

中国新冠肺炎疫情预测建模与理性评估

金启轩

(武汉大学 弘毅学堂, 武汉 430072)

摘要：文章针对当前各界关于疫情前景预测研究存在的局限性,结合中国政府对于新型冠状病毒感染的肺炎的疫情处置进展,以及医学界对于COVID-19研究的最新成果,通过建立科学的数学模型对中国(除湖北省和港澳台地区)30个省级行政区的疫情发展状况予以预测,对疫情的规模进行定量分析,对疫情原始基数和有效传播率进行科学和可靠的区间估算,从而对疫情发展状况作出切合实际且理性的判断,为疫情防控提供参考。

关键词：新型冠状病毒;COVID-19;数学模型;有效传播率

中图分类号：R174

文献标识码：A

文章编号：1002-6487(2020)05-0011-04

0 引言

自新冠肺炎暴发以来,人们对于引发疫情的2019新型冠状病毒(COVID-19)的科学认知逐步加深,应对手段逐步调整,信息掌控逐步丰富。在此发展过程中,科学界和社会各界纷纷对疫情前景开展了预测研究,比如:伦敦帝国理工学院根据截至2020年1月24日的数据,在其官网上发布了统计得出的基本传染数中位数2.6,95%置信区间为(2.1,3.5)。世界卫生组织和香港大学的研究者总结了自2019年底疫情成熟开始的两个月内的确诊病例数据,将 R_0 系数提升至2.68,95%置信区间则被缩小到2.47~2.86^[1]。1月23日,英国兰开斯特大学、格拉斯哥大学及美国佛罗里达大学推出论文,如果疫情不受控制,估计2月4日武汉市感染人数将会在13万人至27万人之间^[2]。香港大学李嘉诚医学院院长梁卓伟与研究团队于2月1日在医学期刊《柳叶刀》上发表报告,以模型推算1月25日前,武汉已经有75800多人感染新型肺炎,其后传播至重庆、北京、上海、广州和深圳;若病毒传播能力不减,武汉疫情将在4月到达高峰,中国多个主要城市可能持续有本地暴发,会随武汉高峰期后的一、二个星期进入高峰^[3]。

2月6日,世界卫生组织发言人贾萨瑞维奇接受了环球时报-环球网记者采访,他认为现阶段尚无法预测中国国内新冠肺炎疫情的“拐点”何时到来,也无法预测这个疫情是否会在中国以外大规模暴发,从而演变成全球性流行病,因为仍有大量关键信息未能为世界科学界所掌握^[4]。

以上预测与估算,大都基于或滞后或片面的信息,缺乏可靠的有针对性的模型基础,有的对于数据的处理并没有与实际情况进行对比分析与矫正。故各种预测结论不但相差巨大甚至相互矛盾。这导致社会民众对于政府处理疫情的能力产生了怀疑。

由于疫情前期诊断流程冗长,核酸检验试剂盒严重不足,以及人手、设备和场地不足,每天确诊的人数受限,所

以,湖北省的确诊人数数据反映的仅是检测能力,而非真实的疫情发展情况。因此,湖北省的“数据淤积”是医学能力和诊断标准的问题,本文针对相关预测估算中的缺陷,选择中国30个省份(剔除湖北省以及港澳台地区)的数据,综合国家对疫情的处置以及对疫病的研究进展予以建模,预测比较符合实际的疫情发展状况,从而为疫情处理提供有价值的参考,消除民众的恐慌心理。

1 疫情评估

以2020年1月23日武汉“封城”为初始节点,对数据进行收集和分析,最终估算1月23日的感染数及分布(中国30省份),以及1月23日前一周开始的各个阶段有效传播率,进而评估疫情的发展规模(结束时间及疫情人数)。

1.1 数据收集

(1)国家卫健委和各省卫健委公开发布的每日确诊疫情数据,见下页表1。

(2)医学界根据前期病例以及mers推断潜伏期最长14天,一般3~7天。

(3)据武汉协和医院的初期临床数据,病人从感染到住院平均时间为9天。

(4)确诊数据公布时间与感染时间的滞后差值:潜伏期+就诊时间+确诊时间+统计时间,后三项前期值较大(如确诊时间前期需要3天以上,后期缩短到1天),后期逐渐减小。

(5)武汉1400万人口,1月23日前流出500万人,“封城”时有900万人。

1.2 模型及处理

(1)中国30省份疫情的性质

疫情初期(暴发期),一般使用指数模型进行评估,但使用指数模型有个最基本的前提,即初始传播基数明确。然而在本次疫情中,由于中间宿主和传播途径尚未确定,对传播方式的研究也未取得确切的量化成果。也就是说,

表1 每日确诊疫情数据 (单位:人)

日期	全国	累计	湖北	累计	武汉	累计	30省份	累计
2.12	15152	60362	14840	48293	13436	32975	312	12609
2.11	2015	45210	1638	33453	1104	19521	377	11757
2.10	2478	43195	2097	31815	1552	18417	381	11380
2.9	3062	40717	2618	29718	1921	16865	444	10999
2.8	2656	37655	2147	27100	1379	14994	509	10555
2.7	3399	34999	2841	24953	1985	13615	558	10046
2.6	3143	31600	2447	22112	1501	11630	696	9488
2.5	3694	28457	2987	19665	1766	10129	707	8792
2.4	3887	24763	3156	16678	1967	8363	731	8085
2.3	3235	20876	2345	13522	1242	6396	890	7354
2.2	2829	17641	2103	11177	1033	5154	726	6464
2.1	2590	14812	1921	9074	894	4121	669	5738
1.31	2102	12222	1347	7153	576	3227	755	5069
1.30	1982	10120	1220	5806	378	2651	762	4314
1.29	1737	8138	1032	4586	356	2273	705	3552
1.28	1459	6401	840	3554	315	1917	619	2847
1.27	1771	4942	1291	2714	892	1602	480	2228
1.26	769	3171	371	1423	80	710	398	1748
1.25	688	2402	323	1052	46	630	365	1350
1.24	444	1714	180	729	77	584	264	985
1.23	259	1270	105	549	70	507	154	721
1.22	571	1011	69	444	62	437	502	567
1.21	149	440	105	375	105	375	44	65
1.20前	291	291	270	270	270	270	21	21

目前的研究,无法提供可用于评估本次疫情的初始传播基数的有效模型和数据,建立在该基础之上的指数模型(无论是静态的还是动态的),可信度较低。由此计算出来的有效传播率和传播周期以及疫情规模,都缺乏足够的依据和可靠度。

对数据观察发现,1月23日前,湖北省之外的省份人传人现象并不“严重”(显见的有效传播率小于1),甚至国外尚未见人传人,该现象亦为世界卫生组织所发现。即便是人传人,大都局限在家庭成员或密切接触的人之间。

考虑到1月23日之后,武汉市及湖北省内的各市先后“封城”,整个国家对疫情高度警惕,这会极大地限制病毒的传播。初步判断,1月23日之后,疫情的有效传播率极大概率的是小于1的,即在传播学上该疫情是处于被控制的状态而非处于暴发扩张状态。为证明该判断,本文对1月23日之后的增长率进行了统计和建模如下:

图1与表2为中国30省份原始数据,选择中国30省份的数据其原因在于:数据可靠、及时;只有人传人;初始基数是确定的(即各省的输入病例之和)。

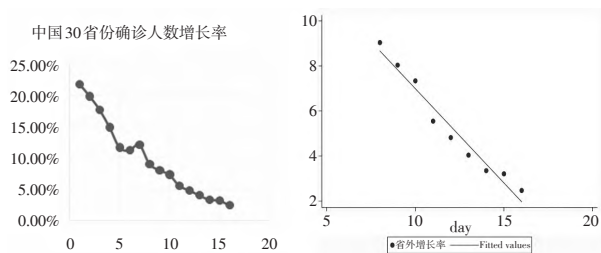


图1 数据拟合

据悉,中国政府于1月25日召开会议,成立了应对疫情的领导小组并向湖北等疫情严重地区派出指导组,推动有关地方的防控一线工作。虽然该重大举措强力推进,但

表2 中国30省份原始数据

日期	模型天数t	总计(人)	确诊日增(人)	增长率(%)
1.21		65	44	67.69
1.22		567	502	88.54
1.23		721	154	21.36
1.24		985	264	26.80
1.25		1350	365	27.04
1.26		1748	398	22.77
1.27		2228	480	21.54
1.28	1	2847	619	21.74
1.29	2	3552	705	19.85
1.30	3	4314	762	17.66
1.31	4	5069	755	14.89
2.10	5	5738	669	11.66
2.2	6	6464	726	11.23
2.3	7	7354	890	12.10
2.4	8	8085	731	9.04
2.5	9	8792	707	8.04
2.6	10	9488	696	7.34
2.7	11	10046	558	5.55
2.8	12	10555	509	4.82
2.9	13	10999	444	4.04
2.10	14	11380	381	3.35
2.11	15	11757	377	3.21
2.12	16	12609	312	2.47

注:起点为1月28日(记为第1天),其中:增长率=增长人数/确诊人数。

实际效果则显现在两天之后。从表2可看出,自1月28日起,确诊人数的增长率呈较为规律的线性下降趋势。

为使模型更准确地预测将来的情况,本文选取1月28日为第1天,对表2中从2020年1月28日开始至2020年2月12日的数据进行分析。图1中第8天后的增长率变化趋势显然与第8天前不同,故本文从第8天即2月4日进行数据拟合,令天数为t,得到的线性回归方程为:

$$\text{增长率} = 15.37\% - 0.84\%t$$

模型拟合优度 R^2 为 96%, 高度拟合,说明这一系列数据呈规律的线性下降趋势。

令增长率为0,可得 $t=18$ (天),天数项 95% 的置信区间为 $(-0.98, -0.69)$,常数项的区间为 $(13.57, 17.17)$ 。可以得出增长率为0时天数最大区间为 $(14, 25)$ 。

计算结果表明中国30省份确诊日增长率大概率在2月21日左右趋近于0。

为提高模型的拟合优度,尝试用时间序列中的3日加权平均移动法进行简单的处理,见下页表3。

再将表3中得到加权平均的数据制得散点图,见图2。

模型拟合优度 R^2 为 96%, 高度拟合,说明这一系列数据呈规律的线性下降趋势。

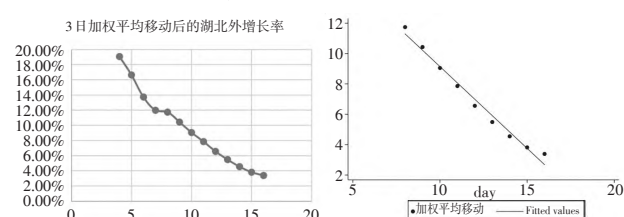


图2 三日平均移动后的中国30省份增长率及拟合

可以明显看到,自第8日起,数据的线性相比之前增强了,再次进行最小二乘法估计得到:

表3 加权平均移动

日期	天数t	中国30省份增长率(%)	3日加权平均移动(%)
1.28	1	21.74	
1.29	2	19.85	
1.30	3	17.66	
1.31	4	14.89	19.07
2.10	5	11.66	16.64
2.2	6	11.23	13.74
2.3	7	12.10	11.98
2.4	8	9.04	11.74
2.5	9	8.04	10.43
2.6	10	7.34	9.05
2.7	11	5.55	7.86
2.8	12	4.82	6.56
2.9	13	4.04	5.49
2.10	14	3.35	4.55
2.11	15	3.21	3.82
2.12	16	2.47	3.39

注:用 g 代表增长率,加权平均移动 $g_4=(3g_3+2g_2+g_1)/(3+2+1)$ 。

增长率=20.00-1.10%t

模型拟合优度 R^2 为98%,高度拟合,说明这一系列数据更加呈规律的线性下降趋势。

令增长率为0,可得 $t=18$ (天),天数项95%的置信区间为(-1.20,-0.95),常数项的区间为(18.32,21.49)。可以得出增长率为0时天数最大区间为(15,23)。

计算结果表明中国30省份确诊日增长率大概率在2月19日左右趋近于0。

注意到加权平均移动法对季节性十分敏感,而本组数据本身不受季节变化的影响。但可以发现,通过三日加权平均法得到的许多个数据比真实数据要大,这可能导致了此分析方法整体预期的前移,所以增长率可能在2月19日左右趋近于0。

考虑到数据的延后性(平均5~11天)即本次疫情于1月23日之后基本被控制在非暴发性扩张状态(有效传播率小于1)。欧、美、港专家所做的指数型扩张模型(有效传播率大于1)是不大可靠的。

(2)中国30省份疫情的规模

对疫情的规模进行定量分析,需要可靠的模型和数据。如上文所述,中国30省份数据初始基数是确定的(即各省份的输入病例之和),只有人传人情况,数据可靠、及时,样本足够大,分布足够广泛,非常适合做可靠的疫情规模的定量分析。

①日增确诊数变化

原始数据见图3,考虑到在中国30省份内人传人主要是二代病例,三代病例较少,四代较为罕见,本文对中国30省份日增确诊数分别用一次和二次函数进行分段拟合处理(见图4、图5)。

一次线性形式:日增确诊人数 $=-59t+1235.5$

拟合优度为 R^2 为95.90%,高度拟合。天数 ≈ 21 时,中国30省份新增确诊人数趋近于0值。

二次函数形式:日增确诊人数 $=3.7t^2-151.9t+1796.3$ (无零点)

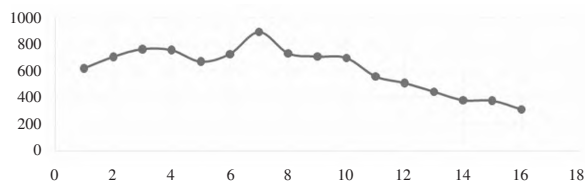


图3 中国30省份日新增确诊人数

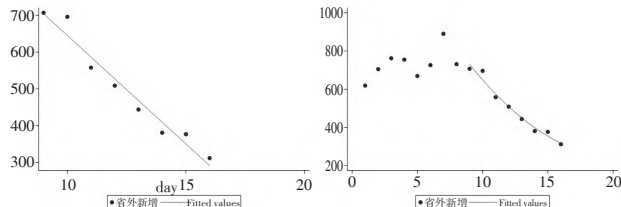


图4 一次拟合

图5 二次拟合

拟合优度为 R^2 为97.42%,数据有可能更符合此二次模型变化趋势。计算其最低点:天数=21

分析计算结果得:中国30省份新增确诊人数可能于第21天(2月17日)趋近于0水平,中国30省份总确诊人数可能趋近于顶峰。

②原始基数和有效传播率估算

统计了部分省份的输入病例与本地感染病例的分布后发现:截至2月3日(当日确诊病例中输入病例与本地感染者各占一半),北京、天津、上海、辽宁、陕西、江苏6省份,共1006人确诊,其中本地感染者不超过约329人(误差10%以内),即本地传染占比约33%。以此为样本估算中国30省份的本地传染者,约2418人,则可计算出1月23日中国30省份原始基数(即湖北输出的病例,1月23日后,湖北基本停止输出人员,也就停止输出病例)约为 $7354-2418=4936$ 。直到2月6日,依然存在确诊输入病例,因此该原始基数的值应略大于4936,即在4936~5150人之间。

根据拟合的日增长确诊人数函数,可以估算本次疫情中国30省份的总感染数是小于14000的,2月17日前后确诊病例可能不会再大幅增长,由此本文估算8日周期等价有效传播率不大于0.62,且4日周期等价有效传播率不大于0.31。

2 分析结果

2.1 疫情性质及原因

1月23日之后,日增长率以 $y=kx+b$ 形式作为模型可高度拟合,并迅速下降。 k 值可理解为国家和整个社会对于病毒的防控力度。防控力度越大, k 的绝对值越大,直线会更快地达到0点,即确诊人数会更快达到峰值。

社会防控病毒力度的影响因素:当地大型医院数、医院床位数、医护人员数量、社会管理人员活动的严格程度、民众的防疫意识等。

国外对疫情的定性之所以与事实相差巨大,源于其对数据的处理存在信息不对称,国外专家不清楚感染基数是多少,所以会建立错误的指数扩张模型,以至于夸大了病

毒的传播速度(夸张到1~3天倍增)和传播能力,忽视了我国政府和人民的防控能力。

计算结果表明中国30省份确诊日增长率大概率在2月21日左右趋近于0,可以说我国政府和民众对疫情的处理非常高效。

2.2 疫情的发展和规模

1月23日之后,中国30省份的日新增感染数在极短的时间(3~7天)内触顶开始下降,下降趋势可以用一次和二次多项式高度拟合,表明本次疫情中国30省份的以二代感染者为主,考虑到二次拟合的拟合度高于一次拟合,说明存在三代及以上感染者,但占比较少。与本文先前的观察一致。说明我国在阻止疫情传播方面非常成功,有效地阻断了病毒的多代传播。

本次疫情的有效传播率小于0.62(8日周期等价值),大大低于国内外专家估算的基础传播率,一方面说明国内外专家基于基础传播率来估算本次疫情的发展规模是不可靠的,因其基础传播率的估算本身缺乏足够可靠的科学依据。另一方面,也说明我国政府和人民投入巨大的人力物力所取得的效果是极其出色的,用外媒预估三分之一左右的时间,将疫情基本控制住。

截止本文(2月15日)完成时,湖北省(不含武汉)累计数与中国30省份累计数之比(1月27日)与1月23日湖北省(不含武汉)原始基数与中国30省份原始基数之比(1月30日)非常接近(误差不到3%),说明湖北省的疫情发展(有效传播率)与全国大体一致。

2.3 严控疫情

日新增感染数与日新增确诊数大约存在5~11天的延迟,前期延迟较大,后期延迟较小,接近5~6天,也就是说,中国30省份的疫情,实际上在2月12日已经被基本控制住(绝大部分的省份已经降至日感染峰值的5%~10%以下)。需要着重说明的是,社会学和医学上的控制住,并不

意味着疫情彻底消除,人们可以放松防疫,相反,越是最后阶段,越要防止疫情卷土重来。

3 结论

截至1月23日,中国30省份疫情有效传播率(8日周期等价)小于0.62;1月23日武汉感染人数为12696人,95%置信区间为(10157,15235),湖北省(不含武汉)感染人数为6389人,中国30省份感染人数为4896人,95%置信区间为(4896,5150)。

中国30省份于2月12日以后,疫情传播基本被控制住,湖北省2月19日后疫情的传播基本被控制住,武汉市则于2月19日至2月26日间基本控制住疫情传播(确诊数据比实际感染数据晚至少5天)。

在疫情传播被控制住后的15天,在采取严格防控措施的手段下,各省市可逐步恢复正常的工作,这样风险较低。

参考文献:

- [1] Wu J T, L et al. Now Casting and Forecasting the Potential Domestic and International Spread of the 2019-nCoV Outbreak Originating in Wuhan, China: A Modelling Study. The Lancet[EB/OL]. [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(20\)30260-9/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30260-9/fulltext), 2020.
- [2] Read J M, Bridgen J R E, Cummings D A T, et al. Novel Coronavirus 2019-nCoV: Early Estimation of Epidemiological Parameters and Epidemic Predictions [R] MedRxiv, Version 2. Updated 27 Jan 2020.
- [3] 香港大学.推算武汉农历年前已逾7.5万新型肺炎个案[N]. Now新闻. 2020-02-02.
- [4] 白云怡,陈青青.世卫组织:尚不可能预测“拐点”何时到来[N]. 环球时报, 2020-02-07.

(责任编辑/易永生)

Novel Coronavirus Pneumonia Epidemic Prediction Modeling and Rational Evaluation in China

Jin Qixuan

(Hongyi Honor College, Wuhan University, Wuhan 430072, China)

Abstract: In view of the limitations of the current research on the prediction of the prospect of epidemic situation, this paper combines with the Chinese government's progress in the outbreak management of novel coronavirus pneumonia and the latest research achievements on COVID-19 virus in the medical community to establish a scientific mathematical model, by use of which, the paper predicts the epidemic situation of 30 provincial administrative regions in China (except Hubei Province and Hong Kong, Macao and Taiwan), quantitatively analyzes the scale of the epidemic outbreak, and makes a scientific and reliable interval estimates on the original baseline and effective propagation rate of the outbreak. Thus the paper makes a realistic and rational judgment on the development of the epidemic, and provides reference for the prevention and control of the epidemic.

Key words: novel coronavirus; COVID-19; mathematical model; effective propagation rate