Low dose health effects usually first appear in the difference in the slope of time trend of incidence and/or mortality rate (Tango, 1994). However, analysis of temporal trend of rare diseases often faces difficulties due to too many zero annual incidence (or deaths) in the small area data.

In this paper, we shall propose a space-time model for detecting such a subtle excess risk around putative sources taking many zero counts into account. The proposed model is illustrated with small area data regarding an excess risk of perinatal deaths associated with maternal residential proximity to nuclear power plants in Japan.

2 Motivated example

Our work was motivated by recent public concern whether perinatal deaths cluster around the nuclear power plants in Japan. As is illustrated in Figure 1, there are 12 major sites of nuclear power plants in Japan, most of which are located at the coastal areas. The study area in our epidemiological study was defined as all the municipalities (such as "city", "ward", "town" and "village") that were located within, or overlapped with, twelve circles of radius 30km from these nuclear power plants. There are about 10-15 municipalities for each study area of individual nuclear power plant and 148 municipalities for the entire study area. The data set used here are extracted from Vital Statistics of Japan (Statistics and Information Department, 1995-1999). For each of 148 manucipalities (i=1,...,148), we have number of perinatal deaths during five years $n_{it}(t = 1, ..., 5)$, number of live births b_{it}, expected number of perinatal deaths e_{it} and the distance(km) d_i from the corresponding nuclear power plant. The annual expected number of perinatal deaths were calulated using the national perinatal death rates stratified by maternal age and type of occupation of household in the corresponding year. A part of these data are shown in Table 1.

3 METHODS

3.1 A space-time model

Consider the situation that the entire study area is divided into m small areas and include the point source. The number of cases in the ith small area and at the tth period is denoted by the random variable N_{it} with observed value n_{it}, $i = 1, 2, ..., m$, $t = 1, 2, ..., T$ and $n = \sum_{it} n_{it}$. Under the null hypothesis of both no clustering around the point source over time and no trends in time, the N_{it} are independent Poisson variables with mean e_{it} :

$$H_0 : E(N_{it}) = e_{it}, \quad (i = 1, ..., m, \ t = 1, ..., T) \tag{1}$$

where the e_{it} are the null expected numbers of cases of the ith small area and the tth period, which are calculated using the rates stratified by potential confounders such as age from some reference population or the entire study area population at the same tth period.

An alternative hypothesis of any clustering or of any temporal trend can be expressed as

$$H_1 : E(N_{it}) = \theta_{it} e_{it}, \quad (i = 1, ..., m, \ t = 1, ..., T), \tag{2}$$

where the θ_{it} denote the space-time specific relative risk. A simple and natural space-time model for detecting a risk around the point source will be

$$\begin{aligned} \log(\theta_{it}) &= h_i(t) + \log\left(1 + \epsilon \int_{a_i}^{t} g_i(u, d_i) du\right) \\ & \quad (i = 1, ..., m, \ t = 1, ..., T) \end{aligned} \tag{3}$$

where $h_i(t)$ denotes the temporal trend unrelated to exposure to the point source, a_i denotes the average starting time of exposure to the point source for the population at risk living in the ith small area, $g_i(t, d_i)$ denotes surrogates for the exposure of the ith small area with distance d_i at the tth period. A simple but generally applicable model as a first-order approximation can be

$$h_i(t) = \xi_i + \gamma t \tag{4}$$

$$g_i(t, d_i) = \tau / d_i \tag{5}$$

where the amount of exposure at the tth period is assumed to be constant and inversely related to the distance d_i from the point source. Then under this model, we have approximately

$$\begin{aligned} \log(\theta_{it}) &\approx \alpha_i + (\gamma + \beta/d_i)t. \\ & \quad (i = 1, ..., m, \ t = 1, ..., T) \end{aligned} \tag{6}$$

Then, the null hypothesis of no excess risks around the point source $H_0 : \epsilon = 0$ is equivalent to the following null hypothesis:

$$H_0 : \beta = 0 \tag{7}$$

For testing the null hypothesis, the likelihood ratio test with 1 degrees of freedom can be applied.

3.2 Procedure for many zero counts

As is shown in Table 1, many observed numbers of perinatal deaths over five years by small area are zero. Of $148 \times 5 = 740$ cells, 350 cells are zero counts. In such a situation, it is not so easy to apply the Poisson regression model (6) directly to these data set since there is a possibility that we cannot obtain any meaningful estimates due to many municipalities with many zeros and also its power is expected to be quite low. One conventional idea for avoiding this sort of difficulty will be to combine small areas having similar distances from the point source. For example, we can consider the following four zones: {A:0-3km, B:3km-10km, C:10km-20km, D:20km- }. But we can consider quite number of ways of "combining" and this sort of arbitrary combining introduces an element of subjective selections and will cause multiple testing problems. To cope with so many zero counts, we shall propose the use of quantiles of distances for combining small areas. For example, if we apply "quartile" for combining 148 municipalities described in the motivated examples, then we can construct four zones such that distances included in each zone are Zone 1=

$\{ d_{(1)}, d_{(2)}, ..., d_{(37)} \}$, Zone 2= $\{ d_{(38)}, d_{(39)}, ..., d_{(74)}\}$, Zone 3= $\{ d_{(75)}, d_{(76)}, ..., d_{(111)}\}$ and Zone 4= $\{ d_{(112)}, d_{(113)}, ..., d_{(148)}\}$, where $d_{(i)}$ denotes the distance from the point source for the municipality which is ith nearest to the point source. Let "k-tile" denotes the $100(1/k, 2/k, ..., (k-1)/k)\%$ quantiles. For example, quartile is "4-tile" and quintile is "5-tile". Then to cope with multiple testing problems due to the selection of k-tiles, we shall propose the following procedure :

Step 0: $k = 3$ and appropriately large K should be prespecified.
Step 1: Devide the study area into k zones by using k-tile and for each $j(= 1, ..., k)$ let

$$\begin{aligned} n'_{jt} &= \sum_{i \in \text{Zone } j} n_{it} && (8) \\ e'_{jt} &= \sum_{i \in \text{Zone } j} e_{it} && (9) \\ d'_j &= \sum_{i \in \text{Zone } j} d_i/(\text{number of small areas in Zone } j) && (10) \end{aligned}$$

Step 2: Apply the model (6) for the k zones:

$$\begin{aligned} \log(\theta'_{jt}) &= \alpha'_j + (\gamma' + \beta'/d'_j)t. && (11) \\ & \quad (j = 1, 2, ..., k, \ t = 1, ..., T) \end{aligned}$$

Step 3: Let $p(k)$ denotes the p-value for $H_0 : \beta' = 0$ based on the likelihood ratio test.
Step 4: If $k < K$ then $k \leftarrow k + 1$ and go to Step 1, otherwise go to Step 5.
Step 5: Proposed test statistic P_{min} is defined as

$$P_{min} = \min_{3 \le k \le K} p(k) \quad (12)$$

The null distribution of P_{min} can be obtained by Monte Carlo permutation of distances $d_i, i = 1, 2, ..., m$.

4 Illustration

To illustrate the proposed model, we shall analyze the data described in section 2. In this application, due to a small number of municipalities around each nuclear power plant and many zero counts, we considered it difficult to examine the variability of site-specific estimate of β. So, we shall here combine all the data irrespective of the site difference.

The over-all trend in mortality rates of perinatal deaths in the study area during 1995-1999 were monotonically decreasing; 7.05, 6.66, 6.37, 6.16, and 6.00 (per 1,000 live births). First of all, we shall show the relationship between the observed perinatal mortality rates combined for 5 years and the distance from the nuclear power plants (Figure 2). Several smoothed regression curves drawn in Figure 2 are estimated by smoothing splines. Apparently, there seems to be no associations (Pearson correlation coefficient is $r = 0.0023$) between five-year perinatal mortality rate and distance and we cannot observe any such trend that high mortality rates tend to cluster near the nuclear power plants.

Second, we shall apply the model (6) to the data irrespective of so many zeros. The resultant estimates($\pm$ S.E.) are $\hat{\gamma} = 0.006(\pm 0.032)$, $\hat{\beta} = 0.302(\pm 0.314)$ and the p-value of likelihood ratio test is 0.336. The deviance was 535 with 590 degrees of freedom.

Third, we shall apply the proposed procedure of combining adjacent small areas by setting that the maximum number of zones is equal to $K = 15$. For example, when the study area was divided into $k = 7$ zones, the annual estimate of relative risk (○) and the trend in the estimated relative risks over five years (solid line) based upon the model

$$\log(\theta'_{jt}) \;=\; \eta_{0j} + \eta_{1j}t, \quad (j = 1, ..., 7).$$

were shown in each of seven zones in Figure 3. Dashed line indicates the line where the relative risk is equal to 1.0. Zone 1 shows a clear increasing trend. Zone 2 also shows a little bit increasing temporal trend. Temporal trend in zone 7 is seen to be almost parallel to the horizontal line. Estimated parameters(± S.E.) of the proposed model (11) are as follows: $\hat{\gamma}' = -0.0628(\pm 0.037)$, $\hat{\beta}' = 1.245(\pm 0.425)$. Figure 4 shows the relationship between distance d_j' and the estimated slopes $\hat{\eta}_{1j}$. A decreasing curve $\hat{\gamma}' + \hat{\beta}'/d_j' = -0.0628 + 1.248/d_j'$, estimated from the model (11) was also imposed (dotted line). This result suggests that the slopes of relative risk trend are inversely related to the distance from the nuclear power plants. The deviance was 19.86 with 26 degrees of freedom. The likelihood ratio test for $H_0 : \beta' = 0$ gives us p-value = 0.00327. Individual results for each $k(= 3, ..., 15)$ was summarized in Table 2. The profile p-value of LRT for k is shown in Figure 5 and we found that $P_{min} = 0.00327$ at $k = 7$. This p-value is the ninth-largest among $999 P_{min}$'s calculated by Monte Carlo permulations of distances. Therefore, the adjusted p-value of $P_{min} = 9/(999 + 1) = 0.009$, indicating a significant increasing trend of perinatal deaths around the nuclear power plants.

The results of application to perinatal data are summarize as 1) The spatial consideration could not detect any clustering of perinatal deaths around the nuclear power plants, but 2) The proposed space-time model suggested the recent increase of perinatal deaths around the nuclear power plants in Japan.

Needless to say, the observed association of recent increase of perinatal deaths with distance from nuclear power plants cannot demonstrate causality since several limitations in the data and methods need to be considered. We assumed that a risk of exposure existed if a woman lived in proximity to nuclear power plants at the time she delivered. However, we had no information on each mother's actual exposure to nuclear power plant and her duration of exposure before delivery. Furthermore, maternal address listed on the birth, foetal and death certificates may not be an accurate measure of exposure. The residence at delivery for the mother may not be the residence of the mother during her first trimester which was considered as the period of greatest concern with respect to chemical exposure. Furthermore, no information was available regarding relative mobility of pregnant women around nuclear power plants, e.g., migration of mothers away from nuclear power plants and migration toward nuclear power plants of unexposed mothers. Recent data on mobility during pregnancy suggest that approximately 20 percent of mothers move between the time of conception and birth (Khoury M. *et al.*, 1988). Misclassification of mother's residential exposure due to these imprecise information would be nondifferential and would lead to bias toward the null.

However, we may have a different explanation that if the rates of migration of mothers with relatively high socioeconomic status away from the vicinity of nuclear power plants were recently increasing, we might have a similar association observed in our analysis. Another different story could be produced by socioeconomic status of household that we could not include in the analysis. Because of chosen or imposed circumstances, people living near nuclear power plants could be subject to social disadvantages. Especially, it is well known that the socioeconomic status of women has a predictive value for low birthweight. It seems to me, however, that recent increase of migration of mothers with relatively high socioeconomic status away from nuclear power plants are unlikely and also socioeconomic difference among people in Japan is not so large as those observed in foreign countries.

5 Discussion

In this paper, we proposed a simple space-time model to detect excess risks around putative sources in which the slope of the log-linear trend in mortality (or relative risk) is inversely proportional to the distance from the point source. Further, to cope with longitudinal data with many zero counts per small area, we proposed a simple and easy-to-interprete procedure of combining adjacent small areas, which is free from subjective amalgamation of small areas and from multiple testing problems by searching optimal "quantiles" for combining resions.

Regarding the function of surrogate for the exposure, $g_i(t, d_i)$, we considered a very simple form τ/d_i among those which is inversely related to distance d_i from a point source and constant over time. Of course, we can consider several other time-dependent non-increasing exponential functions but we face more difficult estimation problems, i.e., estimability of parameters. Bithell *et al.*(1994) and Bethell(1995) proposed the use of reciprocal of distance rank instead of the reciprocal of distance in his spatial linear risk score test since the former has the advantage that it is less dependent on population distribution.

Combining adjacent small areas assumes that there exist similar exposure pattern in the past and similar mortality trend in time among small areas to be combined. However, due to many zero counts we cannot examine this sort of homogeneity among small areas. In this sense, the proposed procedure might have an undesirable risk of eliminating the *differences* existed among small areas to be combined.

Further, these procedures still belong to the category of the simplest formulation, i.e., *distance-only analysis*. Some epidemiologists disagree on this approach on the ground that it obviously oversimplifies the spatial factors of the aetiology. Nonetheless, in the absence of detailed exposure information *in the past* around the putative souces under study and as far as the distance from the source can be considered as a primary spatial factor among others, even geographically insufficient distance-only analysis can still provide useful statistical evidence.

If the main directions of wind rose around the point source is another important factor and we could obtain a reliable data on the frequency distribution of wind directions around the point source over several years, $f(\phi_i)$ where $0 \leq \phi_i < 360^o$ denotes the angle of the ith small area measured from some standard direction, then we can consider the following model instead of (3)

$$\begin{aligned} \log(\theta_{it}) &= h_i(t) + \log\left(1 + \epsilon \int_{a_i}^{t} g_i(u, d_i, \phi_i) du\right) \\ &\approx \alpha_i + (\gamma + \beta f(\phi_i)/d_i)t. \\ & \quad (i = 1, ..., m,\ t = 1, ..., T) \end{aligned}$$

When we consider this model in the case where there are so many zero counts, then the proposed minimum search procedure in section 3.2 can be applied by using "k-tile" of $f(\phi_j)'/d_j'$, not d_j'.

References

Bithell, J.F., Dutton, S.J., Draper,G.J. and Neary, N.M.(1994) Distribution of childhood leukaemias and non-Hodgkin's lymphomas near nuclear installations in England and Wales. *British Medical Journal*,**309**, 501-505.

Bithell,J.F.(1995) The choice of test for detecting raised disease risk near a point source. *Statistics in Medicine*, **14**, 2309-2322.

Diggle, P.J.(1990) A point process modelling approach to raised incidence of a rare phenomenon in the vicinity of a prespecified point. *Journal of Royal Statistical Society, Series A*, **153**, 349-362.

Diggle, P.J. and Rowlingson, B.S.(1994) A conditional approach to point process modelling of elevated risk. *Journal of Royal Statistical Society, Series A*,**157**, 433-440.

Diggle, P.J., Morris, S., Elliott, P., Shaddick, G.(1997) Regression modelling of disease risk in relation to point sources. *Journal of Royal Statistical Society, Series A*,**160**, 491-505.

Dolk H, Shaddick G, Walls P, Grundy C, Thakrar B, Kleinschmidt I and Elliott P(1997) Cancer incidence near radio and television transmitters in Great Britain. I. Sutton Cordfield Transmitter. *Am J Epidemiol*, **145**,1-9.

Dolk H, Elliott P, Shaddick G, Walls P and Thakrar B.(1997) Cancer incidence near radio and television transmitters in Great Britain. II. All high power transmitters. *Am J Epidemiol*, **145**, 10-17.

Elliott, P., Shaddick,G., Kleinschmidt, I, Jelly,D., Walls,P., Beresford, J. and Grundy, C.(1996) Cancer incidence near municipal solid waste incinerators in Great Britain. *British Journal of Cancer*, **73**, 702-710.

Khoury M, Stewart W, Weinstein A, et al.(1988) Residential mobility during pregnancy: implications for environmental teratogenesis. *J Clin Epidemiol*,**41**, 15-20.

Lawson, A.B.(1993) On the analysis of mortality events associated with a prespecified fixed point. *J. R. Statist. Soc. A*, **156**, 363-377.

Stone, R.A.(1988) Investigation of excess environmental risks around putative sources: statistical problems and a proposed test. *Statistics in Medicine*, **7**, 649-660.

Statistics and Information Department(1995-1999), Ministry of Health and Welfare. *Vital Statistics of Japan- Volume 2.*

Tango, T.(1994) Effect of air pollution on lung cancer: a Poisson regression model based on vital statistics. *Environmental Health Perspectives*, **102(Suppl 8)**, 41-45.

Tango, T.(1995) A class of tests for detecting 'general' and 'focused' clustering of rare diseases. *Statistics in Medicine*,**14**, 2323-2334.

Tango, T.(2002) Score tests for detecting excess risks around putative sources. *Statistics in Medicine*, **21**, 497-514.

Waller, L.A., Turnbull, B.W., Clark, L.C., Nasca, P.(1992) Chronic disease surveillance and testing of clustering of disease and exposure: application to leukaemia incidence and TCE-contaminated dumpsites in upstate New York. *Environmetrics*,**3**, 281-300.

Table 1. A part of the data set described in section 2.

Area No	Observed Number of perinatal deaths					Expected Number of perinatal deaths					Distance (km)
	1995	1996	1997	1998	1999	1995	1996	1997	1998	1999	
1	1	0	0	0	0	.11	.09	.10	.07	.11	3.17
2	0	0	0	0	0	.12	.19	.17	.16	.10	27.48
3	1	0	0	0	1	.50	.48	.54	.53	.47	10.30
4	2	0	0	0	3	1.21	1.10	.89	.84	.73	6.38
5	0	0	1	2	0	.42	.35	.41	.39	.28	6.08
6	1	1	0	0	1	.57	.50	.50	.47	.47	8.28
7	0	0	0	0	2	.51	.40	.44	.49	.48	5.75
8	0	0	0	0	1	.51	.43	.41	.52	.41	3.92
9	5	1	4	2	3	6.15	5.82	5.61	5.51	5.16	14.04
10	0	0	0	0	0	.23	.19	.12	.11	.11	11.66
11	0	0	0	1	0	.52	.50	.39	.35	.31	9.57
12	0	0	1	2	0	.97	.81	.83	.71	.82	7.37
13	1	0	0	0	2	.51	.45	.42	.41	.37	13.54
14	0	1	1	0	0	.29	.30	.19	.23	.31	14.59
15	1	0	1	1	0	.76	.74	.63	.63	.52	8.59
16	0	0	0	0	0	.42	.43	.36	.34	.24	4.24
17	0	0	0	1	1	.16	.15	.13	.09	.10	6.55
18	1	3	3	2	1	2.12	2.03	1.90	1.60	1.51	11.07
19	7	3	9	6	4	4.83	4.64	4.64	4.55	4.46	10.52
20	1	1	0	2	0	.36	.26	.24	.23	.21	16.48
21	0	0	0	0	0	.15	.13	.12	.10	.12	9.17
22	0	1	0	0	1	.69	.67	.59	.54	.53	11.36
23	1	1	3	3	5	1.06	.98	.84	.92	.93	5.38
24	1	0	2	3	2	.55	.55	.57	.43	.49	7.43
25	0	0	1	3	2	2.52	2.26	1.94	2.13	1.74	9.94
26	3	4	4	2	1	2.00	2.05	2.09	2.01	1.85	33.48
27	13	8	6	11	9	6.53	6.13	5.89	6.14	5.86	12.13
28	12	11	10	14	11	12.36	12.51	11.18	11.85	10.99	10.06
29	4	3	5	1	5	2.98	2.72	2.67	2.73	2.81	3.55
30	13	11	14	13	10	14.93	14.21	13.97	13.27	12.56	15.18
31	1	1	1	2	3	1.97	1.99	1.72	1.61	1.54	10.32
32	3	6	0	1	3	2.73	2.68	2.51	2.39	2.49	10.62
33	2	1	1	1	0	.96	.84	.73	.69	.65	16.36
34	2	0	0	0	0	.51	.39	.36	.37	.39	11.79
35	1	0	0	2	1	.60	.67	.53	.49	.55	4.58
36	3	2	1	3	0	1.39	1.36	1.15	1.07	1.18	3.02
37	0	6	3	2	0	1.70	1.55	1.34	1.30	1.19	8.52
38	0	0	2	1	1	.82	.84	.78	.65	.64	4.80
39	0	0	2	0	1	.62	.51	.55	.44	.50	3.53
40	1	1	1	1	2	.83	.83	.69	.65	.68	5.91
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
146	1	0	1	0	0	.57	.49	.52	.44	.43	22.79
147	2	1	1	0	1	.70	.67	.74	.59	.61	24.55
148	0	0	0	0	0	.36	.31	.24	.25	.17	22.84

Table 2: Results of the proposed iterative procedure applied to data on perinatal deaths around nuclear power plants in Japan

k	deviance	degrees of freedom	$\hat{\gamma}'$	$\hat{\beta}'$	$p(k)^*$
3	9.11	10	-0.0453	1.030	.0575
4	4.81	14	-0.0694	1.283	.00559
5	11.16	18	-0.0567	1.168	.00939
6	16.90	22	-0.0577	1.183	.00745
7	19.86	26	-0.0628	1.245	.00327
8	20.50	30	-0.0555	1.105	.00491
9	25.57	34	-0.0441	0.993	.0143
10	34.36	38	-0.0362	0.869	.0234
11	29.28	42	-0.0331	0.831	.0263
12	39.10	46	-0.0326	0.817	.0254
13	42.49	50	-0.0276	0.743	.0341
14	57.06	54	-0.0162	0.611	.0883
15	54.29	58	-0.0092	0.512	.141

$*$ p-value based on the likelihood ratio test for $H_0 : \beta' = 0$

Legends for Figures:

Figure 1 : Location of the 12 major nuclear power plants in Japan. Names of power plant company are indicated with the cite name in the parentheses.

Figure 2 : Distance (km)from the nuclear power plants and prenatal death rates during 1995-1999 in the 148 municipalities in Japan. Several regression curves drawn are estimated by smoothing splines.

Figure 3 : Trend in the estimated relative risk (solid line) in each of seven Zones ($k = 7$) and the overall trend in the study area (dotted line).

Figure 4 : Distance (km) from the nuclear power plants and the estimated slopes of seven independent log-linear trend model (11).

Figure 5 : The profile p-value of likelihood ratio test for the hypothesis $H_0 : \beta' = 0$ for $k(= 3, ..., 15)$. The vertical line denotes the optimal k which attains the minimum of the profiel p-value. The p-value of P_{min} was calculated by Monte Carlo repliactions of 999 random permulation of distances.

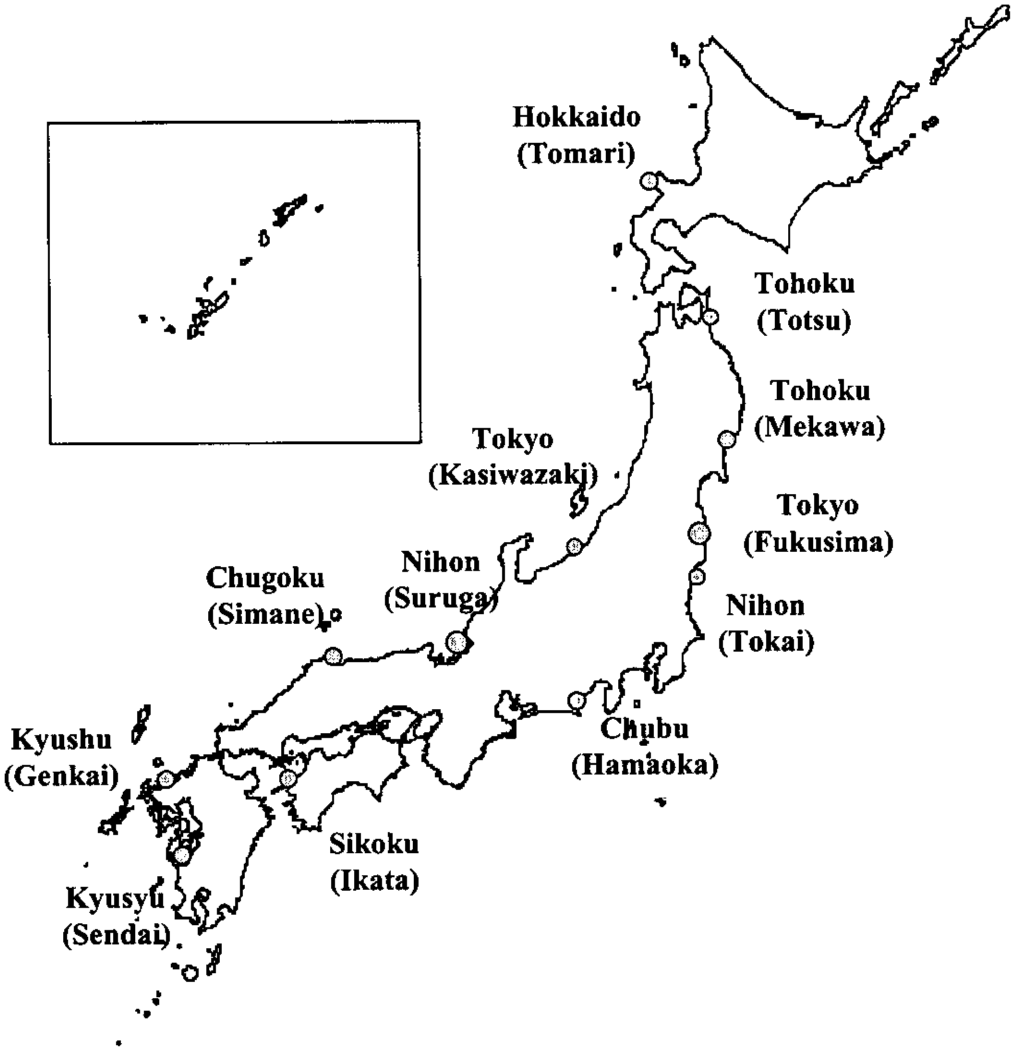

Hokkaido
(Tomari)
Tohoku
(Totsu)
Tohoku
(Mekawa)
Tokyo
(Kasiwazaki)
Tokyo
(Fukusima)
Nihon
(Tokai)
Nihon
(Suruga)
Chugoku
(Simane)
Chubu
(Hamaoka)
Kyushu
(Genkai)
Sikoku
(Ikata)
Kyusyu
(Sendai)

Figure 1