

Contents

Preface	xi
List of Contributors	xiii
1 Introduction: Spatial and syndromic surveillance for public health	1
<i>Andrew B. Lawson and Ken Kleinman</i>	
1.1 What is public health surveillance?	1
1.1.1 Spatial surveillance	1
1.1.2 Syndromic surveillance	2
1.2 The increased importance of public health surveillance	2
1.3 Geographic information, cluster detection and spatial surveillance	3
1.4 Surveillance and screening	4
1.5 Overview of process control and mapping	5
1.5.1 Process control methodology	5
1.5.2 The analysis of maps and surveillance	6
1.6 The purpose of this book	7
1.6.1 Statistical surveillance and methodological development in a public health context	7
1.6.2 The statistician's role in surveillance	7
1.7 The contents of this book	8
Part I Introduction to Temporal Surveillance	11
2 Overview of temporal surveillance	13
<i>Yann Le Strat</i>	
2.1 Introduction	13
2.1.1 Surveillance systems	13
2.1.2 Surveillance attributes	14
2.1.3 Early detection of unusual health events	15
2.2 Statistical methods	16
2.2.1 Historical limits method	16
2.2.2 Process control charts	19
2.2.3 Time-series analysis	22
2.3 Conclusion	28

3 Optimal surveillance 31

Marianne Friséen and Christian Sonesson

3.1	Introduction	31
3.2	Optimality for a fixed sample and for on-line surveillance	33
3.3	Specification of the statistical surveillance problem	34
3.4	Evaluations of systems for surveillance	35
3.4.1	Measures for a fixed sample situation adopted for surveillance	36
3.4.2	False alarms	37
3.4.3	Delay of the alarm	37
3.4.4	Predictive value	39
3.5	Optimality criteria	39
3.5.1	Minimal expected delay	39
3.5.2	Minimax optimality	40
3.5.3	Average run length	40
3.6	Optimality of some standard methods	41
3.6.1	The likelihood ratio method	41
3.6.2	The Shewhart method	43
3.6.3	The CUSUM method	44
3.6.4	Moving average and window-based methods	46
3.6.5	Exponentially weighted moving average methods	46
3.7	Special aspects of optimality for surveillance of public health	48
3.7.1	Gradual changes during outbreaks of diseases	48
3.7.2	Change between unknown incidences	49
3.7.3	Spatial and other multivariate surveillance	50
3.8	Concluding remarks	51
	Acknowledgment	52

Part II Basic Methods for Spatial and Syndromic Surveillance 53

4 Spatial and spatio-temporal disease analysis 55

Andrew B. Lawson

4.1	Introduction	55
4.2	Disease mapping and map reconstruction	56
4.3	Disease map restoration	57
4.3.1	Simple statistical representations	57
4.3.2	Basic models	62
4.3.3	A simple overdispersion model	66
4.3.4	Advanced Bayesian models	67
4.4	Residuals and goodness of fit	68
4.5	Spatio-temporal analysis	71
4.6	Surveillance issues	75

5 Generalized linear models and generalized linear mixed models for small-area surveillance 77

Ken Kleinman

5.1	Introduction	77
5.2	Surveillance using small-area modeling	78

5.2.1	Example	78
5.2.2	Using the model results	79
5.3	Alternate model formulations	80
5.3.1	Fixed effects logistic regression	80
5.3.2	Poisson regression models	81
5.4	Practical variations	82
5.5	Data	83
5.5.1	Developing and defining syndromes	84
5.6	Evaluation	85
5.6.1	Fixed and random effects monthly models	85
5.6.2	Daily versus monthly modeling	92
5.7	Conclusion	93
6 Spatial surveillance and cumulative sum methods		95
<i>Peter A. Rogerson</i>		
6.1	Introduction	95
6.2	Statistical process control	96
6.2.1	Shewhart charts	96
6.2.2	Cumulative sum charts	96
6.3	Cumulative sum methods for spatial surveillance	105
6.3.1	Maintaining a cumulative sum chart for each region	105
6.3.2	Maintaining cumulative sum charts for local neighborhoods around each region	106
6.3.3	Cumulative sum charts for global spatial statistics	110
6.3.4	Multivariate cumulative sum methods	111
6.4	Summary and discussion	112
	Acknowledgments	113
	Appendix	113
7 Scan statistics for geographical disease surveillance: an overview		115
<i>Martin Kulldorff</i>		
7.1	Introduction	115
7.1.1	Geographical disease surveillance	115
7.1.2	Tests for spatial randomness	117
7.1.3	Scan statistics	117
7.2	Scan statistics for geographical disease surveillance	119
7.2.1	Probability models	119
7.2.2	Likelihood ratio test	120
7.2.3	Scanning window	121
7.2.4	Adjustments	122
7.3	Secondary clusters	123
7.4	Null and alternative hypotheses	124
7.4.1	The null hypothesis	124
7.4.2	Spatial autocorrelation	124
7.4.3	The alternative hypothesis	125
7.5	Power	126
7.6	Visualizing the detected clusters	126
7.7	A Sample of applications	127

7.7.1	Cancer surveillance	127
7.7.2	Infectious diseases	129
7.7.3	Other human diseases	129
7.7.4	Veterinary medicine	130
7.7.5	Plant diseases	131
7.8	Software	131
	Acknowledgment	131

8 Distance-based methods for spatial and spatio-temporal surveillance **133**

Laura Forsberg, Marco Bonetti, Caroline Jeffery, Al Ozonoff and Marcello Pagano

8.1	Introduction	133
8.2	Motivation	134
8.3	Distance-based statistics for surveillance	136
8.3.1	MEET statistic	136
8.3.2	The interpoint distribution function and the M statistic	137
8.4	Spatio-temporal surveillance: an example	141
8.4.1	Temporal component	142
8.4.2	Bivariate test statistic	144
8.4.3	Power calculations	145
8.5	Locating clusters	147
8.6	Conclusion	151
	Acknowledgments	152

9 Multivariate surveillance **153**

Christian Sonesson and Marianne Frisén

9.1	Introduction	153
9.2	Specifications	154
9.3	Approaches to multivariate surveillance	155
9.3.1	Reduction of dimensionality	155
9.3.2	Reduction to one scalar statistic for each time	156
9.3.3	Parallel surveillance	157
9.3.4	Vector accumulation methods	160
9.3.5	Simultaneous solution	162
9.4	Evaluation of the properties of multivariate surveillance methods	162
9.5	Concluding discussion	164

Part III Database Mining and Bayesian Methods **167**

10 Bayesian network approaches to detection **169**

Weng-Keen Wong and Andrew W. Moore

10.1	Introduction	169
10.2	Association rules	170
10.3	WSARE	172
10.3.1	Creating the baseline distribution	172
10.3.2	Finding the best one-component rule	174

10.3.3	Two-component rules	174
10.3.4	Obtaining the p -value for each rule	176
10.4	Evaluation	177
10.4.1	The simulator	177
10.4.2	Algorithms	179
10.5	Results	181
10.6	Conclusion	186
11	Efficient scan statistic computations	189
	<i>Daniel B. Neill and Andrew W. Moore</i>	
11.1	Introduction	189
11.1.1	The spatial scan statistic	191
11.1.2	Randomization testing	191
11.1.3	The naive approach	192
11.2	Overlap-multiresolution partitioning	193
11.2.1	Score bounds	196
11.3	Results	197
11.3.1	Comparison to SaTScan	200
11.4	Conclusions and future work	201
12	Bayesian data mining for health surveillance	203
	<i>David Madigan</i>	
12.1	Introduction	203
12.2	Probabilistic graphical models	204
12.3	Hidden Markov models for surveillance: illustrative examples	206
12.4	Hidden Markov models for surveillance: further exploration	210
12.4.1	Beyond normality	210
12.4.2	How many hidden states?	212
12.4.3	Label switching	212
12.4.4	Multivariate extensions	212
12.5	Random observation time hidden Markov models	214
12.6	Interpretation of hidden Markov models for surveillance	220
12.7	Discussion	221
	Acknowledgments	221
13	Advanced modeling for surveillance: clustering of relative risk changes	223
	<i>Andrew B. Lawson</i>	
13.1	Introduction	223
13.2	Cluster concepts	223
13.3	Cluster modeling	224
13.3.1	Spatial modeling of case event data	224
13.3.2	Spatial modeling of count data	230
13.3.3	Spatio-temporal modeling of case and count data	231
13.4	Syndromic cluster assessment	235
13.4.1	The Bayesian posterior distribution	236
13.5	Bayesian version of the optimal surveillance alarm function	239

