

中国武汉和 29 个省新冠肺炎传播趋势

背景:2019 年的新型冠状病毒感染肺炎(COVID19)是一个重大的公共卫生威胁。COVID-19 于 2009 年 12 月在中国武汉出现,并迅速传播到其他地区和国家。传播模式和疾病传播在时间或不同地点之间的变化,部分反映了公共卫生干预的效果,仍有待量化。由于大多数与传播相关的流行病参数都是未知的,我们试图在**最小假设**的情况下**估算实时传播率**,并使用**动态模型预测新病例**。

方法:用中国国家健康委员会报告的病例和交通数据,包括旅行时间的总数通过铁路、飞机,从武汉和汽车出站,我们已经建立了时间序列模型来估计实时新生成的潜在传播率情况下顺序从 1 月 20 日,2020 年 2 月 13 日,2020 年在湖北武汉和其他在中国 28 个省。我们量化了 COVID-19 的**瞬时传播率 (instantaneous transmission)**和**相对繁殖数(R_t) (relative reproduction number)**,并评估了**公共卫生干预是否影响**了各省的病例**传播率**。根据目前的估计,我们已经非常肯定地预测了疾病传播的趋势。

我们估计, R_t 从 4 到 5 降至 1,并保持在 1 以下,而湖北和其他省份在较短的时间内开始增长,然后下降。从 1 月 23 日到 27 日,传播率的比率急剧下降,这可能是由于政府从 2020 年 1 月 23 日开始实施严格的公共卫生干预。感染期平均为 6~9 天。我们预测了 **2020 年 2 月 19 日至 24 日左右各省感染规模趋于稳定的趋势,而武汉的控制日期将推迟一周。(结论)**

解释(预测):在社会和个人层面实施的公共卫生干预措施在预防武汉和其他省份的 COVID-19 暴发方面是有效的。模型预测结果表明, COVID19 将在 2020 年 2 月底左右在中国得到控制。

1. 背景

为了早期评估 COVID-19 暴发的流行潜力以及政府和公共卫生干预的潜在影响,有必要根据时间序列数据实时量化流行病学参数。根据疫情早期的数据,估计武汉地区的**基本繁殖数量 R_0 (basic reproduction number R_0)** 在 2.2 到 4.0 之间[2,3,11,12,13,14]。在实践中,监测传播速率的定量变化和随时间变化的有效繁殖数 R_t ,以揭示控制措施的影响变得至关重要[15、16、17、18、19]。传播性取决于冠状病毒的生物学特性以及可在国家或社会层面干预人群的接触模式。COVID-19 在武汉和中国各省间的传播率动态变化尚不清楚。我们假设随着时间的推移,传播率会显著降低,这与武汉和其他省份的公共卫生干预措施相一致。

这里我们提供一个实时的基于模型的分析来估计趋势遗传性 COVID-19 从 1 月 20 日到 2 月 13 日在武汉,湖北和其他在中国 28 个省和预测转折点达成潜在爆发高原根据监测情况数量以及交通和人口移民数据从武汉到其他城市。我们的方法,不依赖于疾病进展的流行病学参数的假设,这是我们的研究中没有的新病原体,是一个灵活和可推广的方法,直接估计传播趋势的分布

2. 方法

2.1 数据来源

我们从中国国家卫生健康委员会和省级卫生委员会的官方网站(<http://www.nhc.gov.cn>)获得了 2020 年 1 月 20 日至 2020 年 2 月 13 日期间中国 covid19 确诊病例的时间序列数据。包括 23 个时间点的 28 个省(24 个省加上 4 个直辖市, 包括北京、上海、重庆和天津), 以及武汉和湖北主要城市的病例数据。之所以选择 1 月 20 日作为开始日期, 是因为世卫组织于 1 月 17 日发布的官方诊断方案允许对 19 例新病例进行准确和快速诊断。之所以选择 2 月 13 日结束, 是因为武汉和湖北的诊断标准自 2 月 13 日起发生了变化, 将临床诊断的诊断标准包括在内。本研究中所有病例均经国家卫计委病毒核酸检测确认。

武汉通过高速铁路、高速公路和飞机航班与中国其他城市相连。人口流动统计估计大小带来的, 在武汉以外的城市是基于运输相关数据库如下: 1) 铁路和航空旅行数据: 出站的日常数据高速列车从武汉从高速铁路网络获得相应的旅行时间从 12 月 1 日(<http://shike.gaotie.cn>), 2019 年到 2020 年 1 月 23 日。从 2019 年 12 月 1 日至 2020 年 1 月 23 日, 每日出境游航班数量和航空运输小时数均来自 Citytrip 网站(<https://www.ctrip.com>)。我们计算了每天的出行小时数, 等于出境游数量和从武汉到各大城市的铁路和航空运输的出行小时数的乘积。对于一个给定的省份, 我们汇总了该省所有城市的出行总小时数。2) 公路里程数据: 我们从公交站点网络收集公路里程数据, 网址为 <https://www.qichezhan.cn>。它包含高速公路里程从武汉到 16 个城市在湖北。3) 迁移数据: 我们从百度获得了人口迁移数据迁移映射(<http://qianxi.baidu.com>)包括在人口迁移的速度离开武汉到其他省市从 1 月 1 日到 28 日, 2020 年。铁路和航空飞行的总飞行小时数和迁移比例尺由省绘制在图 1 中。本研究采用火车上、飞机上、公路里程和人口迁移量表上的累计时间, 对 2020 年 1 月 20 日本研究 0 时武汉以外省市的潜在疫情规模进行建模。

从图 1 中可以看出, 在所有省份中, 从武汉出发, 通过铁路和飞机出港的时间最多的是广东省。此外, 从武汉移民到河南的人数也是最多的。在湖北省, 黄干和孝干是距离武汉最近的城市, 无论是里程还是移民规模。这些简单的观察结果与我们的结果一致, 即广东省、河南省、黄干县和孝感县在 2020 年 1 月 20 日从武汉输入的原发感染病例估计数量最多。

2.2 建立 COVID-19 的传递性模型

2.2.1 全国 28 个省的模型

Y_{kt}	number of accumulated diagnosed case till day t ,
TR_k	daily traveling hours on trains from Wuhan,
FL_k	daily traveling hours on airplane from Wuhan,
RM_k	highway mileage from Wuhan,
MI_k	volumes of migration from Wuhan from January 1 to 28, 2020,
α_k	number of underlying primary infected cases on January 20, 2020,
W_{kt}	underline number of infected individuals who are infectious,
γ_{kt}	transmission rate defined by $dW_{kt}/W_{k,t-1}$,
η_{kt}	ratio of transmission rate defined by $\gamma_{kt}/\gamma_{k,t-1}$,
m	duration of infectious period (day),

Y_{kt} : 到 t 天 K 市累计的确诊病例数

TR_k : 每天武汉到 K 市的火车班次的时间总数

FL_k : k 市每天从武汉乘飞机流出的时间

RM_k: k 市到武汉的空间距离
 MI_k: 从 1 月 1 日到 2020 年 28 日, K 市来自武汉的移民数量
 α_k : k 市 2020 年 1 月 20 日潜在基础感染病例数
 W_{kt}: k 市第 t 天被感染的潜在病例数
 γ_{kt} : 传输率, $dW_{kt} \setminus W_{k,t-1}$
 η_{kt} : 传输比率, $\gamma_{kt} \setminus \gamma_{k,t-1}$
 m: 传染期的持续时间

k	省份 k
t	第 t 天
Y_{kt}	第 t 天的累计确诊数
TR_k	从武汉出发的每日火车行驶时长
FL_k	从武汉出发的每日航班行驶时长
RM_k	由武汉延伸的高速公路里程
MI_k	从武汉向外的人口迁移量 (2020 年 1 月 1 日至 28 日)
α_k	2020 年 1 月 20 日的基准感染病例
W_{kt}	潜在具有感染性的患病个体数量
γ_{kt}	传播率: $dW_{kt}/W_{k,t-1}$
η_{kt}	传播率的比率: $\gamma_{kt}/\gamma_{k,t-1}$
m	潜伏期 (天数)

下标 k 表示省或市, 下标 t 表示天数。TR_k 和 FL_k 是基于这两个原因构建的。一是人们在火车或飞机上呆的时间越长, 他(她)就越有可能被感染。另一种是感染发生在当地, 因此火车或飞机的数量比乘坐火车或飞机的人口有更多的信息。另外, 在湖北省, 大部分人是开车离开武汉的, 我们用 RM_k 来度量 k 市与武汉的空间距离。

2.2.1 28 个省的模型

首先,我们建立一个指标 α_k 代表 k 省基线感染病例 1 月 20 日,2020 年。特别地,我们将使用 TR_k,FL_k 和 MI_k 来衡量 k 省和武汉之间的关系。我们假设

$$\alpha_k = \beta_1 \times TR_k + \beta_2 \times FL_k + \beta_3 \times MI_k, \text{ for province } k,$$

其中, $\beta = (\beta_1, \beta_2, \beta_3)$ 由 Y_{kt} 的观察值来估计。

到目前为止,我们还不确定影响传播和持久性的关键流行病学参数。因此,我们尽量少做假设。上面定义的符号,很明显, k 省第 t 天的平均新病例是 $\gamma_{kt} W_{kt}$ 。然后,我们假设, k 省第 t 天的平均新病例服从均值为 $\gamma_{kt} W_{kt}$ 的泊松分布

$$dY_{kt} = Y_{kt} - Y_{k,t-1} \sim \text{Possion}(\gamma_{kt} W_{kt}),$$

在中央政府的统一领导下,我们假设对 28 个省来说传输率 γ_{kt} 的趋势是相同的 28 个省,也就是说, $\eta_{kt} = \eta_t$, 是独立于 k 的, 所以

$$\gamma_{kt} = \eta_t \times \gamma_{k,t-1}.$$

为了避免关于流行病演变的假设，我们允许 η_t 是 t 的任意函数。通过逐个点的估计 η_t 和检查超过 t 的结果模式确定了 η_t 的函数形式。 η_t 表示生成的函数形式， $\eta_t = \eta_t(\mathbf{a})$ 。最后，我们注意到

$$dW_{kt} = \gamma_{kt} W_{k,t-1}, \quad W_{kt} = W_{k,t-1} + dW_{kt}.$$

用链式计算，有 $dW_{kt} = \gamma_{kt} \prod_{j=0}^{t-1} (\gamma_{kj} + 1) W_{k0}$ 和 $W_{kt} = \prod_{j=0}^{t-1} (\gamma_{kj} + 1) W_{k0}$, where $W_{k0} = \alpha_k$ and $\gamma_{k0} = 0$.

在实践中，受感染的病人将被隔离并从感染源中移除。用传染期持续时间的 m 表示，我们

$$W_{kt} = \prod_{j=0}^{t-1} (\gamma_{kj} + 1) \alpha_k - I(t > m) \prod_{j=0}^{t-(m+1)} (\gamma_{kj} + 1) \alpha_k.$$

有

令 $\gamma_1 = (\gamma_{11}, \dots, \gamma_{K1})'$ ，所有的参数可以表示成 $\delta = (\gamma'_1, a', \beta')'$ ，loglikelihood（似然函数）可以表示成

$$\begin{aligned} L(\delta) &= \sum_{k=1}^K \sum_{t=1}^T \{dY_{kt} \log(\lambda_{kt}) - \lambda_{kt}\} + C \\ &= \sum_{k=1}^K \sum_{t=1}^T \{dY_{kt} \log[\gamma_{kt} X_k^T \beta \{\prod_{i=1}^{t-1} (1 + \gamma_{ki}) - I(t > m) \prod_{i=1}^{t-m} (1 + \gamma_{ki})\}] \\ &\quad - \gamma_{kt} X_k^T \beta \{\prod_{i=1}^{t-1} (1 + \gamma_{ki}) - I(t > m) \prod_{i=1}^{t-m} (1 + \gamma_{ki})\}\} + C, \end{aligned} \quad (6)$$

C 是独立于 δ 的参数， m 是通过最小化预测误差来确定的。通过 200 次 bootstrap 重复采样得到置信区间[20,21]。

2.2.2 湖北和武汉的模型

湖北的建模和对数似然函数与 28 个省的相似，只是 FL_k 被 RM_k 代替，省被城市代替，因为湖北和武汉之间没有航班，大部分人是乘汽车或公交车离开武汉的。具体地说

$$\alpha_k = \beta_1 \times TR_k + \beta_2 \times RM_k + \beta_3 \times MI_k, \text{ for city } k \text{ in Hubei.}$$

似然函数的建模和武汉和那些对 28 个省是相似的,除了 α_k 是直接通过武汉确诊病例估计。

2.2.3 随时间变化的繁殖数量 R_t (reproduction number) 的计算

When $W_{k,t-1} = 1$, we have $\gamma_{kt} = dW_{kt}$ ，因此 γ_{kt} 是已感染者一天之内传染的新感染者的

平均数。 $\phi_t = \sum_{k=1}^K \gamma_{kt} / K$ 是各省市对应的平均数。（一共有 K 个省市）。因为一个已感染的人可以传染 m 天，他可以传染 $R_t = m\phi_t$ 个人，这确实是随时间变化的繁殖数。

此外，由 Bettencourt 和 Ribeiro(2008)[16]，我们还计算了依赖时间的复制数，即

$$\tilde{R}_t = m \log \left\{ \frac{\sum_k \lambda_{k,t+1}}{\sum_k \lambda_{k,t}} \right\} + 1 \quad \lambda_{k,t} = \gamma_{kt} W_{kt}.$$

2.2.4 预测 NCPI 爆发的潜在转折点

通过最大对数相似度 (6) 估计参数，我们可以估计和预测平均的新病例，

$dW_{kt} = \gamma_{kt} \prod_{j=0}^{t-1} (\gamma_{kj} + 1) \alpha_k$, 然后可计算累计病例 $\tilde{W}_{kt} = \prod_{i=0}^t (\gamma_{ki} + 1) \alpha_k$ 。

根据累积病例, 我们可以预测 COVID-19 暴发的转折点。在本文中, 我们将转折点定义为累

计病例数达到平稳的那一天, 既满足 $|\partial f(v)/\partial v| \leq c_0$, 此时 $f(v) = \frac{\partial \tilde{W}_{kt}/\partial t|_{t=v}}{\partial \tilde{W}_{kt}/\partial t|_{t=v-1}}$ 且 C_0 是一个预定的很小的数, 通过计算我们取 $C_0=2e-03$

3. 结论

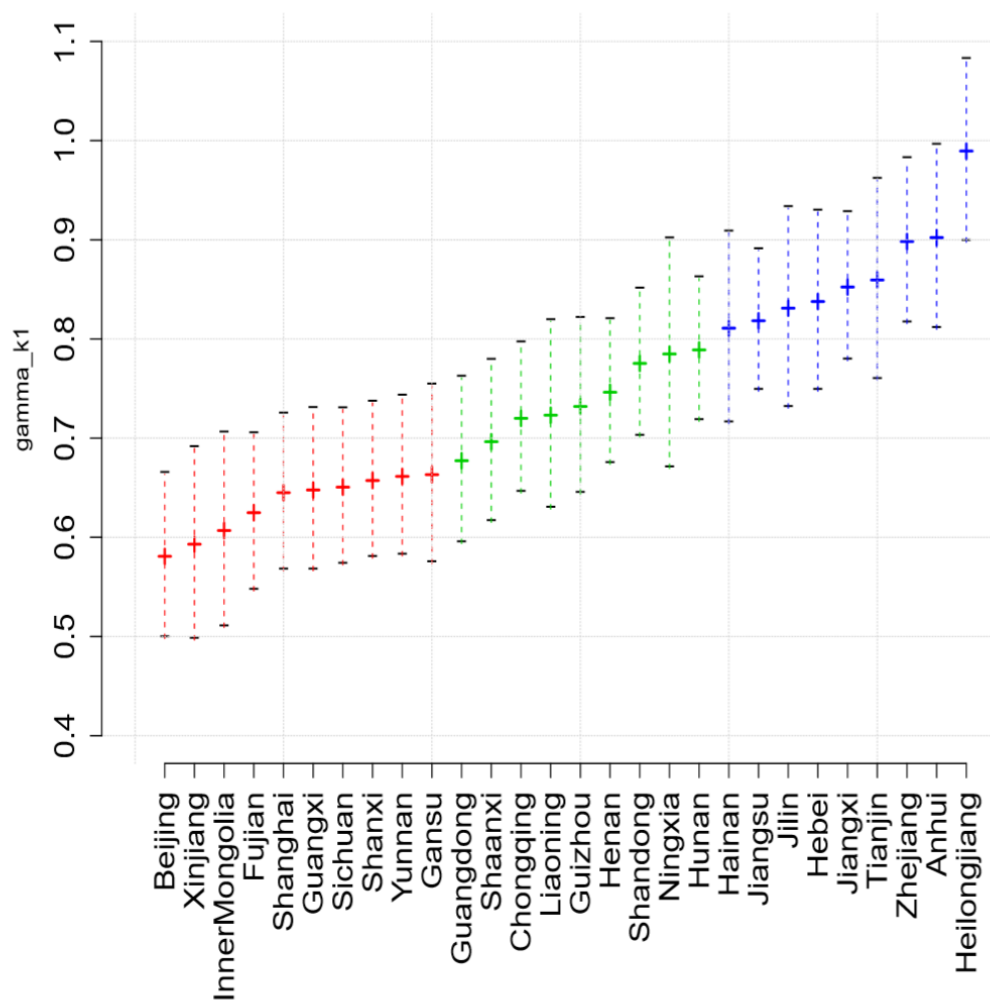
我们使用了一系列的 m 值, 范围从 3 到 23 来适应模型。根据补充图 1(a)、4(a)和 6(a)中预测误差最小的情况, 选择 28 个省和武汉分别为 $m=6$ 和湖北为 $m=9$ 的最优模型, 通过所有 28 个省的补充图 1(b)、4(b)和 6(b), 可得

$$\eta(t) = a_0 + a_1(t - t_1) + a_2(t - t_2)$$

3.1 28 个省的结论

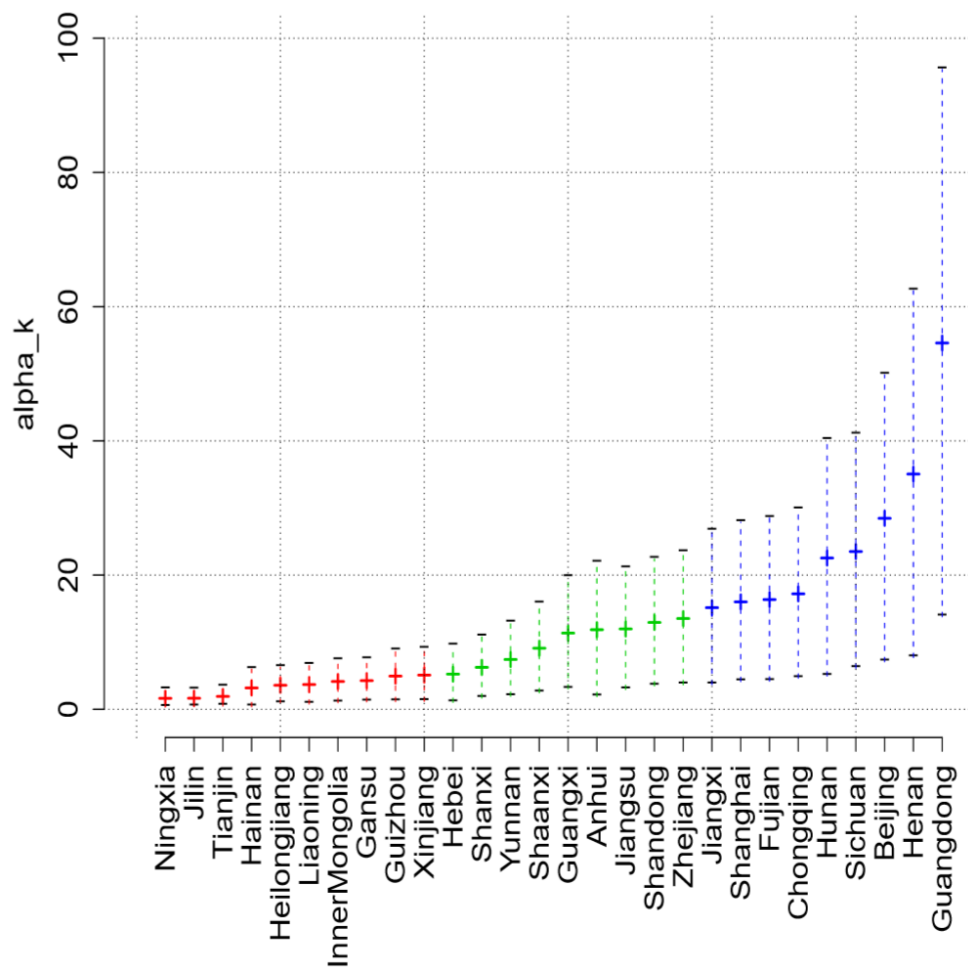
基于 $\eta(t)$ 的函数形式 (上面显示的), t_1 是 1 月 23 日, t_2 是 1 月 27 日, $m=6$, 我们可

以估计 $\delta = (\gamma'_1, a', \beta')'$ 基于 28 个省的 $t=1, \dots, T$, 用第二部分提到的方法。表 1 显示了 β 的估计值。1 月 20 日, 铁路运输和来自武汉的移民对传染病的数量有显著影响, 因此流行病学量表, 但航空运输的系数不显著。



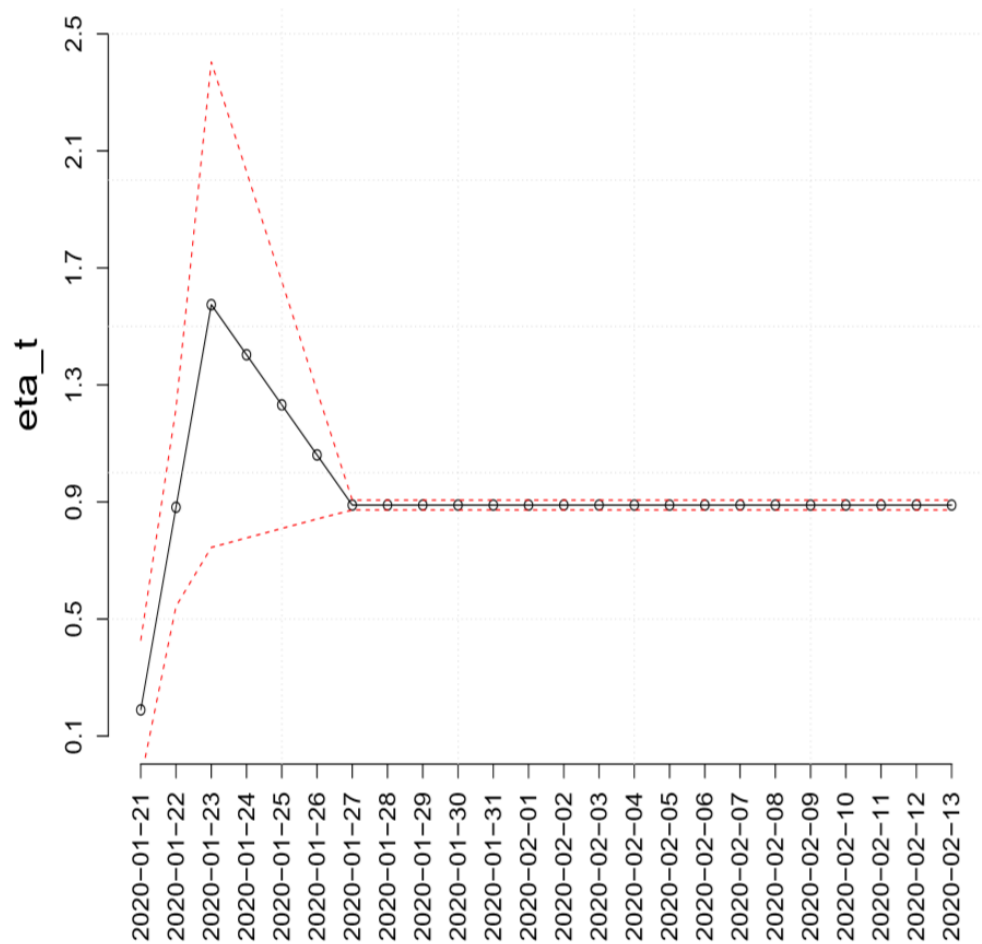
(a)

上图显示了 1 月 20 日 28 个省的 transmission rates γ_{k1} , 因为 $\gamma_{kt} = \gamma_{k1} \prod_{j=2}^t \eta_j$, 并且各省 η_t 相同, 最初的传播率 γ_{k1} 可以用来比较在 28 个省的流行病的预防和控制



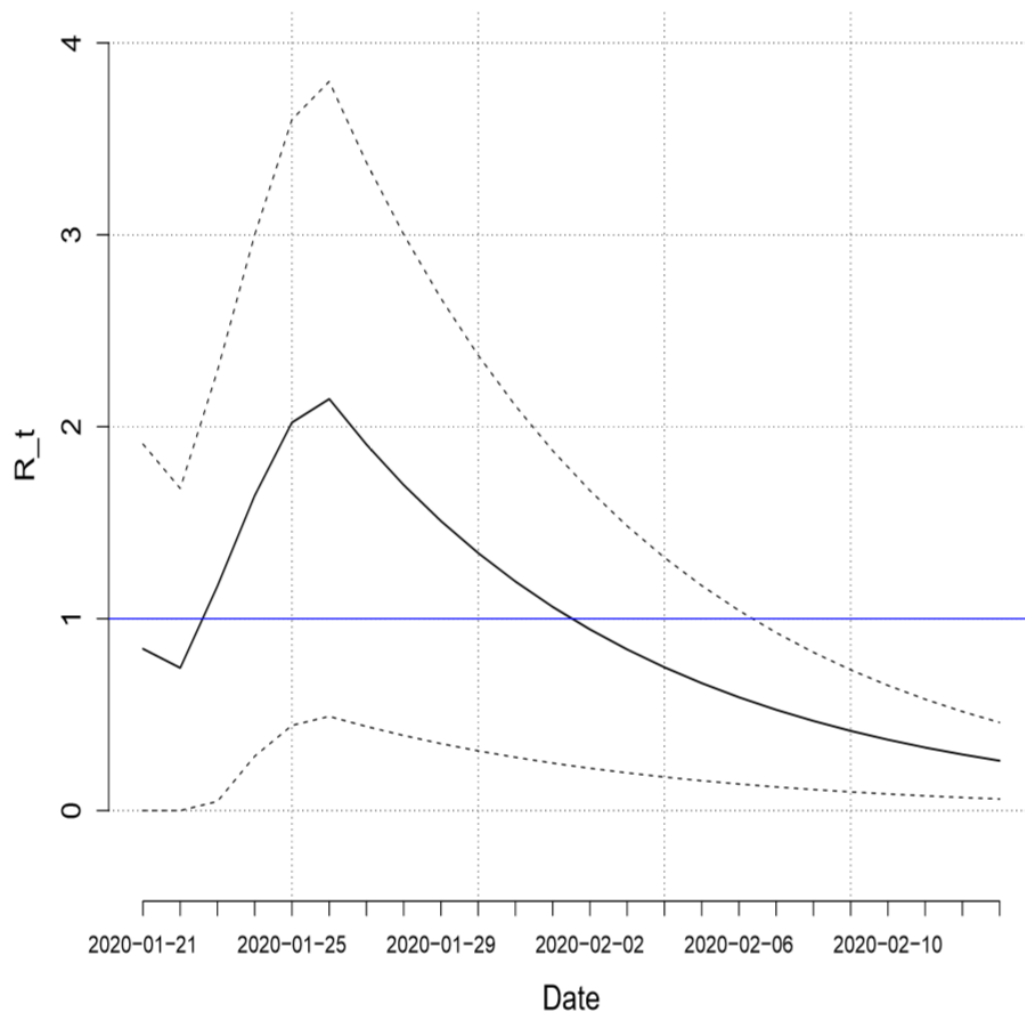
(b)

上图显示了 1 月 20 日潜在的感染病例 α_k 。那些最高的输入性病例 α_k 未必展览 γ_k 传播率最高。例如，北京和四川的潜在病例最多，但传播率较低(图 2(a)中的 0.58 和 0.65);相反，黑龙江和吉林的基础病例最低，但感染率较高。有趣的是，这可能与各省实施的效果相一致;北京和四川以其从 2020 年 1 月 20 日开始的大规模和即时的公共卫生干预而闻名。



(c)

上图显示了随着 t 变化的传输比率 $\eta(t)$,



(d)

图 2(d)显示了 R_t 和 95%置信区间(CI)。繁殖数 1 月 26 日增加到 2.15, 2 月 1 日下降到 1, 2 月 13 日逐渐下降到 0.26。我们也可以计算 \tilde{R}_t 从 Bettencourt and Ribeiro (2008)[16]。 R_t 和 \tilde{R}_t 的形状是一致的, 而 \tilde{R}_t 开始大幅波动。

我们进一步积累病例数, 直到 3 月预测每个省使用上面我们选择的优化模型(图 3 和补充图 2 和 3)。几乎所有省份, 观测值完全属于 95%可信区间的预测区间, 表明一个好的配件, 除了少数病例。例如, 北京的预测值略大于观测值, 这可能是由于基线处对潜在病例的高估。在目前的控制措施没有变化的情况下, 我们观察到所有省份的预测数量在 2 月 19 日到 24 日之间达到了一个高原期(图 3(i)), 在每个图中用蓝色线标记转折点的日期

4. 讨论

在此, 我们使用时间序列数据描述了 2020 年 1 月 20 日至 2 月 13 日在武汉和中国 29 个省暴发的 covid19 疫情的传播动态模式。1 月 21 日至 2 月 2 日, 武汉地区瞬时 R_t 值由 4 ~ 5 降至 1, 而湖北及其他省份在较短时间内呈现先增后降的趋势。从 1 月 23 日到 27 日, 传播率的比率急剧下降, 可能是由于从 2020 年 1 月 23 日开始的严格的公共卫生干预。模型

预测结果显示, 2020 年 2 月 19 日至 24 日, COVID-19 将在各省得到控制, 而武汉的控制日期将推迟一周。

R_t 的时间分布为评价其随时间的演变提供了依据[15,16,22]。湖北和其他省份在 2020 年 1 月 21 日至 27 日期间出现了初期增长, R_t 达到 4, 然后在短时间内下降。 R_t 的增加与变速器的增加一致, 但速度较慢。从 1 月 23 日至 28 日, 传播率的比率有显著下降的趋势。自 2020 年 1 月 23 日起, 中国政府禁止通过航空、铁路和公路进出武汉, 以阻止疫情从武汉蔓延到其他城市。此外, 几乎所有省份在 1 月 23 日和 25 日期间启动了最高级别的公共卫生应急响应, 包括跟踪和隔离与患者的密切接触者、自我隔离等。我们推测, 基于一系列干预效应的巧合, R_t 下降可能是由于自 2020 年 1 月 23 日起实施的一系列降低传播率的干预措施。

对武汉市而言, 虽然 1 月 27 日和 28 日略有上升, 但其生殖数 R_t 总体上呈现下降趋势, 从 2020 年 2 月 2 日的 4.9 下降到 1。与其他地方不同的是, 武汉的疫情可能发生在 1 月初, 数据很少。在我们的研究中, 1 月 21 日的初始 R_t 估计为 4.9, 在此日期之前甚至更高, 为 8.13(数据未显示), 比预期的其他地方的估计值更大。我们还观察到, 1 月 28 日以来, 武汉的传播率比率有所下降, 这表明从 1 月 23 日开始的干预可能对传播率有滞后影响。一些研究估计, 武汉地区的基本繁殖数量 R_0 在 2.2 到 4.0 之间[2,3,11,12], 与 SARS 相比具有高度传染性[13,23,24,25,26,27,28]。[13]的一项研究估计, 12 月 24 日至 1 月 23 日, 武汉 830 例患者的 R_t 值从 8 降至 2.52。这一趋势与我们所观察到的情况相似;然而, 随着相互作用的参与, 我们估计 R_t 在 2020 年 2 月 2 日下降并接近临界阈值 1。

不同省份之间传播率的差异突出了其接触模式和控制措施的重要性, 因为在早期暴发阶段没有获得所需的免疫[4,18]。截至 1 月 20 日, 从武汉输入的潜在病例最多的是广东、河南、北京和四川。北京和四川的传播率最低(从 0.58 到 0.65), 黑龙江、吉林和天津的潜在病例最低, 但传播率高(0.83 到 0.98)。这可能与每个省在减少传播过程中接触人数方面的努力程度有关, 特别是对那些病例最多的人来说具有挑战性。据报道, 对非典型肺炎实施控制的时间高度敏感, 控制措施的实施延迟 1 周导致平均流行规模[16]增加 2.6 倍。因此, 立即采取严格的控制措施对预防疾病传播具有至关重要的作用。

我们预测, COVID-19 将在 2 月 19 日至 24 日左右在各省出现, 一周后在武汉出现。数据拟合的模型对所有区域的拟合效果都很好, 所观察到的病例在图中一般都落在预测界的 95% CI 之内。浙江、贵州两省出现统计异常, 预测误差最大(补充图 3);观察到的病例要么被低估, 要么被高估。这可能是由于感染期持续时间的变化, 或由于我们所掌握的运输数据可能无法完成而导致的基线基础病例估计数量的变化, 这需要进行进一步的研究。

传统的 SIR 或 SEIR 模型[29,30,31,32,33]的估计需要有关疾病进展的流行病学参数的信息, 而在我们的研究中, 新病原体没有这些信息。例如, 研究使用基于 SARS-CoV 先前估计的串行区间估计[2,3,12]。最好从模型本身来估计它们。本文提出的基于概率的方法建模可用于对流行病学参数的实时估计, 且假设量很小。我们假设次要病例是根据泊松分布产生的[22,34]。我们的方法依赖于在感染期间感染病例的潜在统计分布之间的传播推断。它可用于为新病例预测提供统计预期。大量的时间序列数据使我们能够估计每天的比率。因此, 在本研究的计算中, 我们选择对新发传染病使用最小的统计方法, 这是更合适和可推广的。

我们的研究并非没有局限性。首先，我们的结果是基于中国疾病预防控制中心报告的诊断病例。低报很可能发生，这在我们的分析中没有考虑到。截至 2 月 11 日的报告病例中，2020 年 1 月 20 日之前的报告病例仅占 13.8%，其中湖北省[35]的报告病例占 77.6%。由于我们最关心的是 1 月 20 日以后的时间趋势，所以少报率可能不会随着时间变化太大。其次，由于后续阶段实施了一系列干预措施，根据目前的数据，我们无法区分哪些干预措施可能是主要的驱动因素。第三，由于诊断的变化，从 2 月 12 日开始，武汉和湖北的确诊病例急剧增加至 14840 例，这使得我们的预测具有挑战性。因此，我们没有提供 2020 年 2 月 13 日后武汉和湖北的预测病例。

综上所述，我们利用时间序列数据提供了 2020 年 1 月 20 日至 2 月 13 日在中国暴发暴发的传播动态模式。 R_t 的下降趋势以及传播率的比率显示了政府从 2020 年 1 月 23 日开始实施的公共卫生干预的效果。模型预测结果显示，2020 年 2 月 19 ~ 24 日左右，COVID-19 将在全省范围内得到控制，而武汉的控制日期将推迟一周。

中文报告

1、研究目的

估计：新冠肺炎在各地的实时传播强度

预测：新冠肺炎在全国各地和是达到稳定可控的状态

量化：政府的干预措施对疫情发展的影响

2、数据

新冠肺炎数据

来源：国家和各地文件微官网

时间范围：1 月 20 日至 2 月 13 日（13 号后加了临床检测）

地点：24 个省，4 个直辖市，湖北省各市（16），武汉

变量：每日新增确诊人数

交通数据

铁路和航空数据：每日旅行时间=武汉到每个主要城市的出境旅行趟数*铁路和航空运输的旅行时间

公路里程数据：从武汉到湖北 16 个城市的高速公路里程

人口迁移数据：2020 年 1 月 1 日到 28 日离开武汉到其他城市和省份的人口迁移率（百度云上的人口迁移）

3、模型方法

αk 是未知的

η_{kt} 是要估计出来的，这里引入 η_{kt} 是因为它可以当成是 y_{kt} 的导数，一般导数比原函数简单的函数，更好估计，如果直接对传播率估计可能是不稳定的

η_t 是一个变化的变化

● 对1月20日的基准感染病例 α_k 建模

对于每个省份 k ，有

$$\alpha_k = \beta_0 + \beta_1 \times TR_k + \beta_2 \times FL_k + \beta_3 \times MI_k \quad (1)$$

其中， $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)'$ 由观察到的第 t 天的累计确诊数 Y_{kt} 估计得到， TR_k 表示从武汉出发的每日火车行驶时长， FL_k 表示从武汉出发的每日航班行驶时长， MI_k 表示从武汉向外的人口迁移量（2020年1月1日至28日）， $k = 1, \dots, K$ ， $t = 1, \dots, T$ 。

● 新增病例数 dY_{kt} 的分布假定

很明显，省份 k 在第 t 天的平均新增病例与前一天具有感染性的患病个体数量 $(W_{k,t-1})$ 有关，我们假设省份 k 在第 t 天的新增病例服从均值为 $\gamma_{kt}W_{k,t-1}$ 的泊松分布，

$$dY_{kt} = Y_{kt} - Y_{k,t-1} \sim \text{Poisson}(\gamma_{kt}W_{k,t-1}) \quad (2)$$

● 传播率 (γ_{kt}) 建模

由于中央统一部署，各省传播率的变化基本是同步的。因此假设，

$$\gamma_{kt} = \eta_t \times \gamma_{k,t-1} \quad (3)$$

为避免假设过强，我们允许 η_t 是关于 t 的任意函数，将通过数据给出。

● 具有感染性的患病个体数量 W_{kt} 的建模 (1)

最后，我们注意到 W_{kt} 有下面的形式，

$$dW_{kt} = \gamma_{kt}W_{k,t-1}, \quad W_{kt} = W_{k,t-1} + dW_{kt} \quad (4)$$

在链式计算后，可以发现

$$dW_{kt} = \gamma_{kt} \prod_{j=0}^{t-1} (\gamma_{kj} + 1) W_{k0}, \quad W_{kt} = \prod_{j=0}^{t-1} (\gamma_{kj} + 1) W_{k0}$$

其中， $W_{k0} = \alpha_k$ （1月20日基准感染病例）， $\gamma_{k0} = 0$ 。

● 对数似然函数表达式

整合式(1) (2) (3) (5)，得到对数似然函数如下：

$$\begin{aligned} L(\delta) &= \sum_{k=1}^K \sum_{t=1}^T \{dY_{kt} \log(\lambda_{kt}) - \lambda_{kt}\} + C \\ &= \sum_{k=1}^K \sum_{t=1}^T \left\{ dY_{kt} \log \left[\gamma_{kt} X_k^T \beta \left\{ \prod_{i=1}^{t-1} (1 + \gamma_{ki}) - I(t > m) \prod_{i=1}^{t-m} (1 + \gamma_{ki}) \right\} \right] \right. \\ &\quad \left. - \gamma_{kt} X_k^T \beta \left\{ \prod_{i=1}^{t-1} (1 + \gamma_{ki}) - I(t > m) \prod_{i=1}^{t-m} (1 + \gamma_{ki}) \right\} \right\} + C \quad (6) \end{aligned}$$

其中， $\gamma_1 = (\gamma_{11}, \dots, \gamma_{K1})'$ ，所有参数表示为 $\delta = (\gamma_1', a', \beta')$ ， C 是独立于 δ 的常数， m （潜伏期）由最小化预测误差来决定。

- 对1月20日的基准感染病例 α_k 建模

对于湖北省各市和武汉市1月20日的基准感染病例，有

$$\alpha_k = \beta_1 \times TR_k + \beta_2 \times RM_k + \beta_3 \times MI_k \quad (7)$$

其中， TR_k 表示从武汉出发的每日火车行驶时长， RM_k 表示由武汉延伸的高速公路里程， MI_k 表示从武汉向外的人口迁移量（2020年1月1日至28日）， $k = 1, \dots, K$ ， $t = 1, \dots, T$ 。

- 除了 α_k 是直接用武汉的诊断病例估计之外，湖北省各市的建模过程和对数似然函数都与全国其他省的建模类似。

- 注意到 $dW_{kt} = \gamma_{kt} W_{k,t-1}$ 。因此当 $W_{k,t-1} = 1$ 时，第二天新增病人数有 $dW_{kt} \triangleq \gamma_{kt}$ 。因此 γ_{kt} 是一天之内传染性个体所能感染人数，则 $\phi_t = \sum_{k=1}^K \gamma_{kt} / K$ 对应于省市的平均值。

- 由于在潜伏期 m 天内，传染者都能感染他人，那么传染者就能感染 $R_t = m\phi_t$ 个新患者， R_t 就称为与时间相关的再生数。

- 根据Bettencourt and Ribeiro (2008)，我们计算了 \tilde{R}_t

$$\tilde{R}_t = m \log \left\{ \frac{\sum_k \lambda_{k,t+1}}{\sum_k \lambda_{k,t}} \right\} + 1$$

其中， $\lambda_{k,t} = \gamma_{kt} W_{kt}$

- 最大化(6)式的对数似然值后，我们可以估计和预测平均的新病例为

$$dW_{kt} = \gamma_{kt} \prod_{j=0}^{t-1} (\gamma_{kj} + 1) \alpha_k$$

那么累计病例 $\tilde{W}_{kt} = \prod_{j=0}^t (\gamma_{kj} + 1) \alpha_k$

- 拐点是指累计病例数达到平稳期的那天，满足

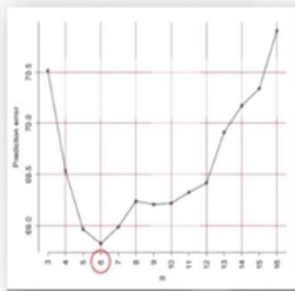
$$|\partial f(v) / \partial v| \leq c_0$$

其中， $f(v) = \frac{\partial \tilde{W}_{kt} / \partial t |_{t=v}}{\partial \tilde{W}_{kt} / \partial t |_{t=v-1}}$ ， c_0 是给定的较小值，这里我们令 $c_0 = 2e - 03$

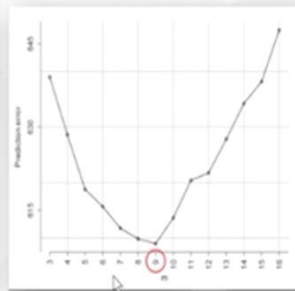
● m 的选择

对于28个省和武汉市, $m = 6$

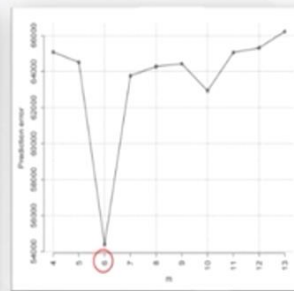
对于湖北省其他市, $m = 9$



(a) 28个省



(b) 湖北省各市

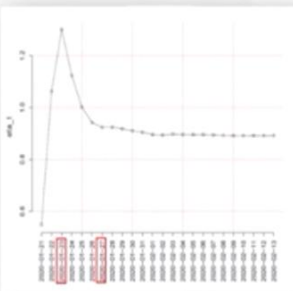


(c) 武汉市

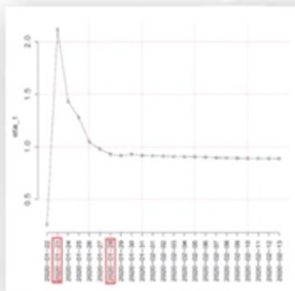
● t_1, t_2 的选择

$$\eta(t) = a_0 + a_1(t - t_1)_- + a_2(t - t_2)_-$$

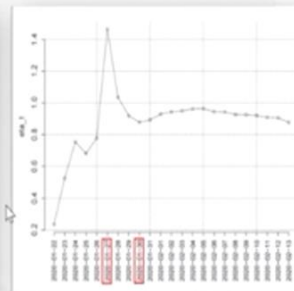
其中, t_1 和 t_2 在28个省、湖北省各市和武汉市是不同的, $t_- = \min(t, 0)$



(a) 28个省



(b) 湖北省各市



(c) 武汉市

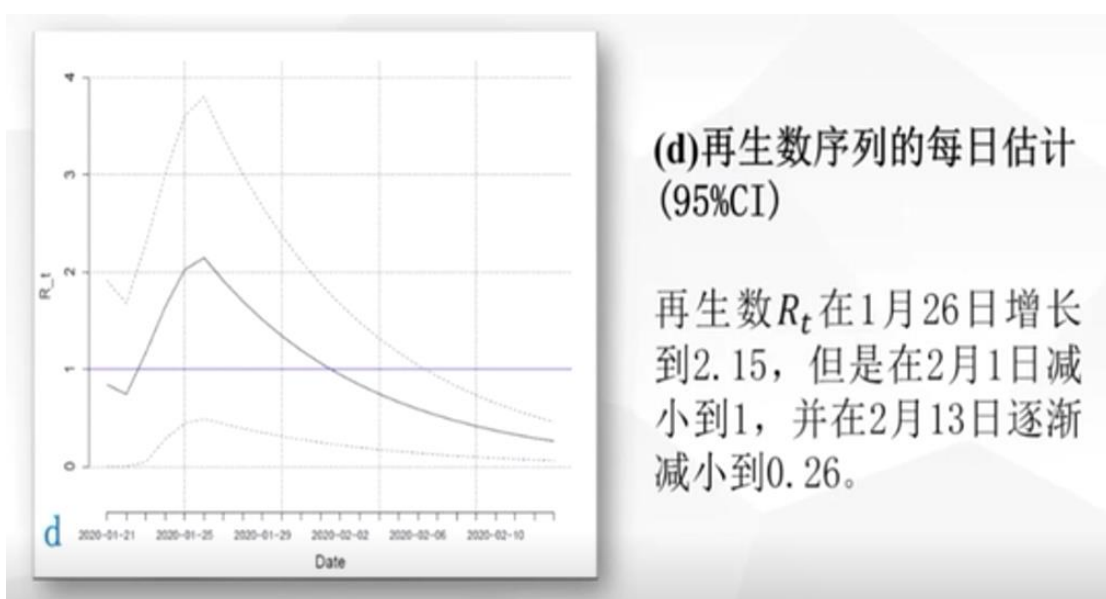
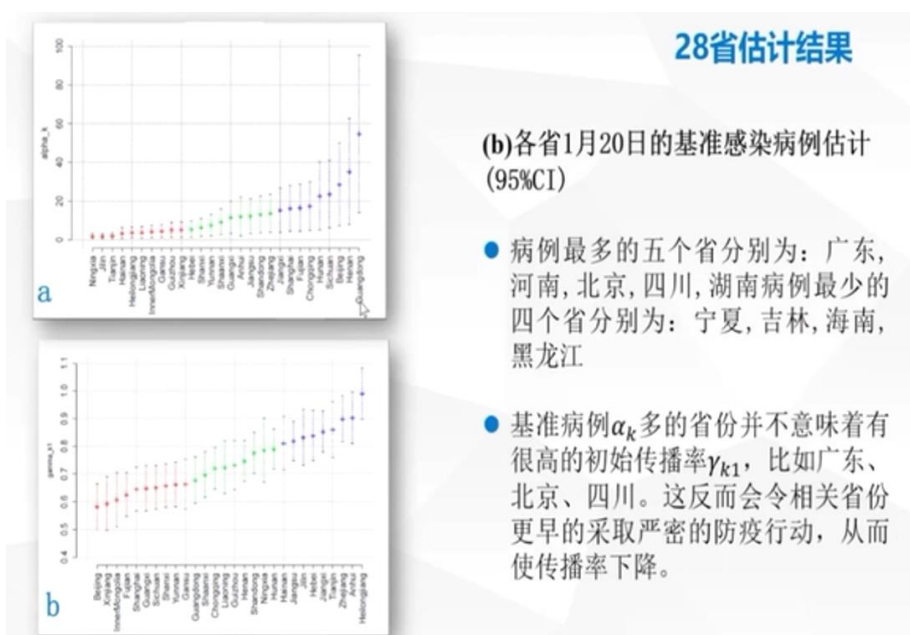
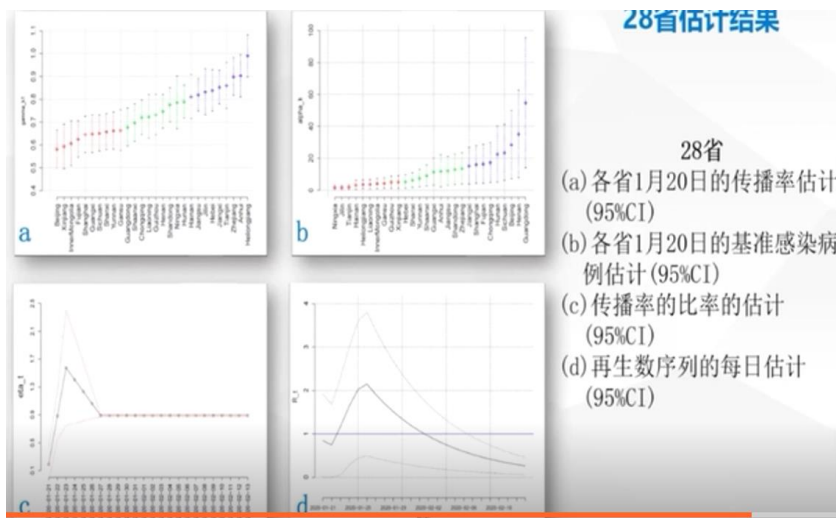
表: 交通方式和人口迁移的影响对基准感染情况的估计

	28省			湖北16市		
	train	airplane	migration	train	highway	migration
Est.	42.628	1e-05	28.740	1e-05	7.781	20.908
SD	17.498	0.853	12.321	0.106	3.719	5.967
p-value	0.015	1.000	0.0197	0.999	0.036	4e-04

● 湖北省内的传染主要受公路和人口迁移的影响;

● 省际传染主要受火车和人口迁移的影响。

这说明了武汉附近城市的离汉方式可能主要是乘坐汽车。



- 政府干预措施起到了积极效应
 - 对于湖北省内和其他省份
从1月21日至27日, R_t 最初上升到4, 但随后在短时间内下降;
从1月23日至28日, η_{kt} 急剧下降。
 - 对于武汉市
 R_t 整体呈下降趋势。
- 我们预测COVID-19在各省内于2月19日至24日之间被控制住, 而武汉会晚一周。
- 传统的SIR或SEIR模型需要流行病学的详细信息, 我们的模型不需要这些信息也可以进行估计预测, 我们的方法所依赖的假设较少。