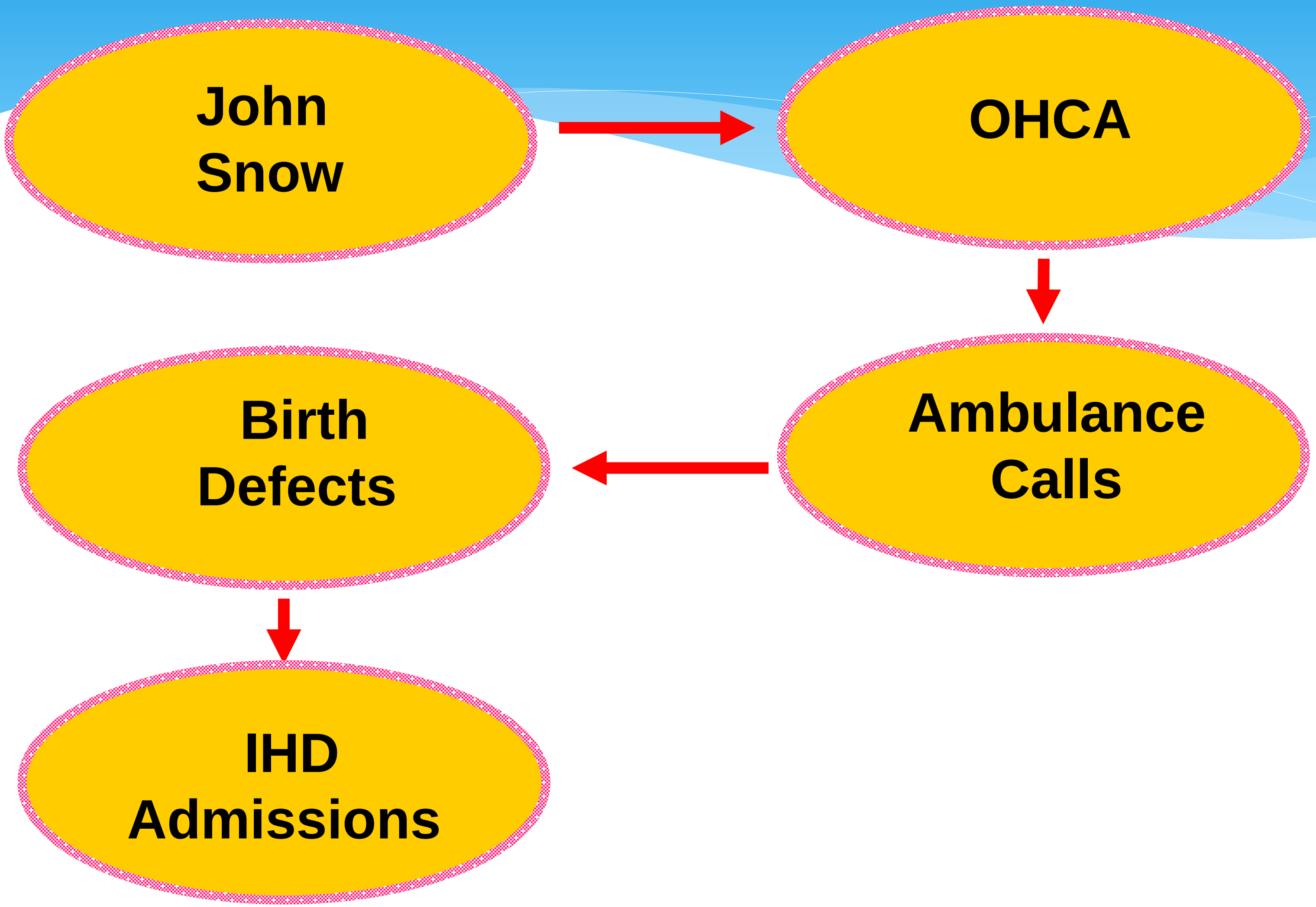
Conditional autoregressive models for geographically sparse outcomes

Dr Arul Earnest PhD,MSc, DLSHTM
Department of Statistics, Data Science and Epidemiology,
School of Health Sciences. Swinburne University of Technology

Road Map

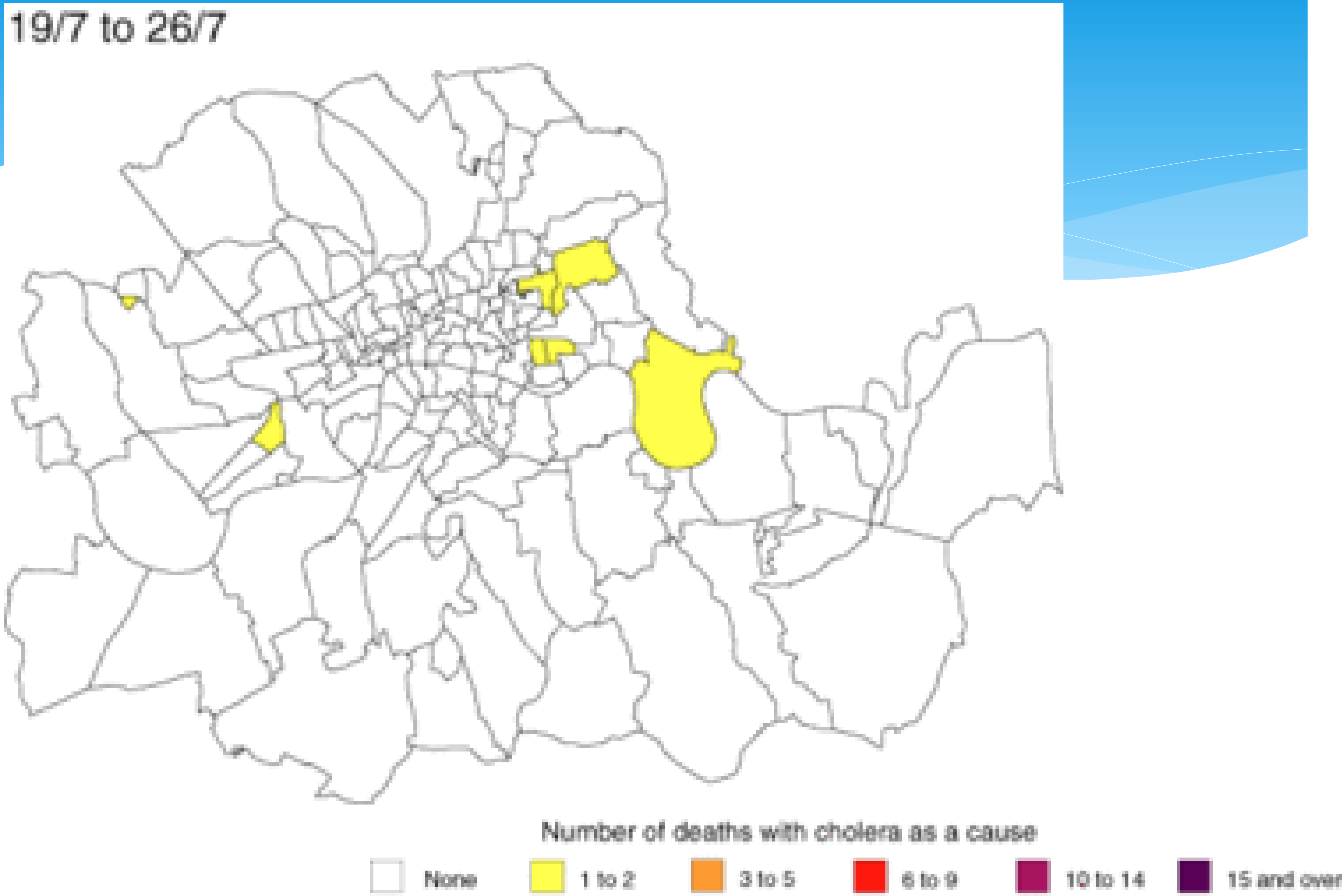

John Snow
OHCA
Ambulance Calls
Birth Defects
IHD Admissions

John Snow

Considered father of epidemiology

Traced the source of a cholera outbreak in Soho, England in 1854.

How the outbreak looked like…



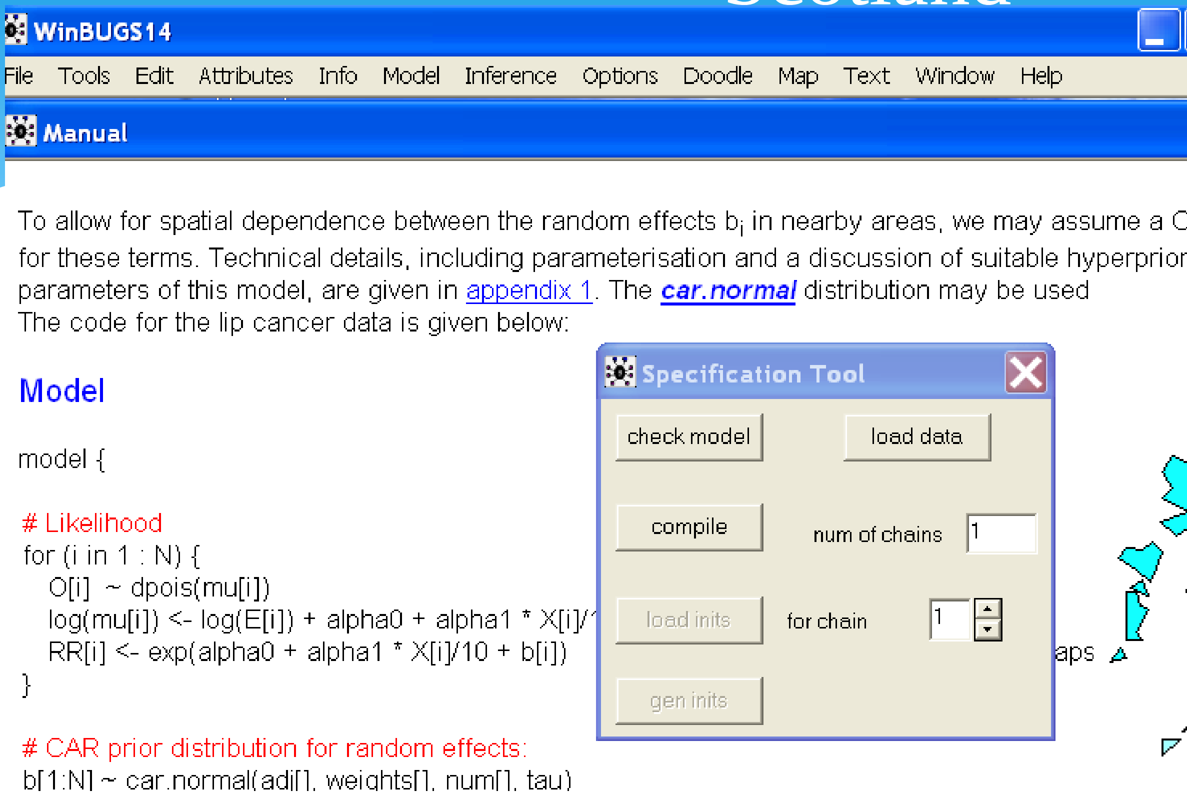
How did he do it?

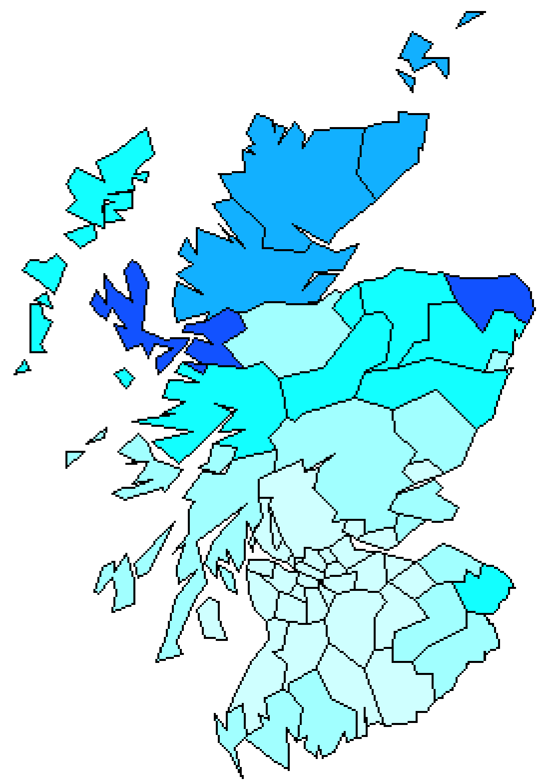
Painstakingly mapped by hand

Notice the cluster of cases along broad street

Modern day models are more sophisticated…

Creating a disease map of lip cancer rates in Scotland





Scottish Lip Cancer Maps

Case Study 1- Heart Attacks in Singapore

Case Study 1

SPATIAL VARIATION AND GEOGRAPHIC-DEMOGRAPHIC DETERMINANTS OF OUT-OF-HOSPITAL CARDIAC ARRESTS IN THE CITY-STATE OF SINGAPORE

Ong ME, **Earnest A**, Shahidah N, Ng WM, Foo C, Nott DJ.
Ann Emerg Med. 2011 Jan 14.

Impact Factor= 4.23

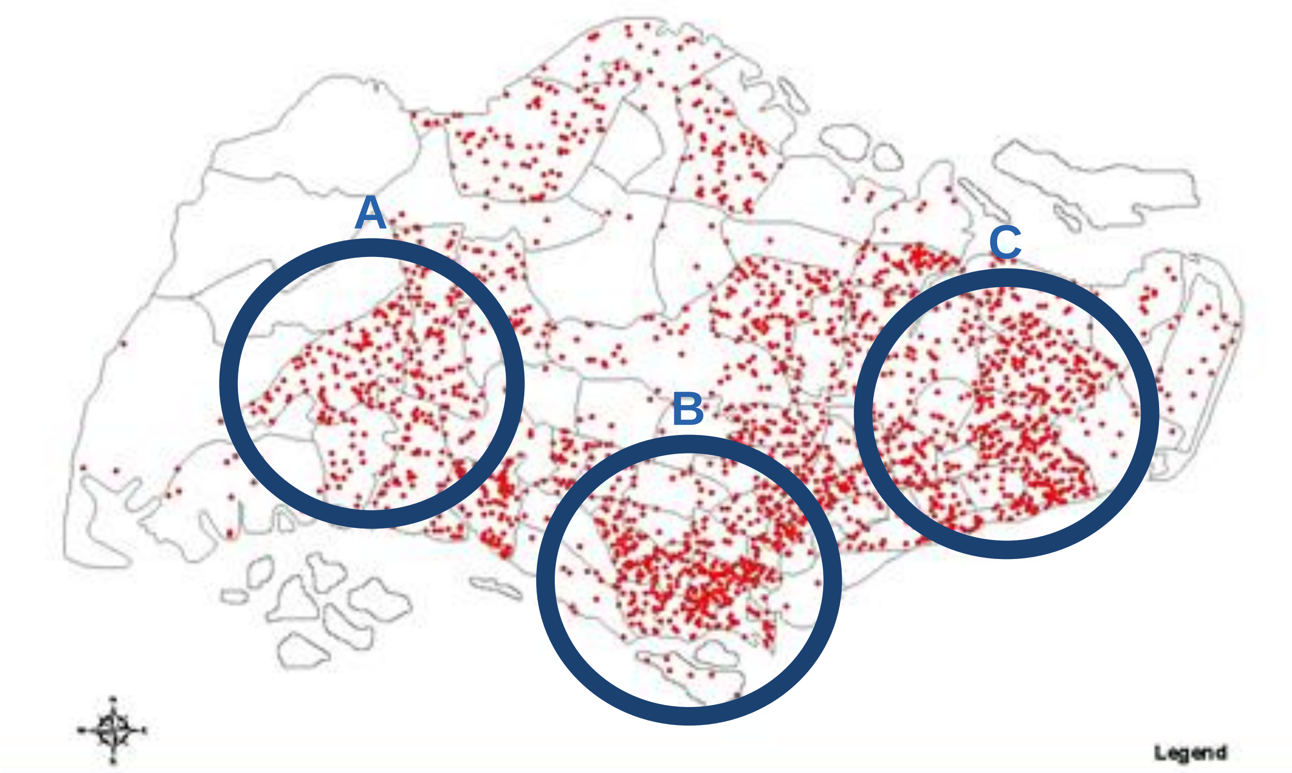
Case Study 1- OHCA

Design:

* Observational ecological study design.
* Relative risk calculated as the ratio of the observed and population standardized expected counts of pre-hospital cardiac arrests
* Conditional autoregressive (CAR) spatial models used to examine the predictors of increased risk at the DGP level.

Case Study 1- OHCA

Which part of Singapore has the highest risk of cardiac arrest? Make a guess



Here's what we really did…

Based on the address of collapse, each case was assigned to a DGP based on the Urban Redevelopment Authority Master Plan 2003.Incomplete street addresses were mapped using an online street directory and map of Singapore.

Here's what we really did…

We calculated the total observed (O_i) cardiac arrests in each DGP by summing up the cases. The expected counts of cardiac arrests were calculated as, E_i=(Pop_i/Totalpop)*Totalcardiac, where Pop_i refers to total population in the $_i$th DGP, and "Totalpop" and "Totalcardiac" refer to the overall number in the population and number of cardiac arrests during the study period.

Here's what we really did…

DGP-specific crude relative risk estimates were calculated as the ratio of the observed and expected counts for each area. The CAR model was then used to smooth these crude relative risks.

And then, we ….

used a Poisson model that incorporated both a spatially structured random effect term as well as a spatially unstructured random term. This model is commonly known as a CAR convolution prior. The expected counts of cardiac arrest were included in the model, as were other covariates.

WinBUGs code

```
# Poisson likelihood for observed counts
    O[i]~dpois(mu[i])
    log(mu[i])<- log(E[i])+bspat[i]+bind[i]

    RR[i]<-exp(bspat[i]+bind[i])

# CAR prior distribution for spatial correlated random effects:
bspat[1:N]~car.normal(adj[],weights[],num[],vspat)

# Normal prior distribution for uncorrelated random effects
for(i in 1:N)
{
bind[i]~dnorm(alpha,vind)
```

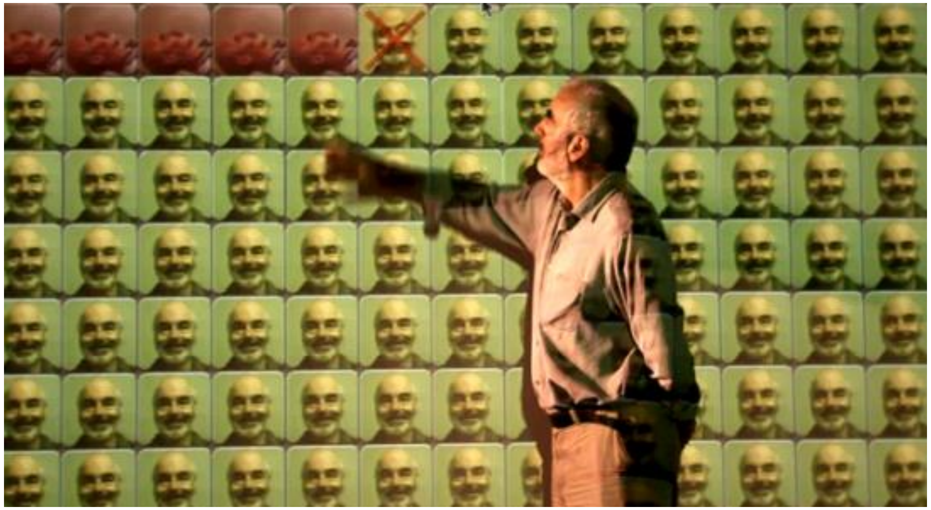
We plug the data and model into a sofware…

WinBUGs is a free software available over the following website:

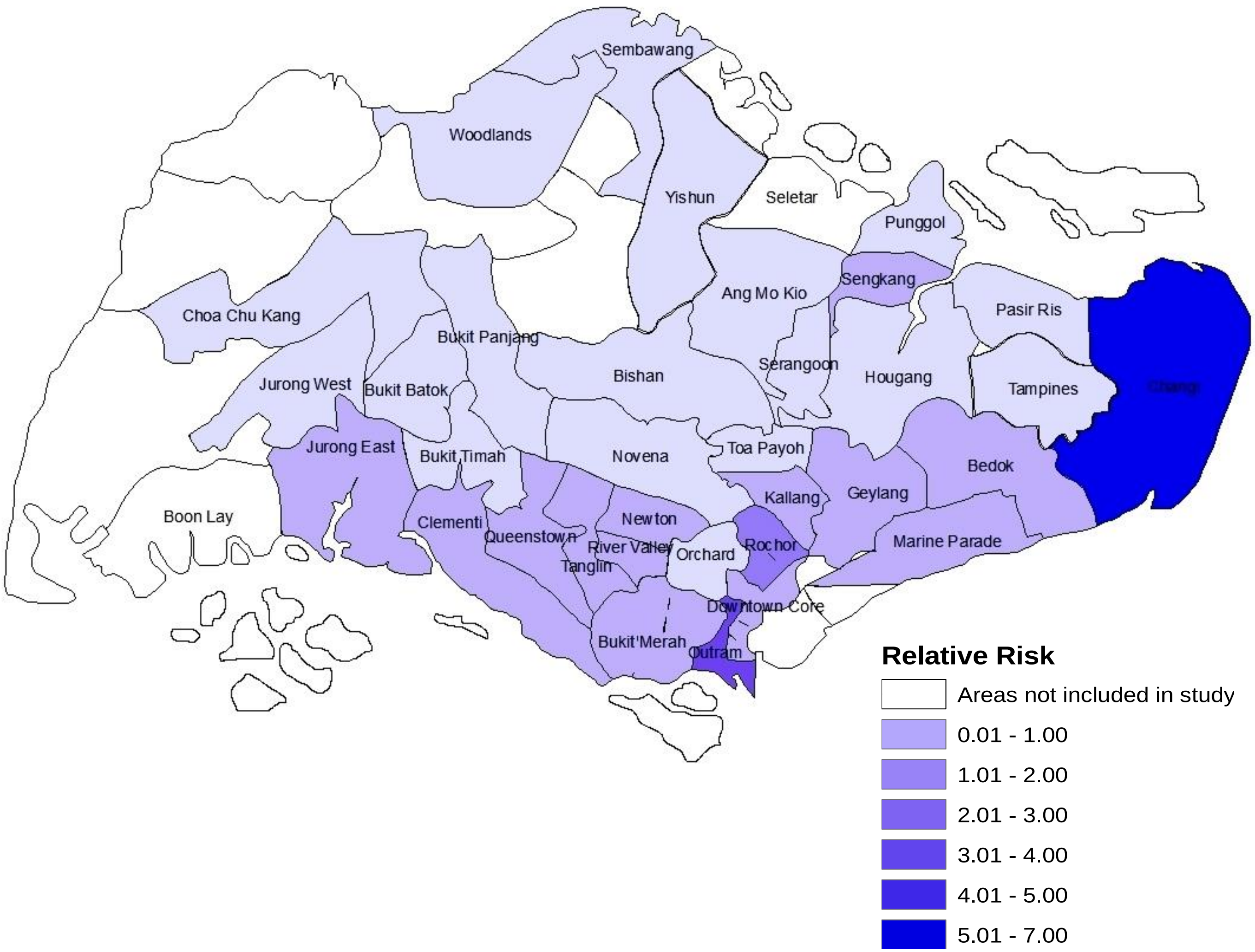
http://www.mrc-bsu.cam.ac.uk/bugs/

Bayesian inference **U**sing **G**ibbs **S**ampling

Developed by a few people from Imperial College and MRC Biostats unit, including this guy..



The results we got…



CAR Regression Analysis

Table 3. Bivariate DGP factors associated with risk of out-of-hospital cardiac arrest.

	All Cardiac Arrests		
Covariates	**RR**	**95% CI**	**Deviance information criterion**
Aged ≥65 y	1.19	1.05–1.35	261.08
Men	1.07	0.91–1.28	260.86
Chinese	0.96	0.93–0.98	260.31
High education (poly and above)	1.12	0.99–1.27	261.35
In 5 rooms and above	1.00	0.99–1.01	261.35
With high personal income (≥$5,000)	0.99	0.97–1.02	261.38
Working	0.93	0.78–1.10	261.21
Senior officials and professionals	0.99	0.98–1.02	261.25
No family nucleus	1.08	1.03–1.13	261.20
With household size ≥5	0.97	0.92–1.02	261.29

Conclusion

1. Spatial variation in OHCA in Singapore
2. This spatial variation in risk not explained by areal-measures of socio-economic status
3. Rather, it is driven by racial, family and age structure of the area we live in
4. Results are being used for health services planning. i.e. these findings will help policy-makers in terms of planning health education programs, ambulance deployment and identifying locations for AEDs.

The Alfred
Caulfield Hospital
Sandringham Hospital

AlfredHealth

Keyword Search Go

For Patients & Visitors | Departments & Services | Research | Careers | About Us | For Staff | For GPs

Alfred Health News

Follow us:

New emergency department to deliver more timely, and child-friendly, emergency care for Bayside

24/09/2014

Modern and spacious treatment areas, and a dedicated zone for children, are among the many improvements achieved through the redevelopment of Sandringham Hospital's emergency department.

Officially opened today by the Minister for Health, The Hon. David Davis, the $6.8 million emergency department expansion also introduces six short stay beds for people needing extended medical observation and establishes an 'urgent care centre' for those who can benefit from GP-like care.

Alfred Health Chief Executive Officer, A/Prof Andrew Way said the increased size of the new department, together with the introduction of an urgent care centre operated by GP-specialists, brings a new level of access to emergency care for the Bayside community.

"Establishing an urgent care centre - located next to our busy emergency department - is an innovative model," A/Prof Way said. "It's about providing timely care for the Bayside community, and supports greater access to support for the 30,000-plus patients who visit our emergency department each year.

News Archive

Case Study 2

Geographical variation in ambulance calls in Singapore is explained by socio-economic status

Earnest A, Tan SB, Shahidah N, Ong ME. Acad Emerg Med. 2012 Feb;19(2):180-8.

Objectives

Primary Aim: map the spatial distribution of ambulance calls, specifically medical and trauma related calls, in Singapore, at the Development Guide Plan (DGP) level, using residential addresses of callers.

Secondary aim: studying the relationship between the risk of medical and trauma related calls with socio-demographic variables measured at the areal level

Methods

Data from 2 sources:

1) Cardiac Arrest and Resuscitation Epidemiology (CARE) study database
2) Singapore Census 2000

Based on their residential address from where the call was made, each case was assigned to a DGP based on the Urban Redevelopment Authority (URA) Master Plan 2003

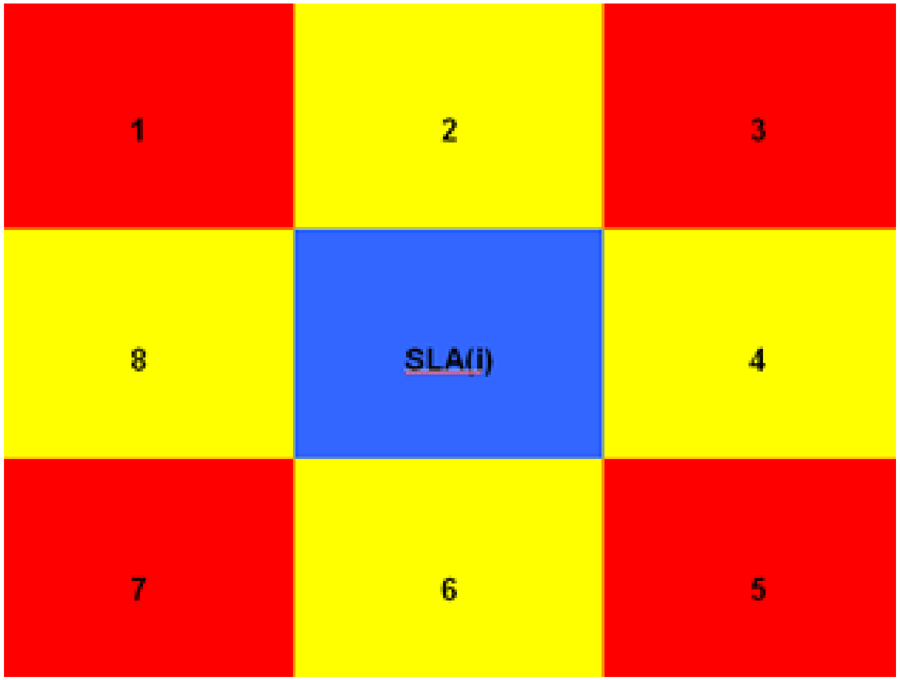
Methods

Some more info on the Bayesian model:

1) Neighbourhood structure based on Queen method of assignment

Figure 1 - Neighbourhood assignment based on adjacency

Note: For Rook method, only neighbours 2,4,6 and 8 assigned to SLA(i)
For Queen method, all neighbours (i.e. 1-8) are assigned to SLA(i)

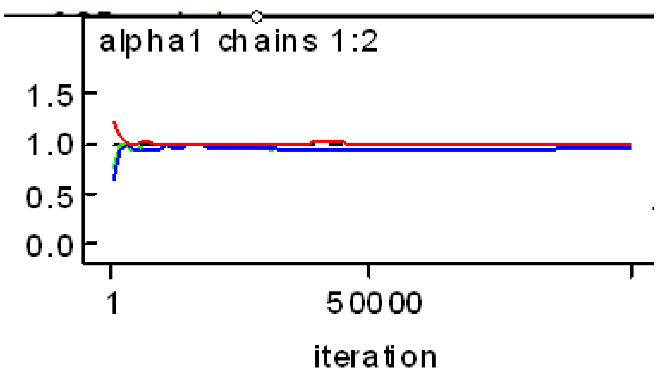


Earnest A, Morgan G, Mengersen K, Ryan L, Summerhayes R, Beard J. Evaluating the effect of neighbourhood weight matrices on smoothing properties of Conditional Autoregressive (CAR) models.
Int J Health Geogr. 2007 Nov 29;6:54.

Methods

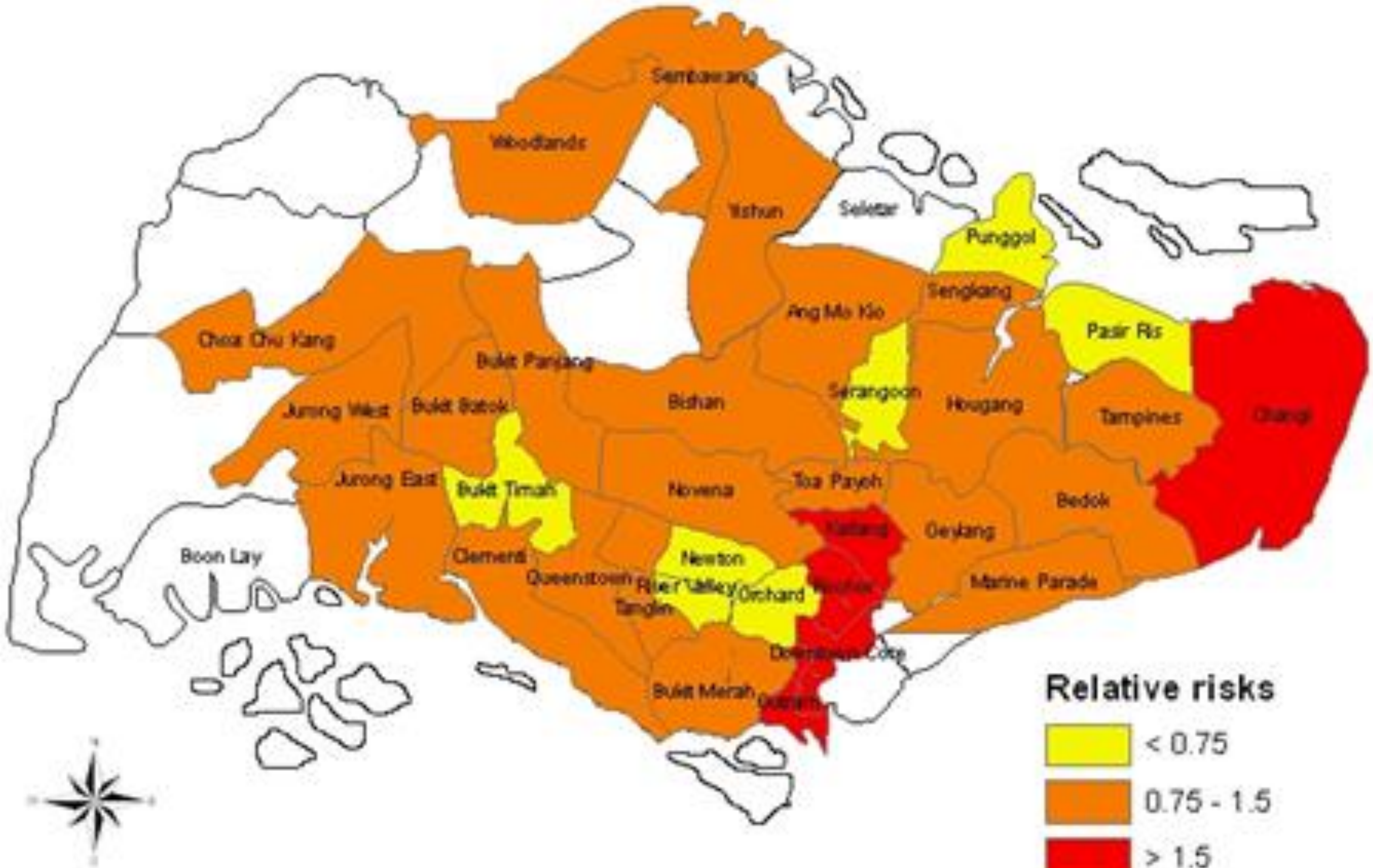
Some more info on the Bayesian model:

1) Neighbourhood structure based on Queen method of assignment
2) Priors were non-informative
3) 2 diffuse chains used and convergence assessed



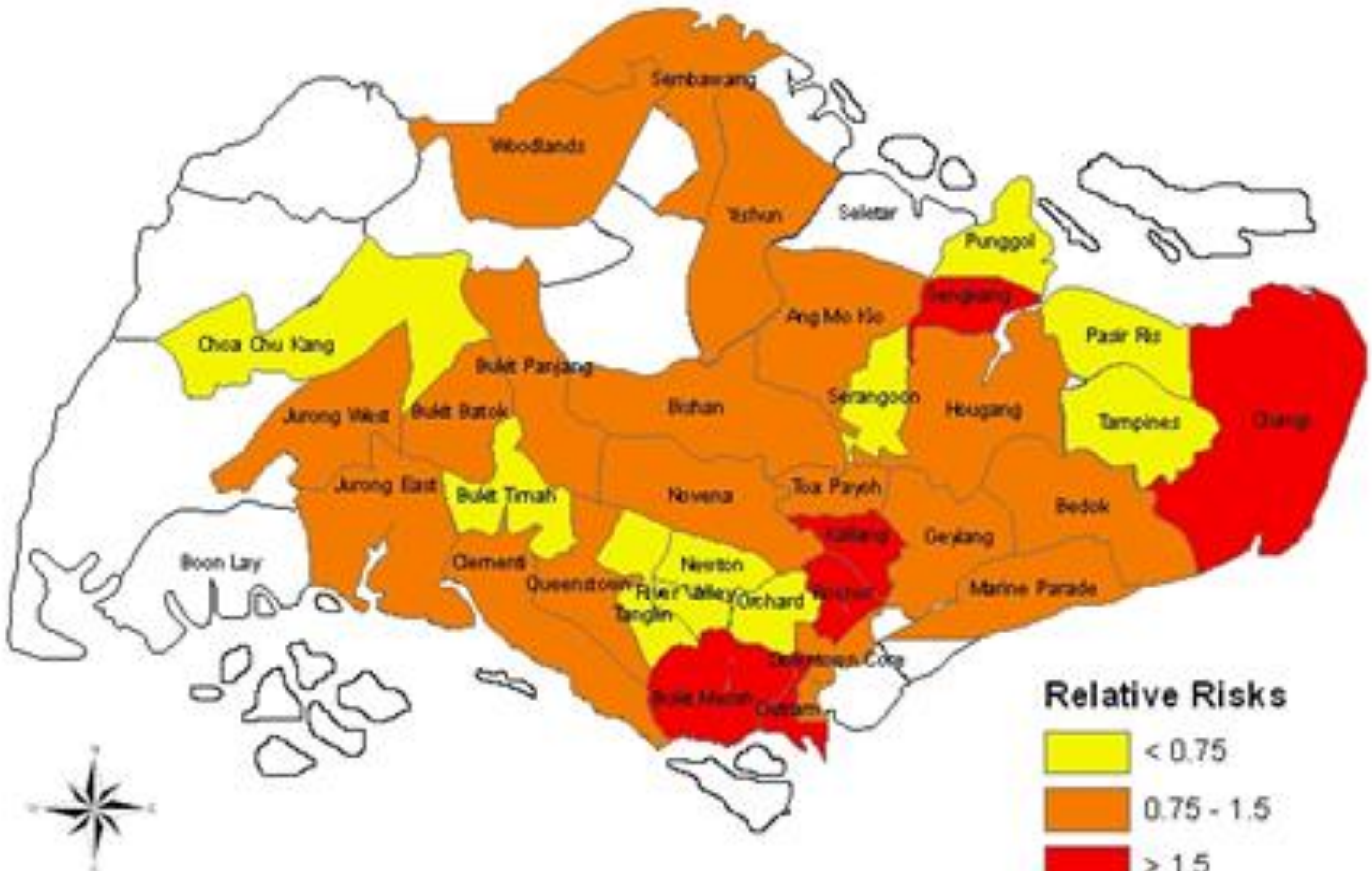
Results

Figure 1. Smoothed relative risk of ambulance calls for trauma cases



Results

Figure 2. Smoothed relative risk of ambulance calls for medical cases



Results

Table 5. Factors associated with risk of medical-related ambulance calls

Factors	RR	95% CI		DIC
Proportion aged 65 and above	0.55	0.12	4.22	341.60
Proportion male	**0.76**	**0.44**	**0.97**	336.70
Proportion Chinese	1.01	0.79	1.40	337.77
Proportion high education	1.68	0.29	5.47	338.19
Proportion living in 5-room flat and above	0.85	0.66	1.04	334.39
Proportion working	0.84	0.55	1.04	335.66
Proportion senior officers and professionals	**0.69**	**0.47**	**0.94**	342.64
Proportion with no family nucleus	2.01	0.96	6.42	341.52
Proportion with household size 5 and above	0.73	0.47	1.84	336.67
Proportion travelling by car alone	**0.71**	**0.57**	**0.98**	331.33
Proportion household income $5000 and above	**0.66**	**0.56**	**0.79**	330.03
Proportion household income $6000 and above	**0.72**	**0.59**	**0.96**	334.21
Proportion household income $7000 and above	**0.73**	**0.58**	**0.96**	335.88
Proportion household income $8000 and above	**0.69**	**0.43**	**0.88**	338.71
Proportion personal income $4000 and above	0.73	0.49	1.45	339.46
Proportion personal income $5000 and above	**0.67**	**0.24**	**0.95**	344.02
Proportion personal income $6000 and above	0.67	0.43	1.04	341.06

Note: relative risks reported for a 10% point increase

Results

Table 6. Factors associated with risk of trauma-related ambulance calls

Factors	RR	95% CI		DIC
Proportion aged 65 and above	2.51	0.88	19.89	286.98
Proportion male	1.07	0.25	3.10	286.75
Proportion Chinese	0.76	0.50	1.19	286.95
Proportion high education	1.88	0.13	10.28	288.86
Proportion living in 5-room flat and above	0.90	0.79	1.06	287.23
Proportion working	0.63	0.35	2.72	287.08
Proportion senior officers and professionals	0.80	0.50	1.07	290.30
Proportion with no family nucleus	1.67	0.61	8.08	287.44
Proportion with household size 5 and above	**0.63**	**0.33**	**0.90**	285.98
Proportion travelling by car alone	0.77	0.59	1.02	286.92
Proportion household income $5000 and above	**0.76**	**0.57**	**0.91**	285.35
Proportion household income $6000 and above	**0.79**	**0.62**	**0.93**	286.55
Proportion household income $7000 and above	**0.77**	**0.57**	**0.97**	288.38
Proportion household income $8000 and above	0.76	0.43	1.03	289.66
Proportion personal income $4000 and above	0.80	0.63	1.09	289.07
Proportion personal income $5000 and above	0.79	0.45	1.03	290.00
Proportion personal income $6000 and above	0.75	0.52	1.06	291.06

Note: relative risks reported for a 10% point increase

Conclusion

* Ambulance calls in Singapore demonstrate a clear spatial gradient
* Risk of making such calls decreases for areas with an increased socio-economic status
* Results can help policy makers target specific populations at risk with focused campaigns as well as more effective ambulance deployment.

Case Study 3- Birth Defects in Australia

Sometimes, we can have events that are rare (sparse) like birth defects

e.g. good Statistician?

, and the CAR model does not really work well in this situation, as it borrows strength from neighbours, which themselves are sparse.

Need a more innovative solution….

Solution

Borrow strength from a related outcome

Earnest A, Beard JR, Morgan G, Lincoln D, Summerhayes R, Donoghue D, Dunn T, Muscatello D, Mengersen K. Small area estimation of sparse disease counts using shared component models-application to birth defect registry data in New South Wales, Australia. Health Place. 2010 Jul;16(4):684-93. Epub 2010 Feb 25.

Some more maps

Figure 1. Crude relative risk estimates of Trisomy 21

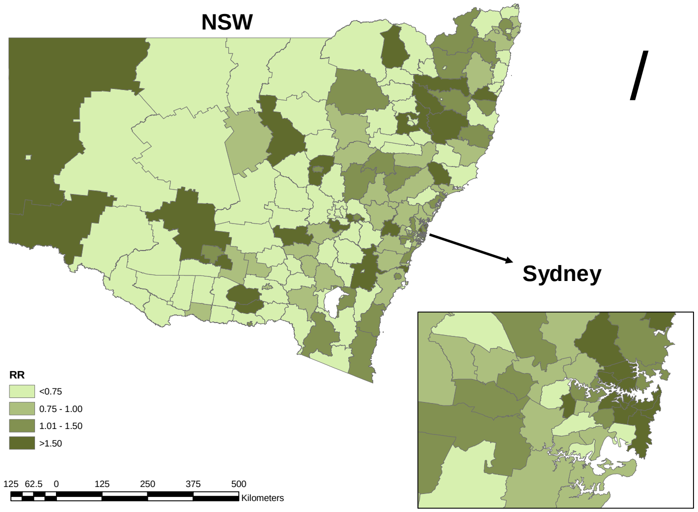


Figure 2. Smoothed relative risk estimates of Trisomy 21 (Shared component ZIP model)

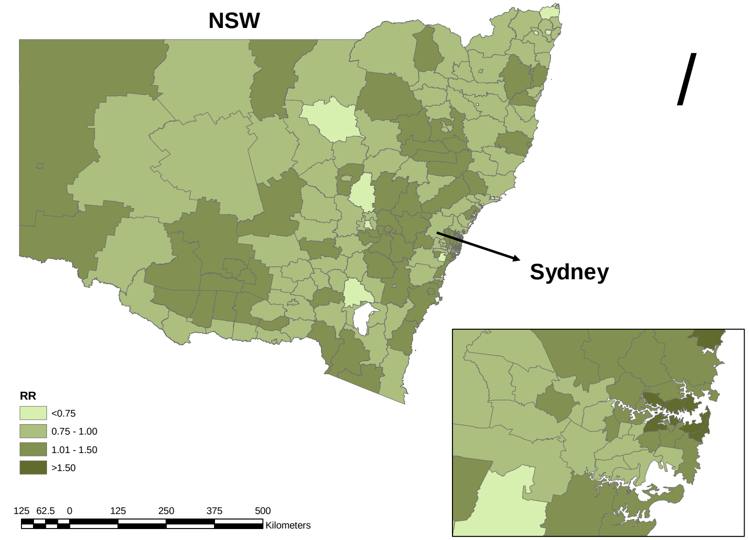


Figure 3. Crude relative risk estimates of caesarean rates

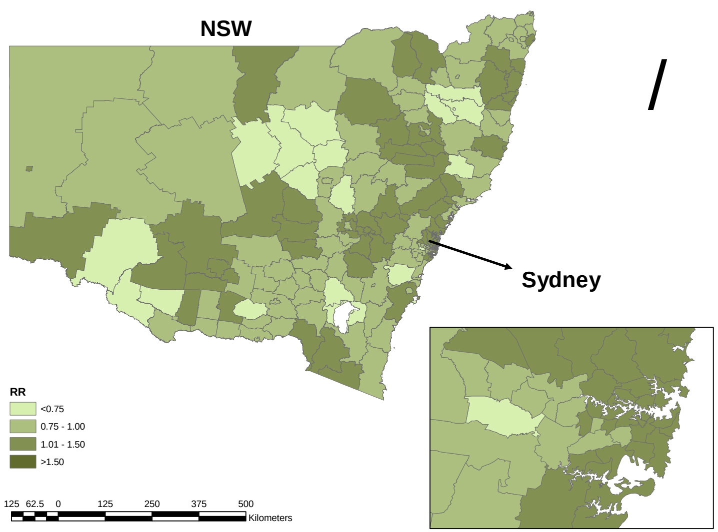
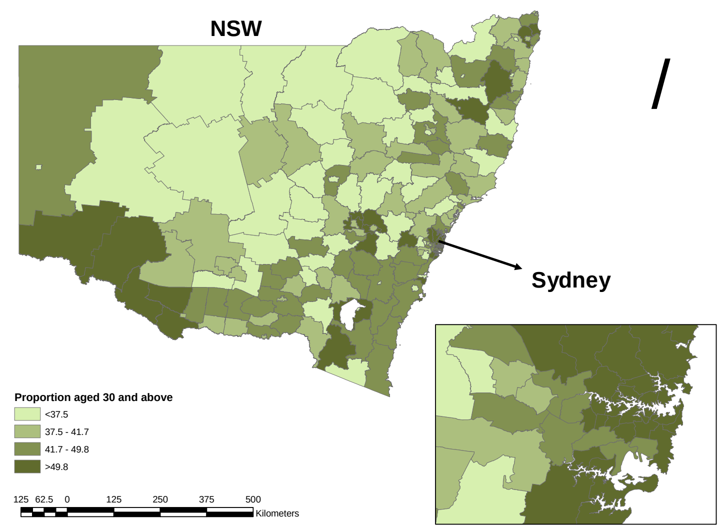


Figure 4. Proportion of births among mothers aged 30 and above



Some (rather painful) WinBUGs code

Appendix 1. Copy of WinBUGS code for the shared component ZIP model

```
model {
# Likelihood
for (i in 1:N) {
for (k in 1:2) {
zeros[i,k] <- 0
zeros[i,k] ~ dpois(mu[i,k])
mu[i,k]<-(1-step(O[i,k]))*(-log(1-p[i,k]   +   p[i,k]*exp(-lambda[i,k])))   +   (step(O[i,k]-1))*(-(log(p[i,k])-
lambda[i,k] + O[i,k]*log(lambda[i,k])-logfact(O[i,k])))
logit(p[i,k]) <- alpha0[k]
log(lambda[i,k])<-log(E[i,k]) + eta[i, k]
}  }
for(i in 1:N) {
# Define log relative risk in terms of disease-specific (psi) and shared (phi) random effects
eta[i,1] <- phi[i] *delta + psi[1, i]     # changed order of k and i index for psi
eta[i,2] <- phi[i] /delta + psi[2, i]     # (needed because car.normal assumes right hand index is areas)
}
# Spatial priors (BYM) for the disease-specific random effects
for (k in 1:2) {
for (i in 1:N) {
psi[k, i] <- bind[k, i] + bspat[k, i]            # convolution prior = sum of unstructured and spatial effects
 # unstructured disease-specific random effects
bind[k, i] ~ dnorm(alpha[k], vind[k])
}
# spatial disease-specific effects
```

Case Study 4- IHD Admissions in NSW

IHD and IRSD

Beard JR, Earnest A, Morgan G, Chan H, Summerhayes R, Dunn TM, Tomaska NA, Ryan L. Socioeconomic disadvantage and acute coronary events: a spatiotemporal analysis. Epidemiology. 2008 May;19(3):485-92.

Impact Factor: 4.0

Aim

Relationship between socioeconomic factors and acute coronary events

Outcomes: deaths from acute myocardial infarction, hospital admissions for acute coronary syndrome and related revascularization procedures

Socio-economic measure: IRSD

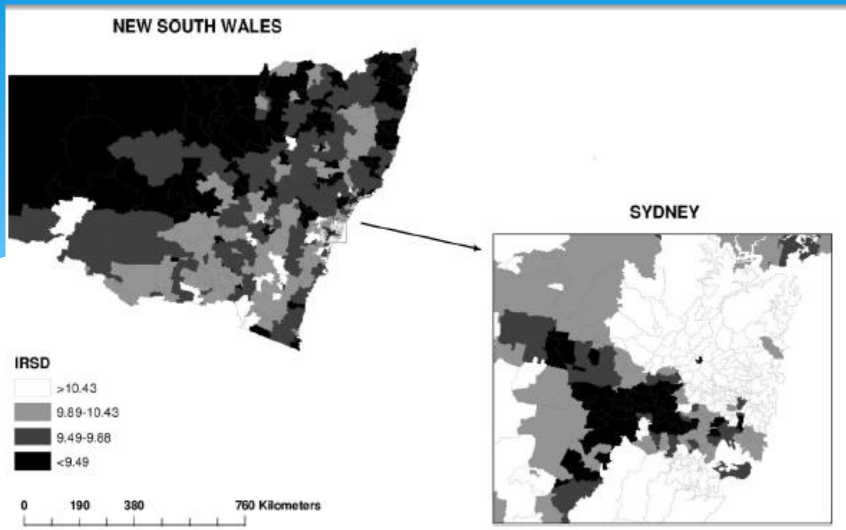
Bayesian Model

Bernardinelli spatiotemporal model: where both the area-specific intercepts and temporal trends were modeled as random effects

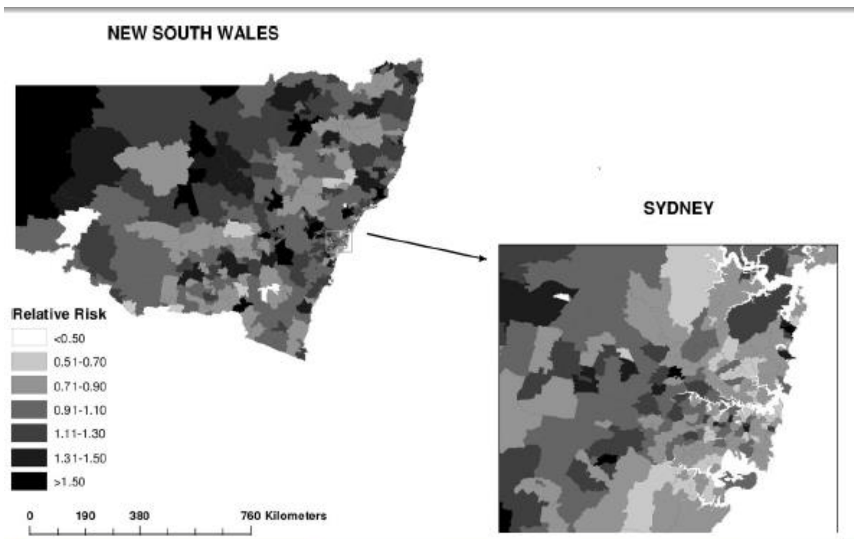
$$O_{ik} \sim Poi(\mu_{ik})$$

$$\log(\mu)_{ik} = \log(E_{ik}) + u_i + v_i + \beta_{1i} * t_k + \beta_{2i} * t_k^2$$

Results



Map of IRSD



Map of AMI mortality, after adjusting for spatio-temporal effects

Results

TABLE 2. Multivariate Conditional Autoregressive Analysis of Socioeconomic Disadvantage and Other Covariates for Acute Coronary Syndrome Admissions, Mortality, and Procedures

	Admissions		Mortality		CABG		Angiography		Interventional Angiography	
Covariate	RR (95% CI)	DIC	RR (95% CI)	DIC	RR (95% CI)	DIC	RR (95% CI)	DIC	RR (95% CI)	DIC
Disadvantage with Indigenous		21,715		15,427		5862		11,172		7612
Disadvantage										
>10.44[a]										
9.89–10.44	1.12 (1.03–1.21)		1.12 (1.03–1.23)		0.93 (0.73–1.20)		0.86 (0.75–0.98)		0.90 (0.74–1.10)	
9.49–9.89	1.27 (1.15–1.39)		1.19 (1.09–1.31)		0.97 (0.75–1.28)		0.81 (0.69–0.94)		0.79 (0.64–0.99)	
<9.49	1.34 (1.21–1.49)		1.36 (1.23–1.51)		0.84 (0.63–1.13)		0.72 (0.61–0.86)		0.68 (0.54–0.87)	
Indigenous proportion										
0–3[a]										
3–10	1.17 (1.08–1.27)		1.06 (0.98–1.15)		1.18 (0.90–1.56)		1.00 (0.86–1.17)		1.00 (0.79–1.25)	
>10	1.21 (1.03–1.44)		1.24 (1.05–1.47)		1.12 (0.54–2.25)		1.21 (0.80–1.83)		1.27 (0.71–2.23)	
Disadvantage with Sydney postal areas		21,712		15,432		5862		11,166		7602
Disadvantage										
>10.44[a]										
9.89–10.44	1.12 (1.03–1.24)		1.11 (1.02–1.21)		0.97 (0.76–1.22)		0.87 (0.77–0.98)		0.90 (0.76–1.09)	
9.49–9.89	1.31 (1.18–1.47)		1.19 (1.08–1.30)		1.04 (0.79–1.35)		0.82 (0.73–0.93)		0.79 (0.66–0.97)	
<9.49	1.43 (1.29–1.62)		1.39 (1.26–1.53)		0.93 (0.69–1.23)		0.75 (0.65–0.87)		0.70 (0.57–0.88)	
Sydney postal areas	1.33 (1.06–1.56)		0.95 (0.84–1.09)		1.39 (0.75–2.25)		1.11 (0.80–1.54)		1.61 (1.09–2.46)	

[a]Reference category. For socioeconomic disadvantage, reference category is least disadvantaged, that is, most advantaged.

Conclusion

Strong association between socioeconomic disadvantage and mortality from acute MI. This relationship appears to operate both by increasing the risk of developing disease and by reducing the chance of receiving optimal care.

Found substantial spatial variation in acute MI mortality, hospital admissions for acute coronary syndrome,and related hospital procedures.

Conclusion

Areas with high levels of socioeconomic disadvantage showed higher rates of both hospital admissions and mortality.

After accounting for increased admission rates, these same areas showed lower rates of interventions such as angiography and interventional an- giography, although rates of coronary artery bypass grafts (a more established intervention) were not associated with socioeconomic disadvantage.

Final Conclusions

Bayesian CAR model can incorporate spatial correlation in data

Free software exist

Computational advances mean faster time to run MCMC chains

Uses non-informative priors (appeals to the many frequentists among us)

Thank you!

Arul Earnest

arul_earnest@hotmail.com