

传染病监测数据与 R 软件

张 兵

zhangbing4502431@outlook.com

广东省公共卫生研究院

广东省疾病预防控制中心

2017 年 4 月 2 日

① 传染病监测数据

② 传染病相关特性

③ 传染病监测预警

④ 气候变化与传染病

⑤ 疫苗接种与传染病

⑥ 自动化报告与传染病

⑦ 总结与设想

现场
防控

科学
研究



Source : Modeling infectious disease dynamics in the complex landscape of global health (Science 2015)

传染病特征

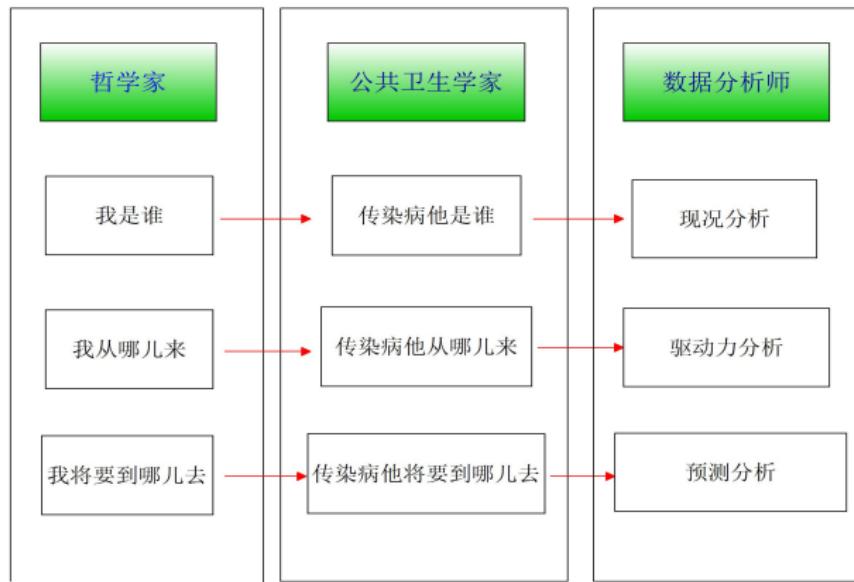
被忽视了的**时间、空间和人口分布特征**。

- 空间特性包括空间依赖性、空间异质性、尺度效应等: → 个案的地址信息
- 时间特性包括短期波动、长期趋势、季节性和周期性: → 发病时间或针对时间
- 人口分布特征主要是指年龄、性别等特性: → 出生日期和性别

传染病流行规律: 在自然与社会多源复杂因子作用下传染病流行的时空变化及其相关因素的时空共变轨迹。

如何做

该如何分析传染病数据？→ 以问题为导向



做什么

如何解释传染病的相关特性

如何解释传染病的区域异质性

如何更快更早识别传染病暴发

如何评估干预措施（疫苗接种）的效果

· · · ·



传染病相关特性研究

何种特性

- 传播能力
- 传播速度
- 报告率
- 年龄分布

基本再生数

基本再生数（Basic reproduction number， R_0 ）是指一个感染者进入到全部为易感者的人群中，在其患病期内能够感染的人群总数。

- R_0 最初由 Alfred Lotka 和 Ronald Ross 提出。
- George MacDonal 在 1952 年首次将 R_0 应用于流行病学领域（蚊媒传染病疟疾）
- R_0 定量地反映了传染病的传播能力， R_0 越大，其传播能力越强。若 $R_0 > 1$ ，则传染病将持续存在，反之，则疾病将灭绝。
- 疫苗接种率 $> 1 - (1/R_0)$ ，才能有效阻止传染病流行。Why？

部分疾病的基本再生数：

- | | |
|---------------|----------------|
| ● 麻疹：12-18 | ● 流行性腮腺炎：4-7 |
| ● 白喉：6-7 | ● HIV/AIDS：2-5 |
| ● 天花：5-7 | ● 百日咳：5.5 |
| ● 脊灰：5-7 | ● SARS：2-5 |
| ● 埃博拉：1.5-2.5 | ● 流感：2-3 |

其他再生数

在 R_0 的基础上衍生出有效再生数、病例再生数和即时再生数。

- Effective reproduction number (有效再生数)：当目标人群中不全是易感人群时，平均每个感染者在其患病期间内能够感染的人群总数。
- Case reproduction number (病例再生数： R_t^c)：从 t 时刻开始，经过一个传代间隔后，平均每个感染者能感染人群总数。
- Instantaneous reproduction number (即时再生数 R_t)：假设外环境不变情况下，在 t 时刻的感染者平均感染人群总数。

即时再生数，可作为传染病干预效果评估的一个有效指标。

Source : A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics (American journal of epidemiology 2013)

实现工具

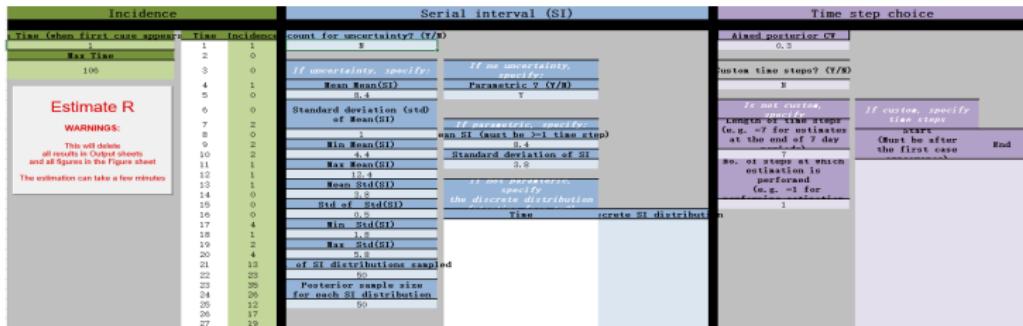
- 基于统计模型，具体可见 R_0 和 EpiEstim 程序包（[简介](#)）
- 基于数据模拟，依靠传染病动力学模型，计算下代再生矩阵。

EpiEstim 程序包的 EXECL 版本：<http://tools.epidemiology.net/EpiEstim.xls>

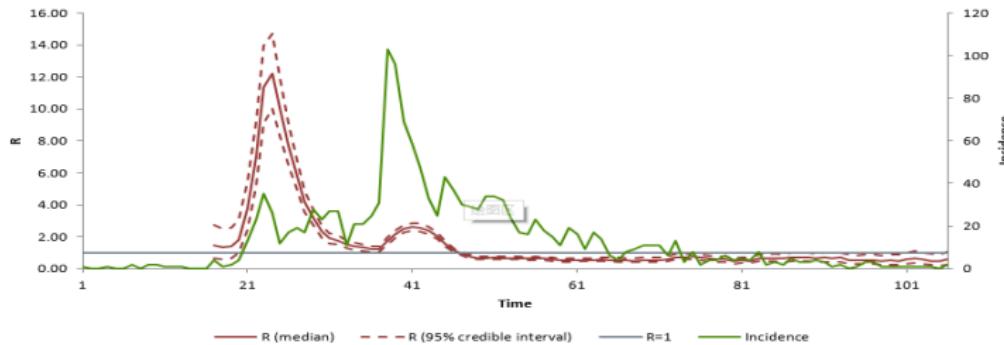
This excel spreadsheet allows estimating of the instantaneous reproduction number for an epidemic for which the incidence is observed and assumptions on the serial interval distribution are made.

Click on the "Estimate R" button in the "Data" sheet to see an example of estimation.

操作与结果



R averaged over time periods (posterior median and 95%CrI)



传播速率

- 传播速率 (rate of spread) 反映传染病在整个人群中的蔓延速度, 而传染病在个体间的传播速率则用传代间期 (generation interval) 表示。
- 传播速率与基本再生数间可通过传代间期 (generation interval) 有效地连接起来。

$$R = e^{rT_c - (1/2)r^2\sigma^2}$$

(T_c 为平均传代间期, σ 为传代间期的标准差, r 为传播速率)

- 在传染病暴发期间, 寻找高传播速度的时段并及时采取措施, 对控制传染病的发生有重要影响。

实现方式: GI程序包, 发布在 Github 上, 需要基于 devtools 进行安装。

Source: Intrinsic and realized generation interval in infectious-disease transmission(Proc. R.Soc. B 2015)

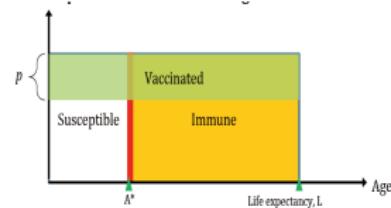
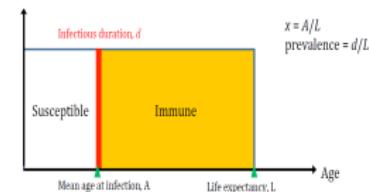
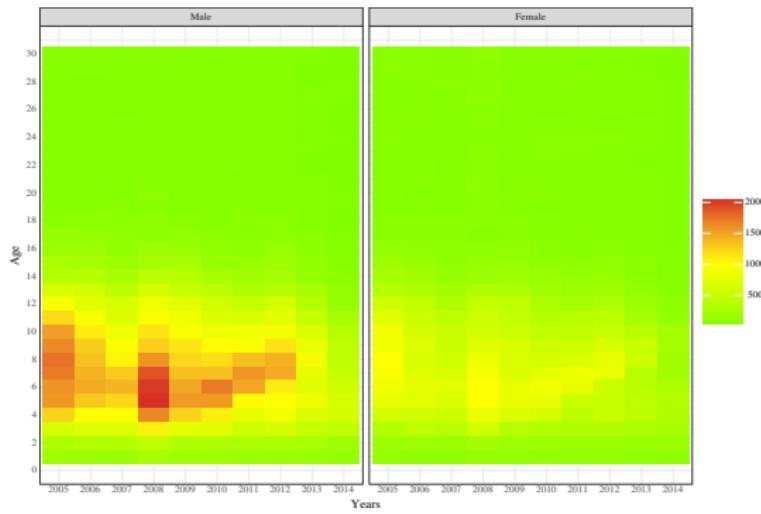
报告率

- 报告率 (reporting rate) 为易感者感染后去医院就诊并被登记成病例的比例。
- 不同传染病的报告率不一致，与疾病严重程度、监测系统敏感性等因素有关。
- 正确评估传染病的报告率，可有效评估传染病的疾病负担。

实现方式：构建传染病动力学模型，并假设状态 E 到状态 I 的转化人数 (P_{SE}) 与报告病例数 (case) 存在某种关联。如 $\text{case} = \rho P_{SE}$ ，其中 ρ 为报告率。

Source: Plug-and-play inference for disease dynamics: measles in large and small populations as a case study (J. R. Soc. Interface 2010)

年龄分布：以流行性腮腺炎为例



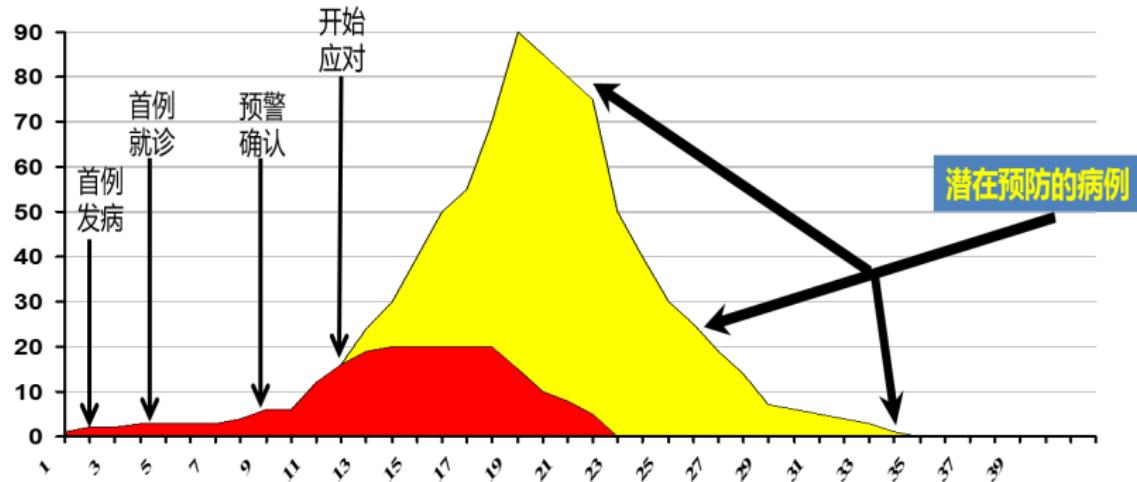
Vaccination increases mean age at infection: $A^* = A/(1 - P)$

A 和 A^* 分别表示疫苗接种前后的平均感染年龄， P 表示疫苗接种率

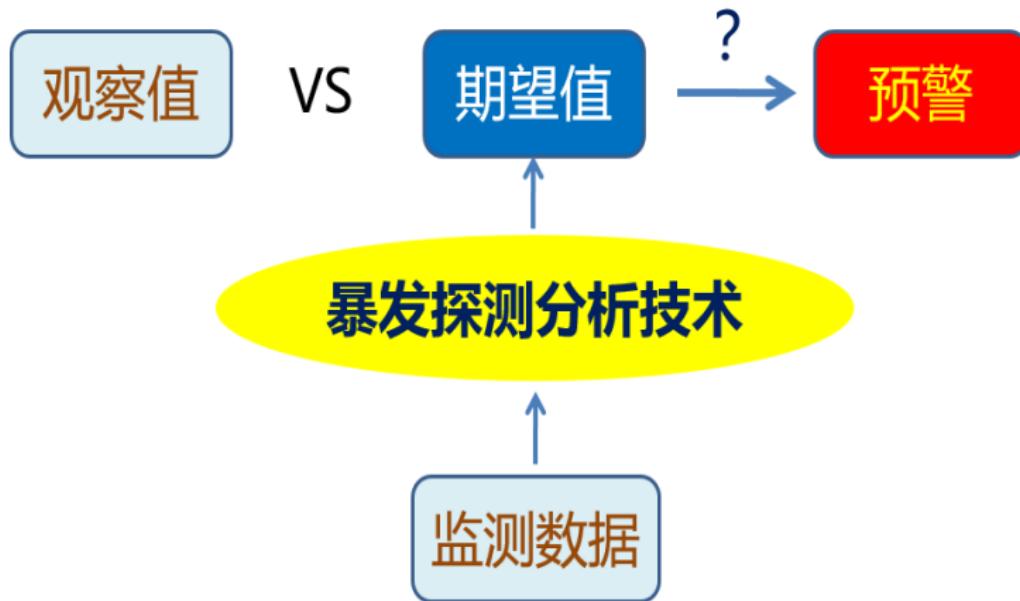
传染病监测预警

暴发探测

- 暴发探测是指识别传染病的时间、空间和人群维度上表现出具有流行病学联系的聚集性特征。
- 暴发探测并及时预警应对有助于疾病的预防和控制。



暴发探测原理



时间维度主流方法

Function	References
bayes	Riebler (2004)
boda	Manitz and Höhle (2013)
bodaDelay	Salmon <i>et al.</i> (2015)
categoricalCUSUM	Höhle (2010)
cdc	Stroup, Williamson, Herndon, and Karon (1989); Farrington and Andrews (2003)
cusum	Rossi, Lampugnani, and Marchi (1999); Pierce and Schafer (1986)
earsC	Fricker <i>et al.</i> (2008)
farrington	Farrington <i>et al.</i> (1996)
farringtonFlexible	Farrington <i>et al.</i> (1996); Noufaily <i>et al.</i> (2012)
glrbn	Höhle and Paul (2008)
glrpois	Höhle and Paul (2008)
outbreakP	Frisén, Andersson, and Schiöler (2009); Frisén and Andersson (2009)
pairedbinCUSUM	Steiner <i>et al.</i> (1999)
rki	Not available – unpublished
rogerson	Rogerson and Yamada (2004)

Surveillance 包集成了主流暴发探测方法，包括统计过程控制类和参考值类两类算法

Source: Monitoring Count Time Series in R: Aberration Detection in Public Health Surveillance (Journal of Statistical Software 2016)

累计和控制图法 (CUSUM)

原理：通过对观察值与期望值的差值进行累加后，放大数据出现的波动，可更快速、灵敏地探测到休哈特控制图法无法识别的微小异常。

- 短期 EARS C1 方法：

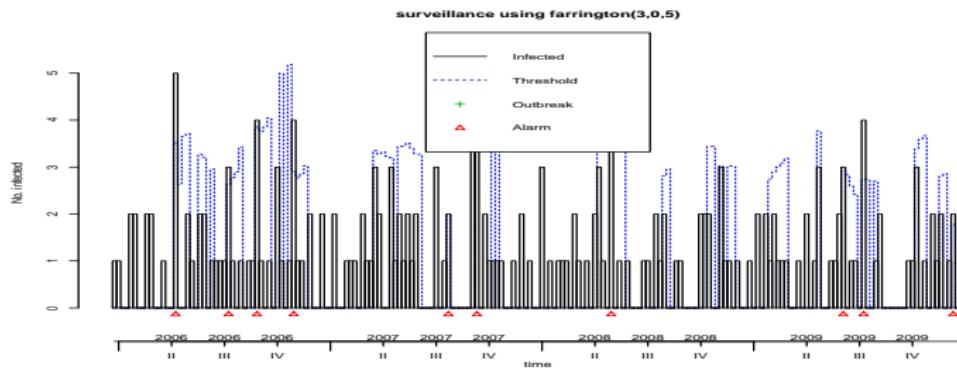
$$\bar{Y}_1(t) = \frac{1}{7} \sum_{i=t-1}^{t-7} Y(i) \text{ 且 } S_1^2(t) = \frac{1}{6} \sum_{i=t-1}^{t-7} [Y(i) - \bar{Y}_1(i)]^2$$
$$U_t = \bar{Y}_1(t) + Z_{1-\alpha} S_1(t)$$

若 $Y_1(t) > U_t$ ，则提示 $Y_1(t)$ 为异常值。

- 在 EARS C1 (度量时刻：t-7 到 t-1) 的基础上衍生出 EARS C2 (度量时刻：t-9 到 t-3) 和 EARS C3(度量时刻：t-11 到 t-3)
- C3 灵敏度最高，C2 次之，C1 最低。

Farrington 模型

- 英国 Farrington, C. 教授于 1996 年首次提出。
- 基于广义线性模型（GLM）对病例时间序列做暴发探测。
- Farrington 相关更新模型，已被开发到欧洲部分国家的监测体系里。



Source: A statistical algorithm for the early detection of outbreaks of infectious disease
(J. R. Statist. Soc 1996)

Farrington 模型三部曲

- 第一步：用观察数据拟合 log 连接的过度离散 GLM。为了排除季节趋势的影响，只使用预警时间点前几年附近几周的数据作为参考值。为了提高敏感性，预警时间点前 26 周的数据也排除。
- 第二步：用上述 GLM 模型预测预警时间点的值及其上限。旧方法假设是正态分布，新方法假设是负二项分布。
- 第三步：实际观察值与预测值上限比较，如果超过预测值上限，则发布预警。

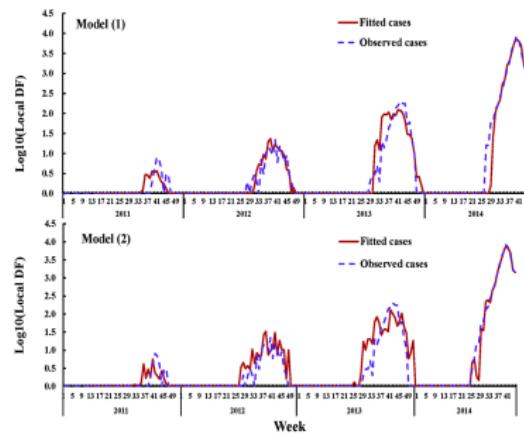
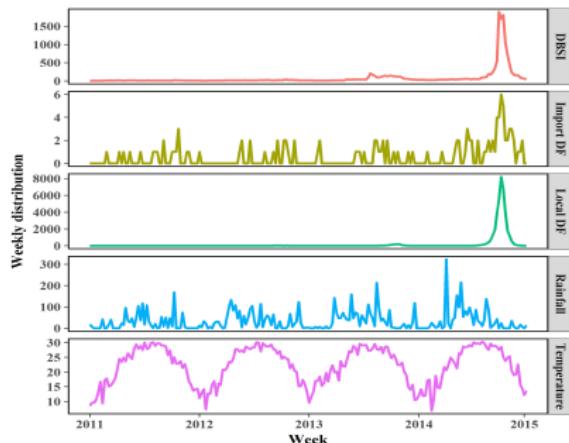
```
phase2013 <- which(epoch(flустs) >= as.numeric(as.Date("2015-02-03")))

control2 <- list(range = phase2013, b = 4, w = 3, weightsThreshold =
2.58,pastWeeksNotIncluded = 26, pThresholdTrend = 1, thresholdMethod =
"nbPlugin")

flu.noufaily <- farringtonFlexible(flустs, control2)
```

大数据与暴发探测

网络搜索数据应用：Google Trend、百度指数、微信指数、维基百科指数等指标纳入传染病暴发探测是否可行？

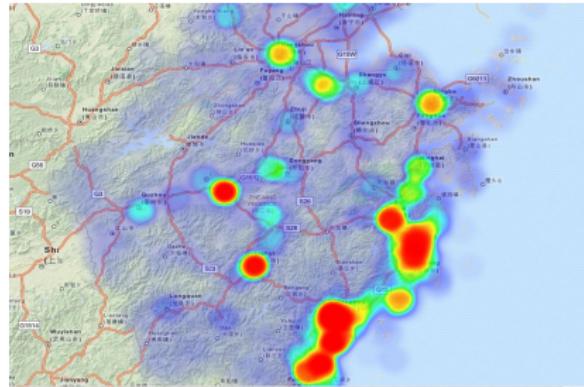


Source: Dengue Baidu Search Index data can improve the prediction of local dengue epidemic: A case study in Guangzhou, China (PLoS NTD 2017)

空间聚集性探测

发现传染病在空间层面是否出现聚集特性 (Hotspots) 及聚集强度:

- 空间聚集模型 : Moran's I、Geary's C、K 函数
- 时空聚集模型 : Knox 分析、时空扫描统计量、Rogerson 统计量
(DCluster 程序包、Surveillance 程序包、rsatscan 程序包)



空间回归模型：区域异质性

基于回归模型解释传染病在空间或时空分布的异质性 (Spatial Epidemiology):

- 空间回归模型：空间滞后模型、空间误差模型、空间杜宾模型等
- 时空回归模型：贝叶斯时空模型、时空多层分模型、时空点密度模型

**All models are wrong, but some
are useful ...**

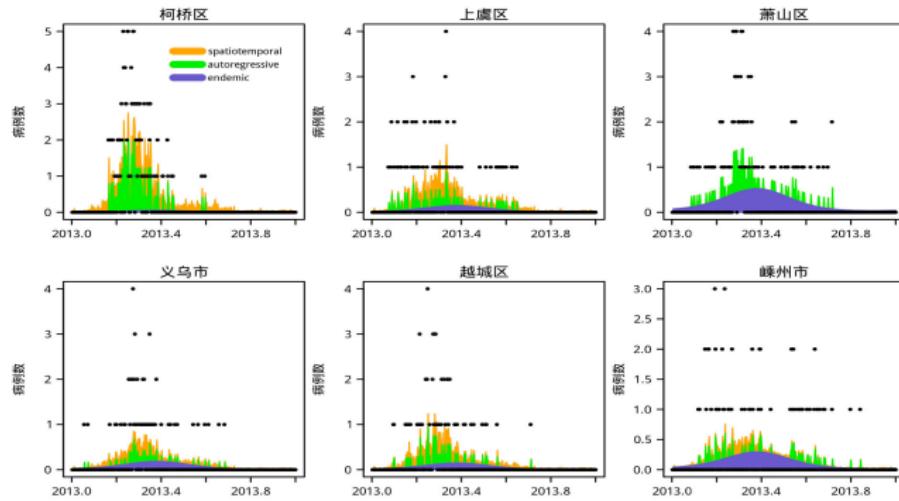
*“The models can provide a means to
systematically compare alternative intervention
strategies, determine the most important issues
in decision-making, and identify critical gaps in
current knowledge.”*

McKenzie, F.E. 2004. Emerging Infectious Diseases.

统计模型，想说爱你不容易

时空多层分模型

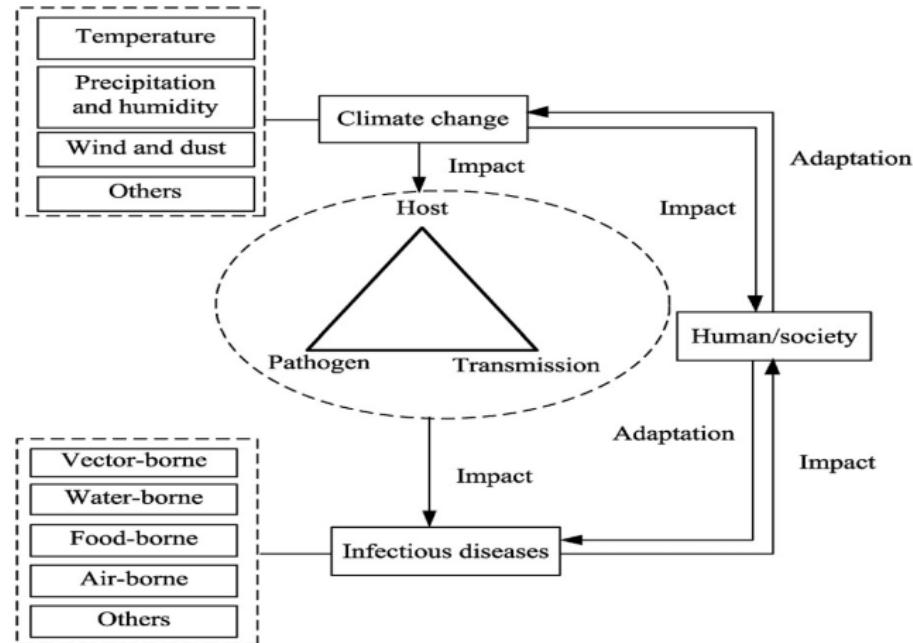
时空多层分模型，是将各地区传染病发病风险分为**本地成分**(本地风险)、**空间成分**(来自临近区域的风险)和**时间自相关成分**(过去时段传染病发病对现阶段的影响)



Source: 浙江省 2013 年麻疹疫情时空分布特征 (中华流行病学 2016)

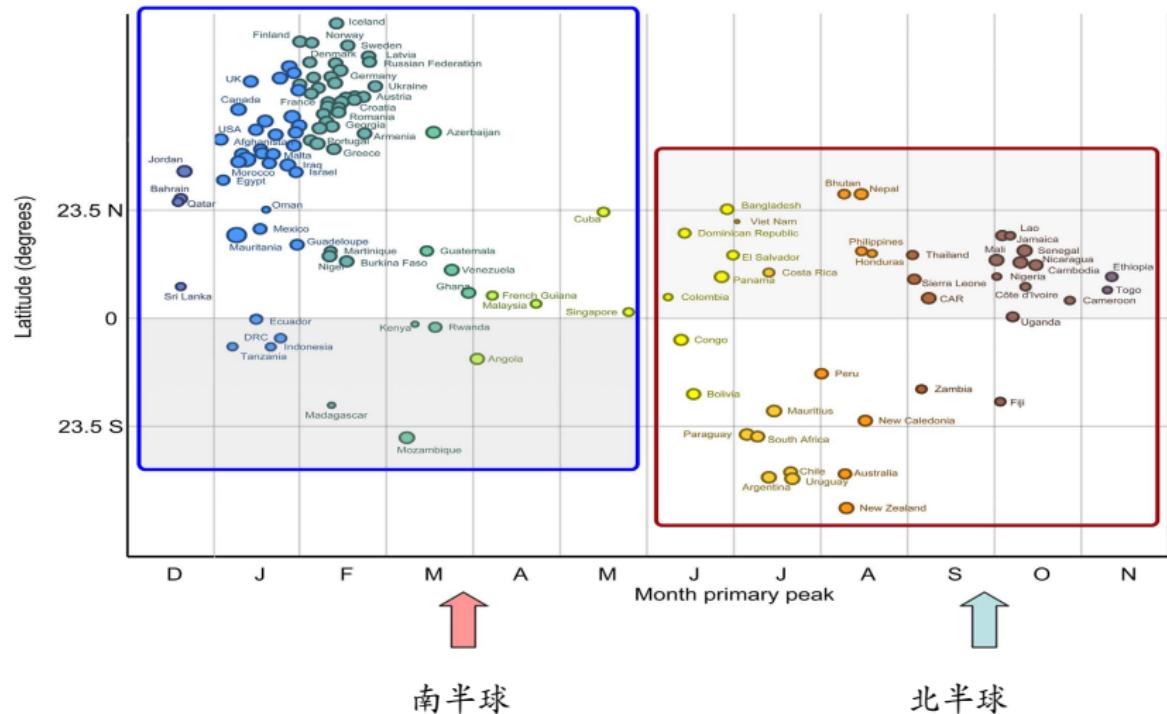
气候变化与传染病

气象因素与传染病



Source: Impact of climate change on human infectious diseases: Empirical evidence and human adaptation (Environment International 2016)

季节性差异：以流感为例



Source: A global map of hemispheric influenza vaccine recommendations based on local patterns of viral circulation (Scientific Reports 2015)

常规统计方法

广义相加模型

·论著·

广义相加模型拟合气象因素与猩红热发病的关联性

吴昊进, 林君萍, 徐校平, 吴晨, 鲁琴宝, 丁督训

浙江省疾病预防控制中心, 浙江 杭州 310051

摘要: 目的 探索气象因素与猩红热发病的关联性。方法 收集浙江省 2005—2014 年猩红热发病数据以及同期气温、气压等气象资料, 应用广义相加模型分析气象因素与猩红热发病之间的关联程度和形式。结果 平均水汽压、日照时数与猩红热发病间存在负相关; 降水量、平均气压、平均风速、平均气温与猩红热发病间均有存在较复杂的非线性关系, 其中平均气压、平均风速和平均气温对猩红热发病影响呈猩红二次曲线关系。平均气压 < 10180 (0.1 hPa) 时是正效应; 在 107~23.6 (0.1 m/s) 风速范围内, 对猩红热发病有较小的正效应; 平均气温在 < 250 (0.1 °C) 时, 对猩红热发病的影响负效应逐渐减弱。结论 气象因素与猩红热发病之间存在较复杂的非线性关系, 降水量、气压、风速以及温度与猩红热可能存在关联。

关键词: 广义相加模型; 猩红热; 气象因素; 关联性

中图分类号: R188; R515.1 文献标识码: A 文章编号: 1007-0831(2016)01-0005-05

分布非线性滞后模型

·论著·

分布滞后非线性模型评估室外温度及湿度对手足口病发病的影响

吴一峰¹, 赵凤敏², 贺天华², 崔一帆¹, 许国雷²

1. 宁波市江北区疾病预防控制中心, 浙江 宁波 315021; 2. 宁波市疾病预防控制中心

摘要: 目的 利用模型估算室外温度和相对湿度对手足口病发病的影响。方法 收集宁波市江北区 2011~2014 年手足口病病例资料以及同期气象资料, 采用分布滞后非线性模型 (DLNN) 计算日均温度和日均相对湿度变化与手足口病发病率数据的相对强度。结果 将温度、相对湿度以及寒暑假以不同的组合进行模型分析, 结果显示在独立模型中, 温度和湿度可以独立对手足口病发病率产生影响。而在综合模型中, 日均温度 (18~25 °C) 或 5 °C 及以下, 相对湿度在 60% 以下, 手足口病发病率增加, 并存在滞后效应, 相对湿度在 85%~100%, 温度 1~4 天时对手足口病发病率有保护作用。结论 室外温度和湿度在一定范围内的变化可影响手足口病发病, 但该变化产生影响有不同时间的滞后效应。

关键词: 温度; 湿度; 手足口病; 温滞后作用

中图分类号: R512.5 文献标识码: A 文章编号: 1007-0831(2016)02-0121-04

几何方法和动力学模型

Convergent cross-mapping



Author Affiliations

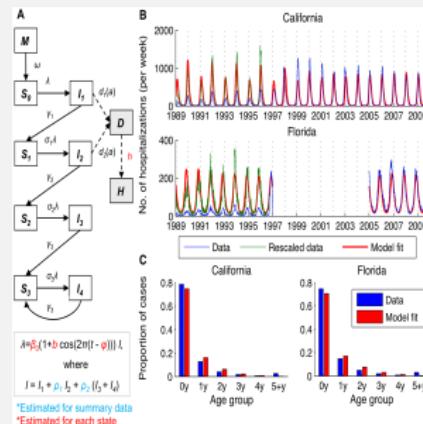
Edited by Alan Hastings, University of California, Davis, CA, and approved September 16, 2016 (received for review May 13, 2016)

Abstract | Full Text | Authors & Info | Figures | SI | Metrics | Related Content | PDF | PDF + SI

Significance

Patterns of influenza outbreak are different in the tropics than in temperate regions. Although considerable experimental progress has been made in identifying climate-related drivers of influenza, the apparent latitudinal differences in outbreak patterns raise basic questions as to how potential environmental variables combine and act across the globe. Adopting an empirical dynamic modeling framework, we clarify that absolute humidity drives influenza outbreaks across latitudes. Find that the effect of absolute humidity on influenza is U-shaped, and show that this U-shaped pattern is mediated by temperature. These findings offer a unifying synthesis that explains why experiments and analyses disagree on this relationship.

动力学模型



Source: Environmental Drivers of the Spatiotemporal Dynamics of Respiratory Syncytial Virus in the United States (PLoS Pathogens 2015)

特殊气候对传染病的影响

- 极端降雨与食源性疾病
- 高温热浪或寒潮与呼吸道传染病
- 极端天气与蚊媒传染病如寨卡

影响病原体的生长繁殖？影响宿主的免疫系统，还是改变人群的活动模式？

难点：作用靶点效应的分离

Source: Effects of Extreme Precipitation to the Distribution of Infectious Diseases in Taiwan, 1994–2008 (PLoS ONE 2012)

Analytical studies assessing the association between extreme precipitation or temperature and drinking water-related waterborne infections: a review (Environmental Health 2015)

模型拟合

基于隐马尔科夫模型，通过 pomp 程序包求解参数值及其 95% 置信区间。

为什么要用 pomp 这个程序包？多种新算法如迭代粒子滤波集成：

- The Impact of a One-Dose versus Two-Dose Oral Cholera Vaccine Regimen in Outbreak Settings: A Modeling Study (PLOS Medicine 2015)
- Inference for dynamic and latent variable models via iterated, perturbed Bayes maps (PNAS 2014)
- Unraveling the Transmission Ecology of Polio (PLoS Biol 2015)
- Inapparent infections and cholera dynamics (Nature 2008)
- Natural immune boosting in pertussis dynamics and the potential for long-term vaccine failure (PNAS 2011)
- Identifying the Interaction Between Influenza and Pneumococcal Pneumonia Using Incidence Data (Science 2013)

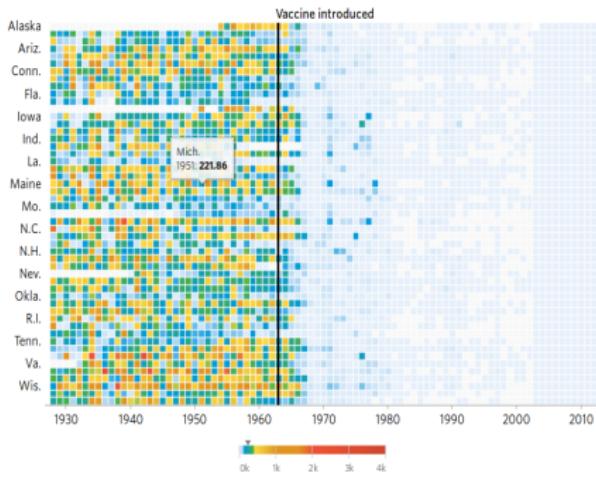
疫苗接种与传染病

Proactive intervention

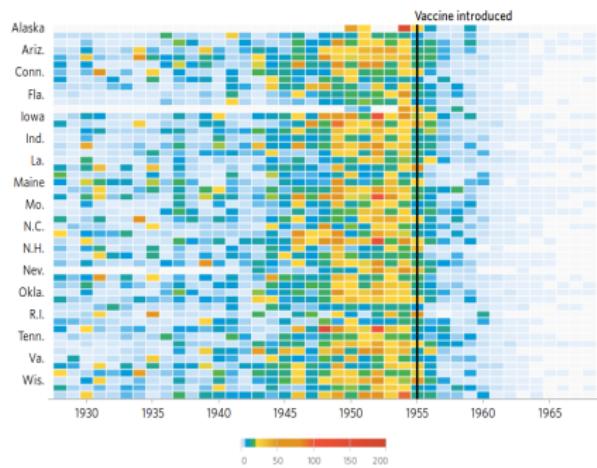
疫苗接种，是预防和控制传染病最经济、有效的公共卫生干预措施。

效益究竟如何？

Measles



Polio

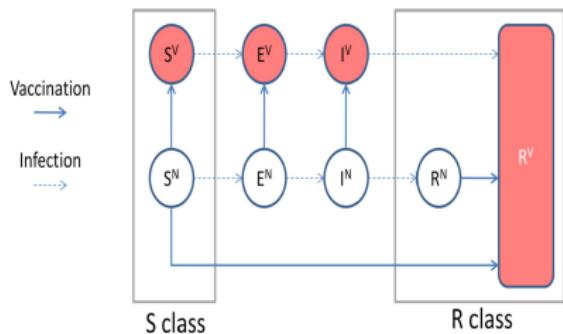


基于模型评估

无法忽视的**人群免疫效应**:

疫苗接种不仅保护接种者，也会降低其他易感者感染风险。

flu-evidence-synthesis: 不同接种策略下流感疫苗效果评估



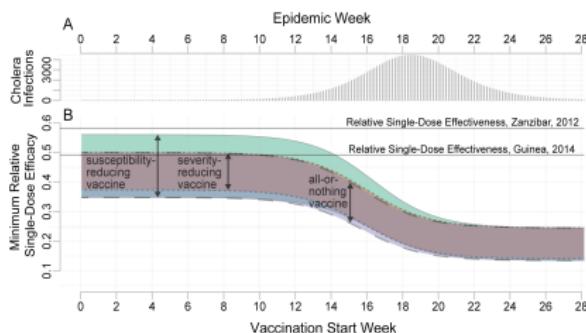
Using a contemporary evidence-synthesis approach, we use virological, clinical, epidemiological, and behavioural data to develop an age- and risk-stratified transmission model that reproduces the strain-specific behaviour of influenza over 14 seasons in England and Wales, having accounted for the vaccination uptake over this period. We estimate the reduction in infections and deaths achieved by the historical programme compared with no vaccination, and the reduction had different policies been in place over the period. **We find that the current programme has averted 0.39 (95% credible interval 0.34–0.45) infections per dose of vaccine and 1.74 (1.16–3.02) deaths per 1,000 doses.** Targeting transmitters by extending the current programme to 5–16-y-old children would increase the efficiency of the total programme, resulting in an overall reduction of 0.70 (0.52–0.81) infections per dose and 1.95 (1.28–3.39) deaths per 1,000 doses. In comparison, choosing the next group most at risk (50–64-y-olds) would prevent only 0.43 (0.35–0.52) infections per dose and 1.77 (1.15–3.14) deaths per 1,000 doses.

Source: Assessing Optimal Target Populations for Influenza Vaccination Programmes: An Evidence Synthesis and Modelling Study (PLoS Med 2013)

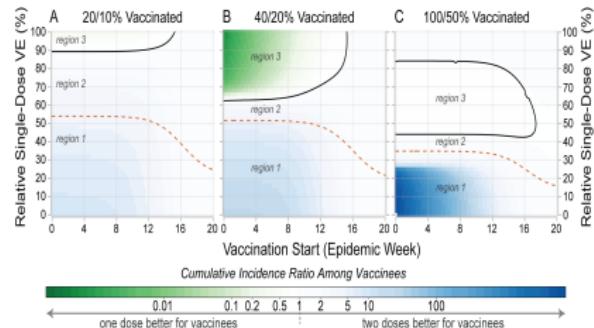
Reactive intervention

应急接种该如何做？

- When : 什么时候干预
- How: 如何干预



- Who: 对哪种人群进行干预
- Region: 临近区域是否需要干预

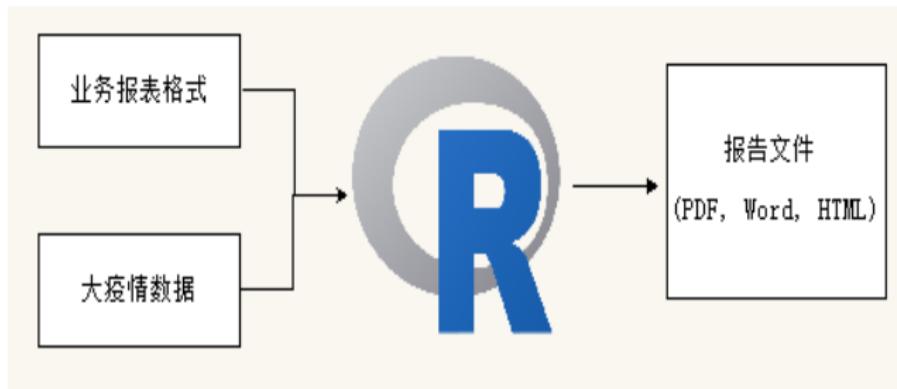


常规接种策略在应急接种时，是否可行？→ 精准公共卫生的落脚点在哪？

自动化报告与传染病

自动化报告

自动化报告 (automation report)，只需提供大疫情网下载数据，常规的业务报告即刻出现 → 人工智能在公共卫生领域的应用



可行性：业务报告格式固定 + 大疫情数据格式固定

总结与设想

总结与设想

- 传染病防控：基于监测数据服务于疾病防控

- 开发传染病预警平台，更快更早识别传染病的暴发
- 建立实时干预效果评估平台，及时了干预措施对解传染病的影响
- 自动化分析报告平台，数据分析、疫情分析报告结构化处理

- 科学研究：精细化(靶点好、数据精、分析方法细)。

- 传染病基本特征分析，若传播力、传播速度、报告率等
- 传染病干预效果评估（长期和短期），尤其是疫苗接种效果评估
- 传染病影响因素探讨，气象因素或其他因素究竟作用于传染病哪个环节

谢谢，欢迎指正！

个人主页：www.spatial-r.com

邮箱：zhangbing4502431@outlook.com

公众号：SpatialR