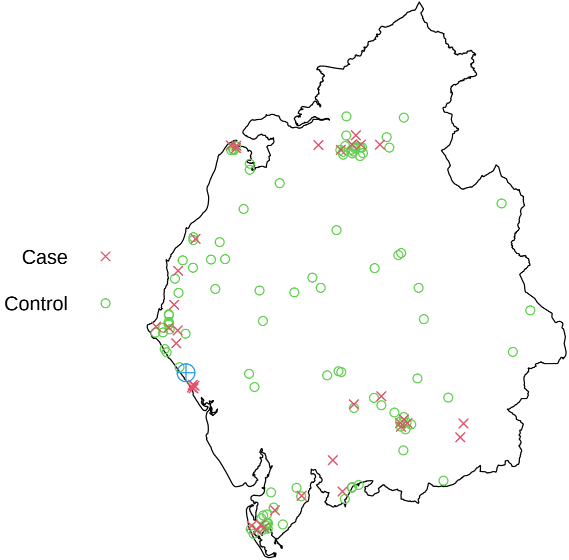
Statistical Analysis of Spatial Relative Risk

A 40-year tale of methodological development

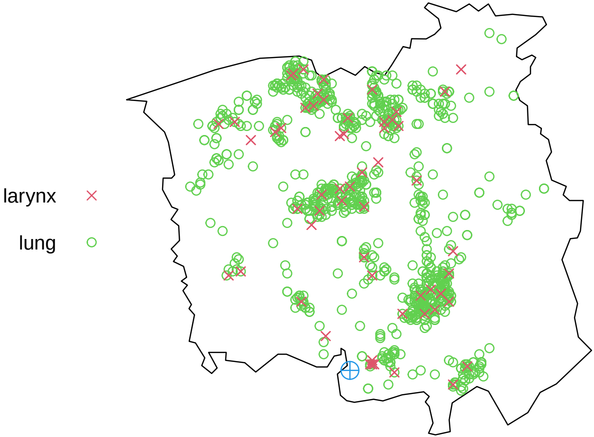
Martin Hazelton <martin.hazelton@otago.ac.nz>
Department of Mathematics & Statistics, University of Otago

EA Cornish Memorial Lecture, University of Adelaide, 17 November 2025

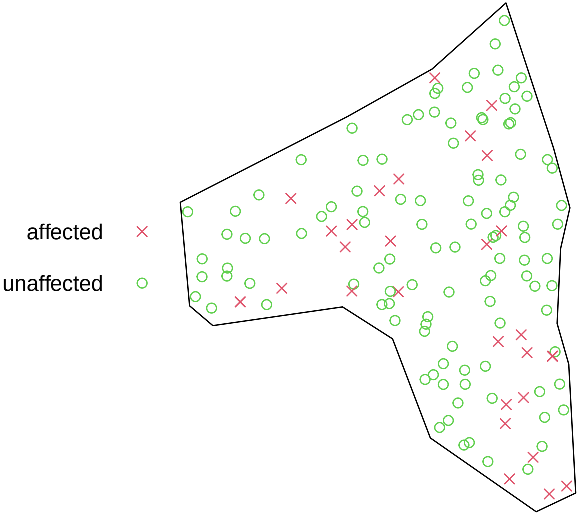
Example 1: Childhood leukemia in Cumbria, UK (1969-1982)



Example 2: Larynx cancer in Chorley-Ribble, UK (1974-1983)



Example 3: Graves in a medieval burial ground, Germany



Spatial Relative Risk

- How to explore geographical variation in risk from case-control data?
- Observations lie in compact spatial region W.
- f is density function for spatial coordinates of cases; g for controls.
- For $\boldsymbol{x} = (x_1, x_2)^{\mathsf{T}} \in W$, the **relative risk function** (Bithell, 1990) is

$$r(\boldsymbol{x}) = \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})}.$$

- Describes only relative spatial differences, not overall intensity.
- Typical to work with: $\rho(\boldsymbol{x}) = \log(r(\boldsymbol{x})) = \log(f(\boldsymbol{x})) - \log(g(\boldsymbol{x}))$.
- $r(\boldsymbol{x}) = 1 \Leftrightarrow \rho(\boldsymbol{x}) = 0$ is null; $\rho(\boldsymbol{x}) > 0$ for elevated risk at $\boldsymbol{x}$.

Bithell, J.F. (1990). *Statistics in Medicine* **9**, 691–701.

An Underlying Model

- Data: marked point pattern $\{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)\}$ on W.
  - $y = 1$ if case; $y = 0$ if control.
- Number cases $n_1 = \sum_{i=1}^{n} y_i$, number controls $n_2 = \sum_{i=1}^{n}(1 - y_i)$.
- Model: case (control) locations form a random sample given n_1 (n_2)
- Independence plausible for cancer examples.
- Conditional independence model possible for burial example.
  - Case and controls from spatial Cox processes.
- Relative risk function primarily a tool for exploratory data analysis.

Kernel Smoothing

- In practice we need to estimate case/control densities f and g.
- Use nonparameteric approach.
- Kernel density estimation:

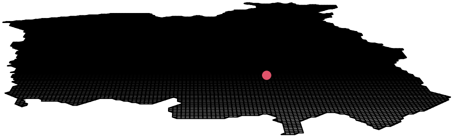
$$\hat{f}(\boldsymbol{x} \mid h) = \frac{1}{n_1} \sum_{i=1}^{n} y_i K_h(\boldsymbol{x} - \boldsymbol{x}_i)$$

$$\hat{g}(\boldsymbol{x} \mid h) = \frac{1}{n_2} \sum_{i=1}^{n} (1 - y_i) K_h(\boldsymbol{x} - \boldsymbol{x}_i).$$

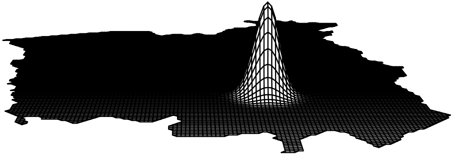
- Kernel K is isotropic density; scaled kernel $K_h(\boldsymbol{x}) = h^{-2} K(\boldsymbol{x}/h)$.
- Bandwidth h controls degree of smoothing.
- Relative risk estimate $\hat{r}(\boldsymbol{x}) = \hat{f}(\boldsymbol{x} \mid h) \Big/ \hat{g}(\boldsymbol{z} \mid h)$.

Kernel Smoothing Illustrated (Chorley larynx cases)

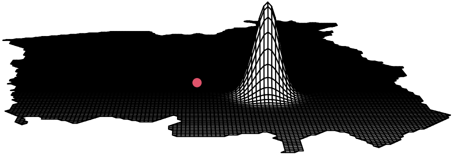
Kernel Smoothing Illustrated (Chorley larynx cases)



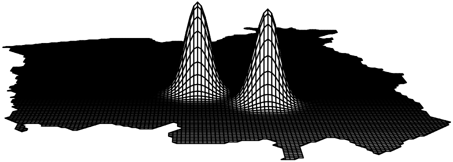
Kernel Smoothing Illustrated (Chorley larynx cases)



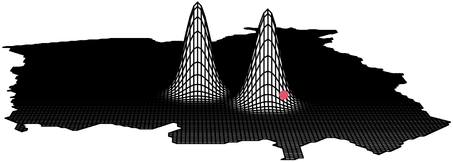
Kernel Smoothing Illustrated (Chorley larynx cases)



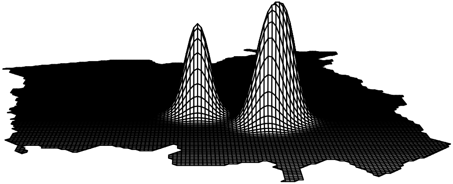
Kernel Smoothing Illustrated (Chorley larynx cases)



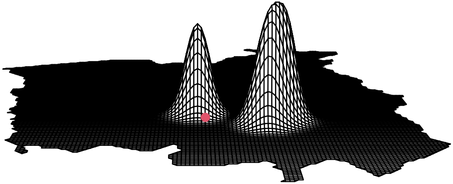
Kernel Smoothing Illustrated (Chorley larynx cases)



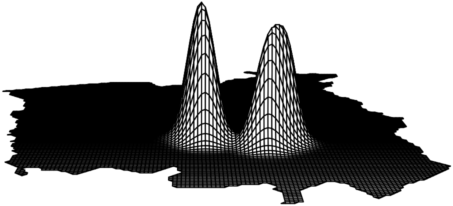
Kernel Smoothing Illustrated (Chorley larynx cases)



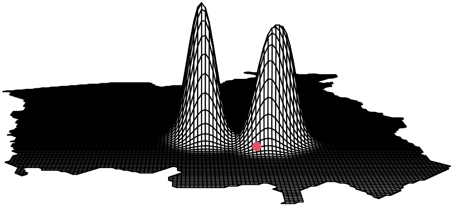
Kernel Smoothing Illustrated (Chorley larynx cases)



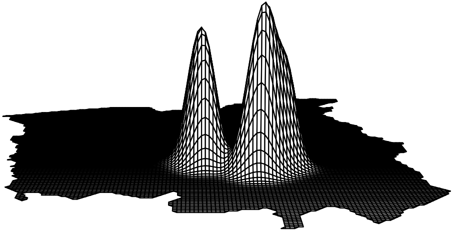
Kernel Smoothing Illustrated (Chorley larynx cases)



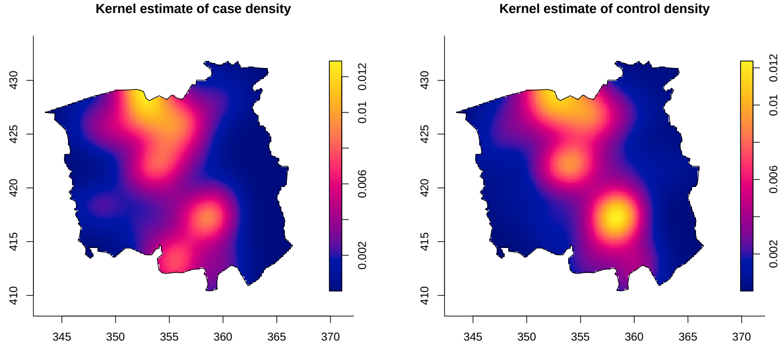
Kernel Smoothing Illustrated (Chorley larynx cases)



Kernel Smoothing Illustrated (Chorley larynx cases)



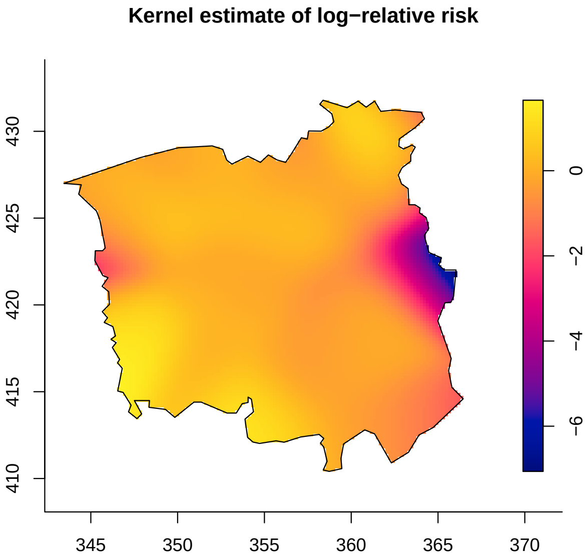
Kernel Density Estimates for Chorley



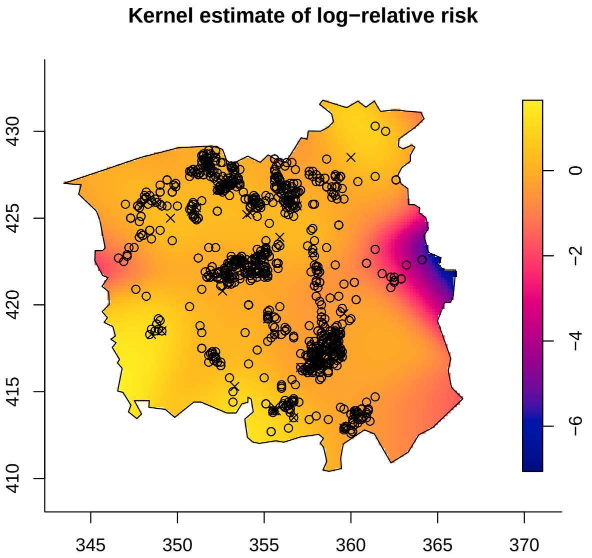
Kernel estimate of case density

Kernel estimate of control density

Log-Relative Risk Estimate for Chorley



Log-Relative Risk Estimate for Chorley



Properties of the Estimator

- Approximations to mean and variance $\hat{\rho}(\boldsymbol{x})$ available through standard asymptotics.
- Asymptotic regimen: $n \to \infty$ with $n_1 = \lfloor \psi n \rfloor$, $n_2 = \lfloor (1-\psi) n \rfloor$ for fixed $0 < \psi < 1$.
- Assume common bandwidth h used to estimate f and g, with $h \to 0$ and $nh^2 \to \infty$ and $n \to \infty$.
- Defining $k_2 = \int x_1^2 K(\boldsymbol{x})\, d\boldsymbol{x}$, $R(K) = \int K(\boldsymbol{x})^2\, d\boldsymbol{x}$,

$$\mathrm{E}[\hat{\rho}(\boldsymbol{x})] \approx \rho(\boldsymbol{x}) + \frac{h^2 k_2}{2} \left\{ \frac{\nabla^2 f(\boldsymbol{x})}{f(\boldsymbol{x})} - \frac{\nabla^2 g(\boldsymbol{x})}{g(\boldsymbol{x})} \right\}$$

$$\mathrm{Var}(\hat{\rho}(\boldsymbol{x})) \approx \frac{R(K)}{nh^2} \left\{ \frac{1}{f(\boldsymbol{x})\psi} + \frac{1}{g(\boldsymbol{x})(1-\psi)} \right\}$$

Kelsall, J.E., & Diggle, P.J. (1995). *Bernoulli* **1**, 3–16.

Choice of Bandwidth

$$\mathrm{E}[\hat{\rho}(\boldsymbol{x})] \approx \rho(\boldsymbol{x}) + \frac{h^2 k_2}{2}\left\{\frac{\nabla^2 f(\boldsymbol{x})}{f(\boldsymbol{x})} - \frac{\nabla^2 g(\boldsymbol{x})}{g(\boldsymbol{x})}\right\}$$

$$\mathrm{Var}(\hat{\rho}(\boldsymbol{x})) \approx \frac{R(K)}{nh^2}\left\{\frac{1}{f(\boldsymbol{x})\psi} + \frac{1}{g(\boldsymbol{x})(1-\psi)}\right\}$$

- Large bandwidth required to control variance in areas of sparse data.
- Bias important at risk bumps – small *h* avoids excessive blurring.
- Various methods proposed:
  - Rules of thumb – e.g. Terrell (1990)'s oversmooother, Scott (1992)'s isotropic selector;
  - Cross-validation (Kelsall & Diggle, 1995)
  - Weighted cross-validation (H, 2008)

Terrell, G.R. (1990). *Journal of the American Statistical Association*, **85**, 470–477

Scott, D.W. (1992) *Multivariate Density Estimation.*, Wiley

Kelsall, J.E., & Diggle, P.J. (1995). *Bernoulli* **1**, 3–16

Hazelton, M. L. (2008). *Statistics in Medicine*, **27**, 2269–2272

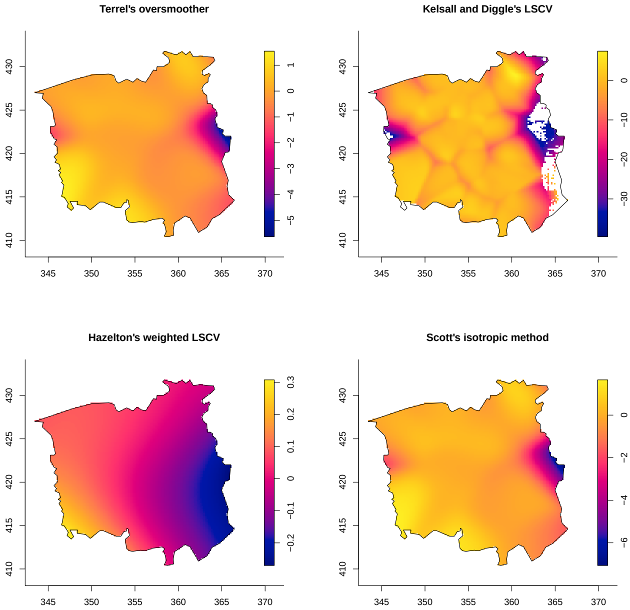
Aside: Bandwidth Choice and Edge Correction

- Weight lost over boundary from kernel functions centred at points near edge of region.
- Perennial problem in spatial statistics.
- Can correct by rescaling estimates near boundary:

$$\hat{f}(\boldsymbol{x})/q_h(\boldsymbol{x}), \quad \hat{g}(\boldsymbol{x})/q_h(\boldsymbol{x})$$

- Scaling factors $q_h(\boldsymbol{x})$ cancel when using common bandwidths for cases and controls.

Comparison of Estimates for Chorley Data



Terrel's oversmoother

Kelsall and Diggle's LSCV

Hazelton's weighted LSCV

Scott's isotropic method

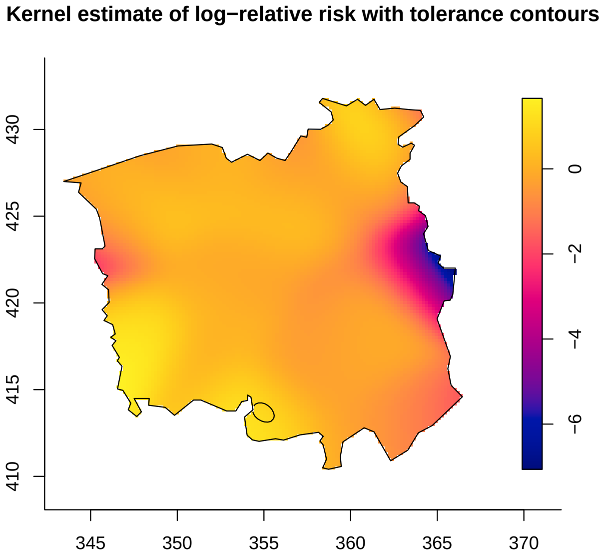
Tolerance Contours

- Aim to embellish relative risk plots to distinguish significant features from stochastic artefacts.
- Idea (e.g. Kelsall & Diggle, 1995b): for each point $\boldsymbol{x} \in W$, would we reject $H_0 \colon \rho(\boldsymbol{x}) = 0$?
- Test can either use Monte Carlo permutation (Kelsall & Diggle, 1995) or asymptotic distribution of $\hat{\rho}$ (H & Davies, 2009).
- Generates field of p-values across W.
- Add to plots contours corresponding to e.g. 5% significance level.

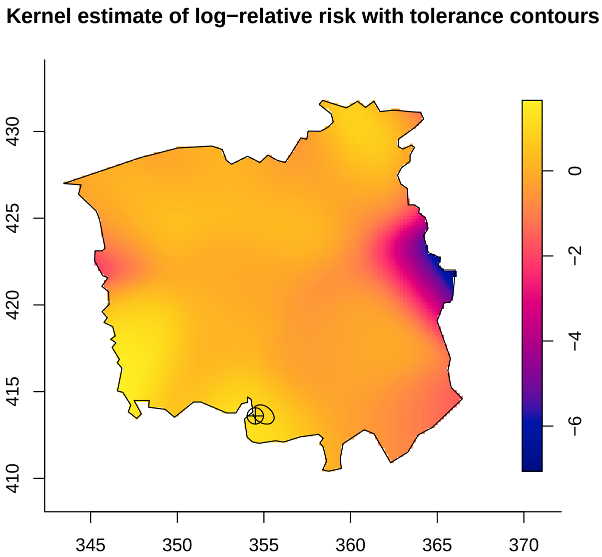
Kelsall, J.E., & Diggle, P.J. (1995b). *Statistics in Medicine* **14**, 2335–2342

Hazelton, M. L., & Davies, T.M. (2009). *Biometrical Journal*, **51**, 98–109

Tolerance Contours for the Chorley Data



Tolerance Contours for the Chorley Data



Spatially Adaptive Smoothing

- A fixed amount of smoothing across W works poorly for inhomogeneous distributions.
- Solution: allow bandwidth to vary across space.
- Focussing pro tem on case density for simplicity,

$$\hat{f}(\boldsymbol{x}) = \frac{1}{n_1} \sum_{i=1}^{n} y_i K_{h_i}(\boldsymbol{x} - \boldsymbol{x}_i).$$

- Following Abramson (1982), effective specification is

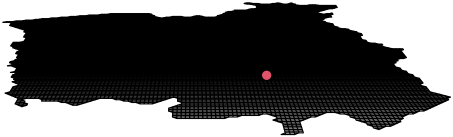
$$h_i = \frac{h_0}{\alpha(\boldsymbol{x}_i)}$$

where h_0 is global bandwidth, $\alpha(\boldsymbol{x}_i) \propto f(\boldsymbol{x}_i)^{1/2}$ is local factor.
- Implement in practice using pilot estimate of f.

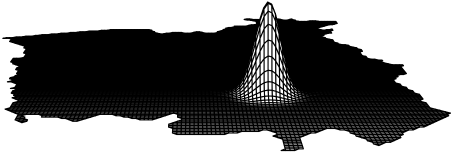
Abramson, I.S. (1982). *Annals of Statistics*, **10**, 1217–1223.

Adaptive Kernel Smoothing Illustrated (Chorley cases)

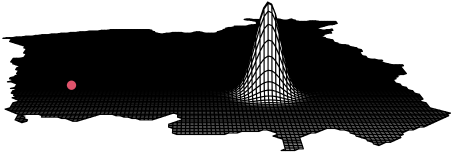
Adaptive Kernel Smoothing Illustrated (Chorley cases)



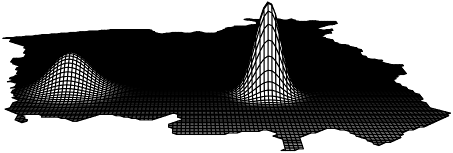
Adaptive Kernel Smoothing Illustrated (Chorley cases)



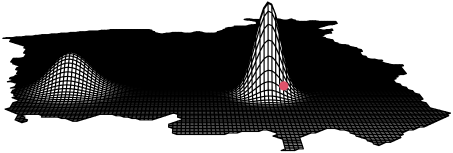
Adaptive Kernel Smoothing Illustrated (Chorley cases)



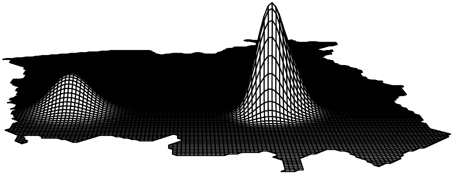
Adaptive Kernel Smoothing Illustrated (Chorley cases)



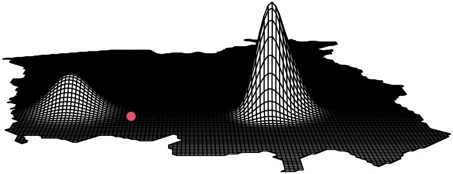
Adaptive Kernel Smoothing Illustrated (Chorley cases)



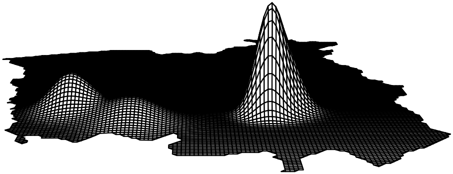
Adaptive Kernel Smoothing Illustrated (Chorley cases)



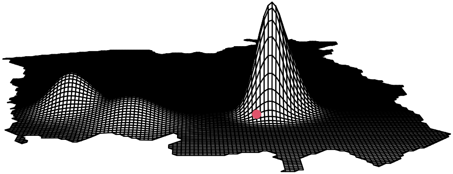
Adaptive Kernel Smoothing Illustrated (Chorley cases)



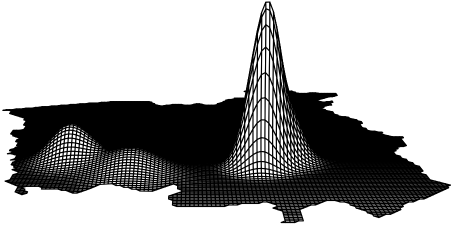
Adaptive Kernel Smoothing Illustrated (Chorley cases)



Adaptive Kernel Smoothing Illustrated (Chorley cases)



Adaptive Kernel Smoothing Illustrated (Chorley cases)



Spatially Adaptive Relative Risk Estimation

- Estimate f and g using adaptive bandwidths.
- Can use a common local bandwidth function $\alpha(\cdot)$ between cases and controls, or separate ones (Davies, Jones, H., 2016).
- Asymptotic properties:

$$\mathrm{E}[\hat{\rho}(\boldsymbol{x})] \approx \rho(\boldsymbol{x}) + h_0^4 c_b \left\{ \frac{A_f(\boldsymbol{x})}{f(\boldsymbol{x})^2} - \frac{A_g(\boldsymbol{x})}{g(\boldsymbol{x})^2} \right\}$$

$$\mathrm{Var}[\hat{\rho}(\boldsymbol{x})] \approx \frac{c_v}{n h_0^2}$$

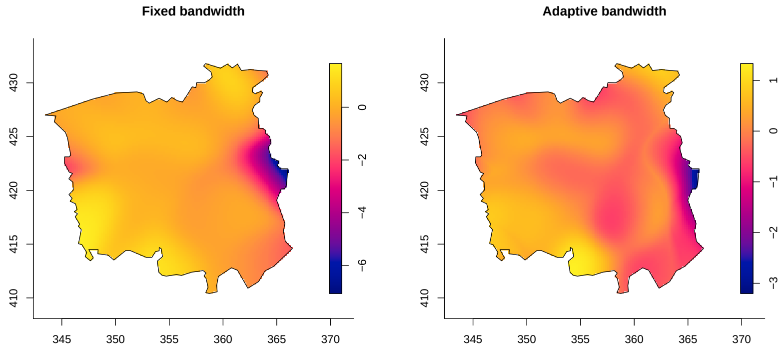
  A_f, A_g bounded functionals of f, g; c_b and c_v constants (Davies & H, 2010).

- Observe (approximate) variance stabilization.

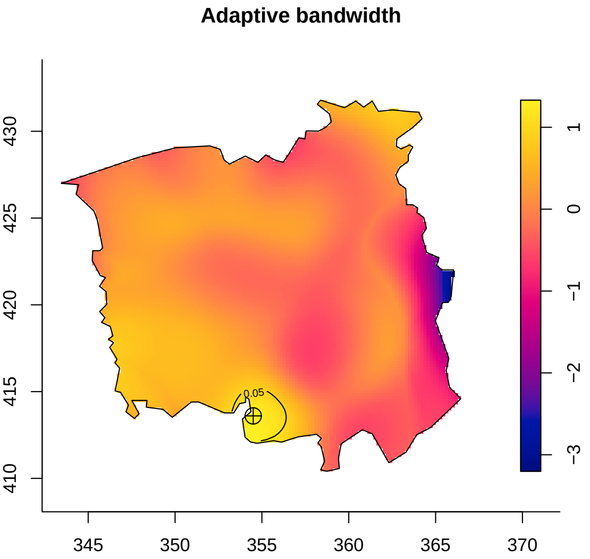
Davies, T.M. & Hazelton, M.L. (2010). *Statistics in Medicine*, **29**, 2423–2437.

Davies, T.M., Jones, K., & Hazelton, M.L. (2016). *Computational Statistics & Data Analysis*, **101**, 12–28.

Adaptive Log-Relative Risk for Chorley



Adaptive Log-Relative Risk for Chorley (tolerated)



Software

- R package `sparr`: Spatial and Spatiotemporal Relative Risk (Davies & Hazelton, 2011)
  - First released 2010
  - Maintained by Tilman Davies
  - Current version 2.3-16 contains all methods from this talk
- R package `smacpod`: Statistical Methods for the Analysis of Case-Control Point Data
  - First released 2014
  - Maintained by Joshua French (UC Denver)
- Function `relrisk` in `spatstat`
  - Fast, but less flexible

Davies, T.M. & Hazelton, M.L. (2011). *Journal of Statistical Software* **39**, 1–14.

Shrinkage Estimation

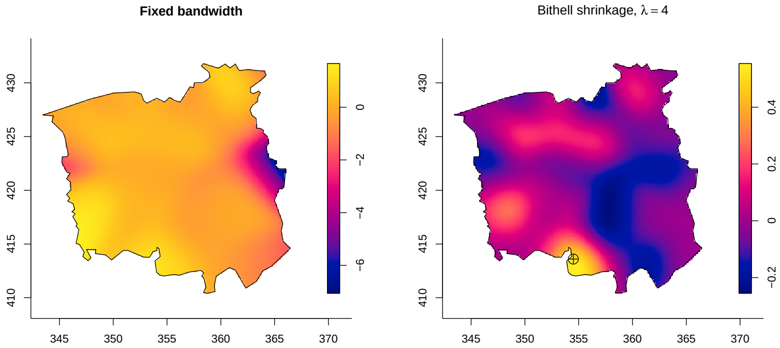
- Null value is $r(\boldsymbol{x}) = 1 \Leftrightarrow \rho(\boldsymbol{x}) = 0$.
- Idea: shrink estimate towards null value in areas of sparse data.
  - Insufficient evidence there to warrant non-null estimate.
- Idea dates back to Bithell's (1990) seminal article:

$$\hat{r}(\boldsymbol{x} \mid h, \lambda) = \frac{\lambda k_0/n_1 + \hat{f}(\boldsymbol{x} \mid h)}{\lambda k_0/n_1 + \hat{g}(\boldsymbol{x} \mid h)}$$

- $k_0 = K_h(\mathbf{0}) = K(\mathbf{0})/h^2$.
- λ an interpretable tuning parameter, controlling degree of shrinkage.
- For finite λ, shrinkage implemented as above does not impact asymptotic properties.

Bithell, J.F. (1990). *Statistics in Medicine* **9**, 691–701.

Bithell Shrinkage Log-Relative Risk for Chorley



Challenges with Bithell Shrinkage

- Theoretical justification?
- How do choose shrinkage parameter λ?
- Resulting estimate of log-relative risk still plagued by distracting artefacts.

Local Likelihood

- Consider estimation at $\boldsymbol{x} \in W$. Write $b \equiv b_{\boldsymbol{x}} = \rho(\boldsymbol{x})$.
- Local constant estimator is $\rho(\boldsymbol{z}) = b$ for $\boldsymbol{z}$ in neighbourhood of $\boldsymbol{x}$.
- $\mathrm{P}(Y = 1 \mid \boldsymbol{z}, n_1, n_2) = p = n_1 e^b / (n_2 + n_1 e^b)$.
- Local log-likelihood (Tibshirani & Hastie, 1987)

$$\begin{aligned} L(b) &= \sum_{i=1}^{n} \log\left(\mathrm{P}(Y_i = y_i \mid \boldsymbol{x}_i)\right) K_h(\boldsymbol{x} - \boldsymbol{x}_i) \\ &= \sum_{i=1}^{n} \{y_i \log(p) + (1 - y_i) \log(1 - p)\} K_h(\boldsymbol{x} - \boldsymbol{x}_i) \\ &= b n_1 \hat{f}(\boldsymbol{x}) - [n_1 \hat{f}(\boldsymbol{x}) + n_2 \hat{g}(\boldsymbol{x})] \log\left(1 + \frac{n_1}{n_2} e^b\right) + c \end{aligned}$$

- $L(b)$ maximized by standard kernel estimator $\hat{\rho}(\boldsymbol{x}) = \hat{b}$.

Tibshirani R & Hastie T. (1987) *Journal of the American Statistical Association* **82**, 559–567.

Lasso Shrinkage Estimation (H, 2023)

- Penalize negative local likelihood by L_1 (lasso) penalty (Tibshrani, 1996):
$$Q(b) = -L(b) + \lambda k_0 |b|$$
- Minimize $Q(b)$ for lasso estimator $\hat{\rho}_L(\boldsymbol{x}) = \hat{b}$.
- Rationale: lasso estimators will shrink $\hat{\rho}_L(\boldsymbol{x})$ to exactly zero for sufficiently large λ.
- Unusually for lasso estimates, $\hat{r}_L(\boldsymbol{x}) = e^{\hat{b}}$ available in closed form

$$\hat{r}_L(\boldsymbol{x}) = \begin{cases} \frac{\hat{f}(\boldsymbol{x}) - \lambda k_0/n_1}{\hat{g}(\boldsymbol{x}) + \lambda k_0/n_2} & 1 < \frac{\hat{f}(\boldsymbol{x}) - \lambda k_0/n_1}{\hat{g}(\boldsymbol{x}) + \lambda k_0/n_2} \\ \frac{\hat{f}(\boldsymbol{x}) + \lambda k_0/n_1}{\hat{g}(\boldsymbol{x}) - \lambda k_0/n_2} & 0 < \frac{\hat{f}(\boldsymbol{x}) + \lambda k_0/n_1}{\hat{g}(\boldsymbol{x}) - \lambda k_0/n_2} < 1 \\ 1 & \text{otherwise.} \end{cases}$$

Hazelton, M. L. (2023). *Statistics in Medicine*, **42**, 4556–4569.

Tibshirani, R. (1996). *Journal of the Royal Statistical Society Series B*, **58**, 267-288.

Choice of Shrinkage Parameter

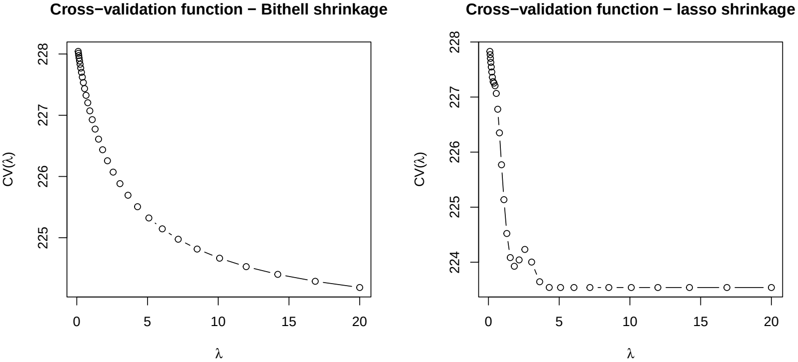
Rule of Thumb

- For lasso method, method effectively changes λ cases at $\boldsymbol{x}$ into controls.
- Suggests $\lambda = 4$.
- $\hat{\rho}$ shrunk to zero except in locations $\boldsymbol{x}$ where we would tend to reject $H_0 : \rho(\boldsymbol{x}) = 0$.

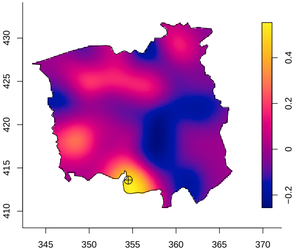
Cross-Validation

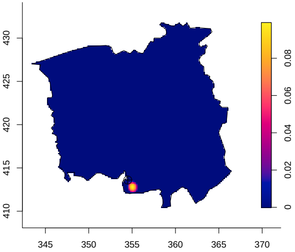
- Employ leave-one-out cross-validation based on Bernoulli log-likelihood.
- Choose left-hand local minimum in cases of multiple extrema.
- Can also be applied to Bithell shrinkage.

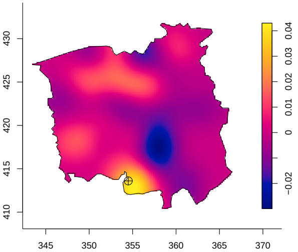
Application to Chorley Dataset: CV function

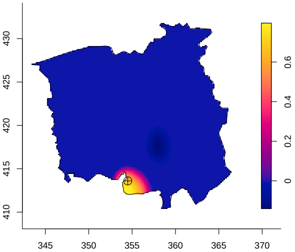


Application to Chorley Dataset: log-relative risk

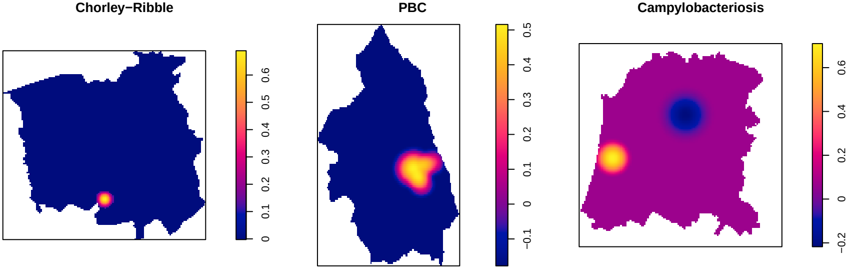

Bithell shrinkage, $\lambda = 4$


lasso shrinkage, $\lambda = 4$

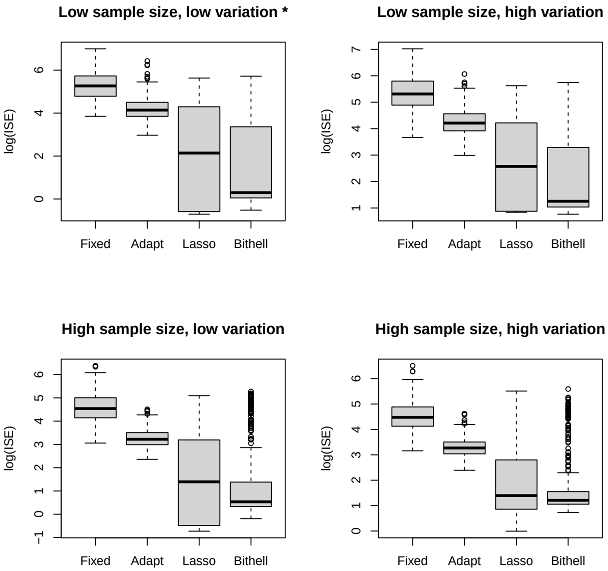

Bithell shrinkage, $\lambda = 100$


lasso shrinkage, $\lambda = 1.82$

Some Numerical Results: Test Problems



Some Numerical Results: Results for Problem 1



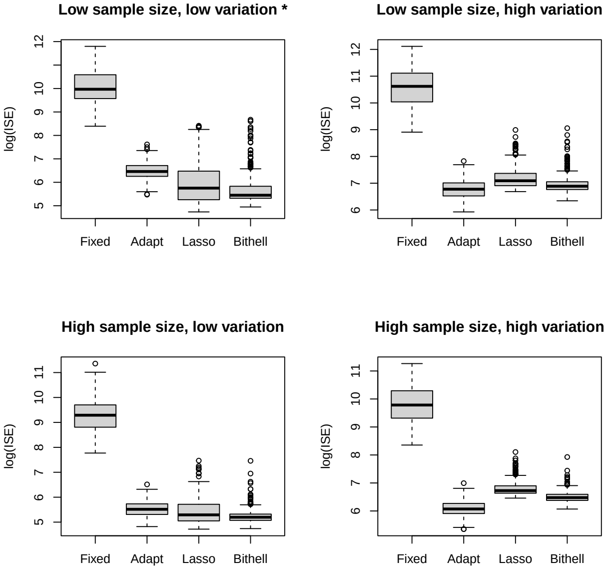
Low sample size, low variation *

Low sample size, high variation

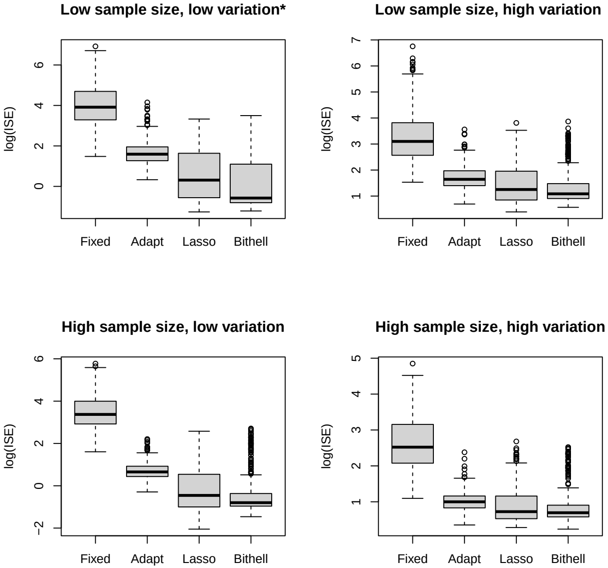
High sample size, low variation

High sample size, high variation

Some Numerical Results: Results for Problem 2



Some Numerical Results: Results for Problem 3



Low sample size, low variation*

Low sample size, high variation

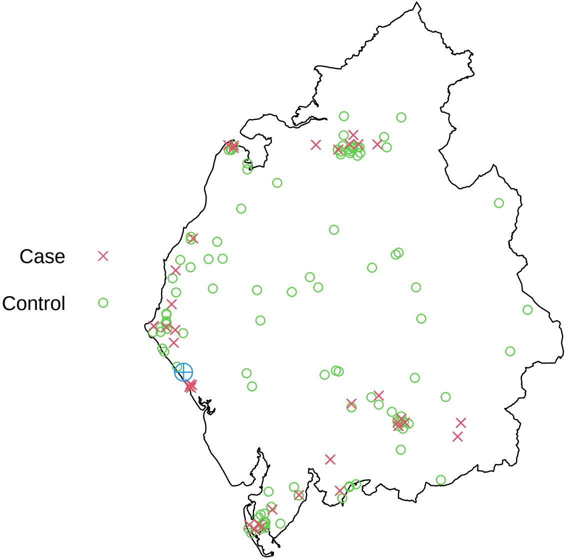
High sample size, low variation

High sample size, high variation

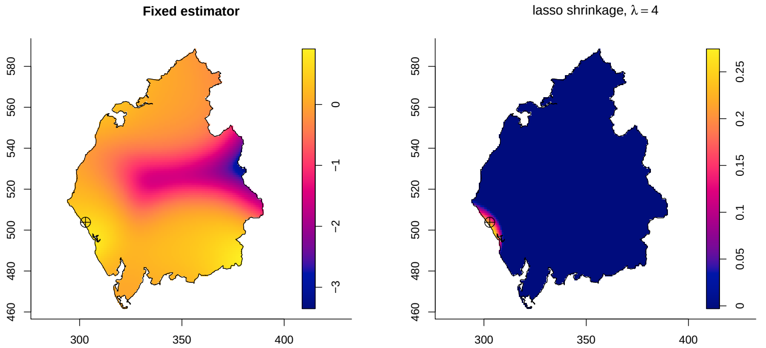
Overview of Results

- Shinkage helps.
- All methods implemented using common bandwidth.
- Shrinkage methods may benefit from smaller bandwidth, since noise controlled in areas of sparse data.
- Pilot study combining adaptive smoothing with shrinkage indicates substantional further improvement.
  - Work with Zoë Halls.

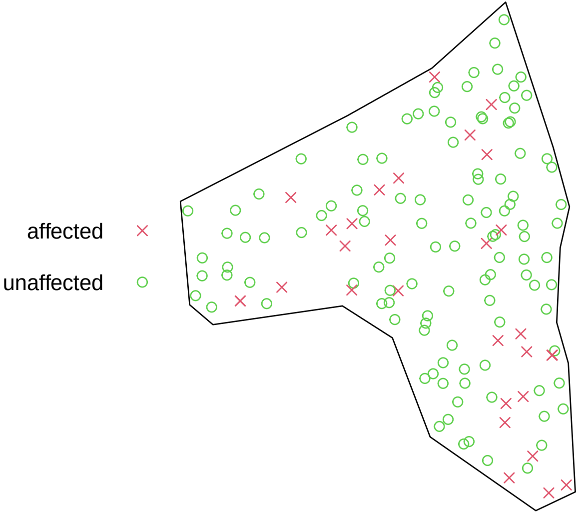
Application to Cumbria Childhood Leukemia Data



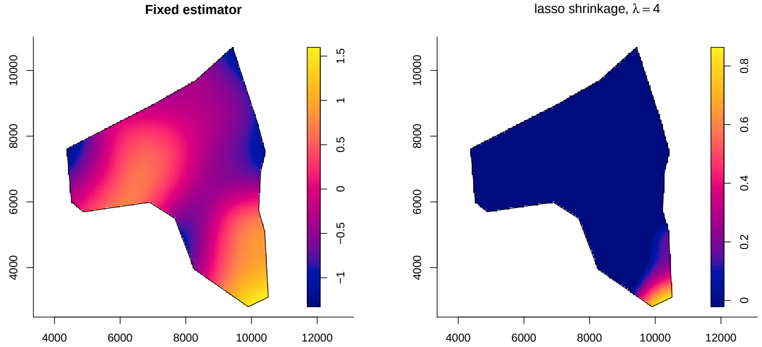
Application to Cumbria Childhood Leukemia Data



Application to Medieval Burial Ground Data



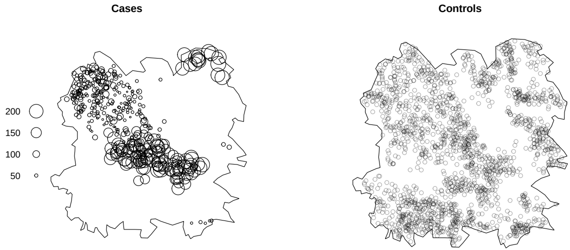
Application to Medieval Burial Ground Data

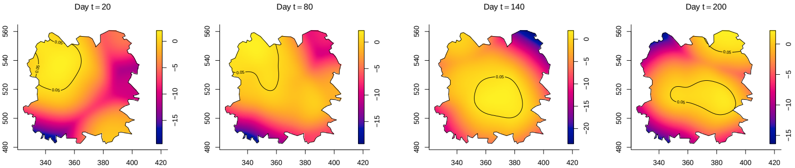

Fixed estimator

lasso shrinkage, $\lambda = 4$

Extension: Spatio-Temporal Relative Risk

(Foot and Mouth disease in Cumbria, 2001)

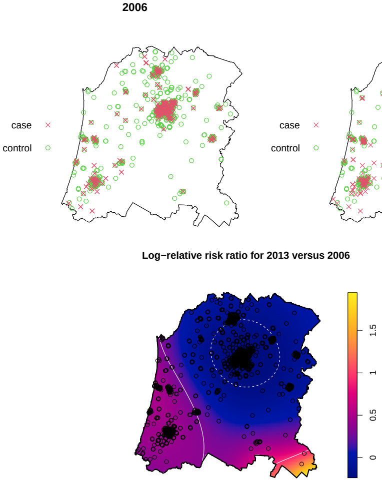




Fernando, W.T.P.S, & Hazelton, M.L. (2014). *Spatial and Spatio-Temporal Epidemiology* **8**, 1–10.

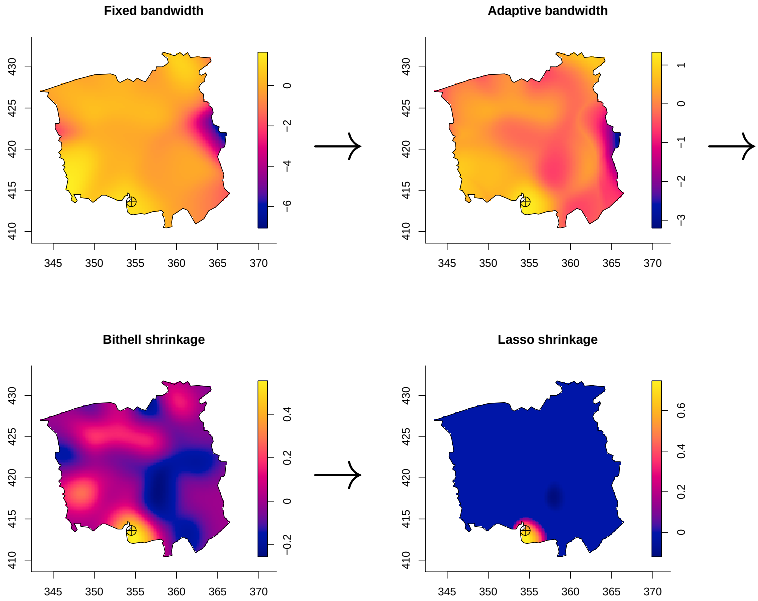
Extension: Testing for Changes in Spatial Relative Risk

(Campylobacteriosis in the Manawatū NZ, 2006 versus 2013)



Hazelton, M.L. (2017). *Statistics in Medicine* **36**, 2735–2749.

Parting Thoughts: Methodological Evolution



Parting Thoughts

- It starts with a good idea for tackling an interesting problem.
- Improvements are incremental.
- Practice drives theory.
- Leverage the research-teaching nexus.
- Release accessible software promptly.

Acknowledgements

Students Past and Present
Elspeth Mackay (Honours, University of Western Australia)
Sarojinie Fernando (PhD, Massey University)
Tilman Davies (PhD, Massey University)
Khair Jones (Honours, Massey University)
Zoë Halls (Honours, University of Otago)

Funding
Marsden Fund, Royal Society of New Zealand

In Memoriam: Dr John Francis Bithell (1939–2020)