



传染病信息学概论

主编 宋宏彬 郝荣章 曾大军



科学出版社

国家科学技术学术著作出版基金资助出版

传染病信息学概论

主 编 宋宏彬 郝荣章 曾大军

科学出版社

北 京

内 容 简 介

本书概述了传染病信息学框架、体系和相关技术基础,详细阐述了传染病时空数据分析,传播动力学模型方法,复杂网络中传染病传播建模的原理、方法和应用。在传染病监测预警方法基础上,结合生物信息学及互联网大数据等方面的发展,重点介绍了基于基因组学和互联网大数据的传染病信息学研究进展及相关预测预警新技术,并在传染病监测技术基础上进一步介绍了传染病信息学在维护全球生物安全以及扩展应用到生物监测的有关情况,最后总结了传染病信息学当前面临的主要挑战以及在全球化时代传染病信息学 2.0 的发展方向。

本书可供传染病信息学专业科研人员及学生阅读和参考。

图书在版编目(CIP)数据

传染病信息学概论 / 宋宏彬, 郝荣章, 曾大军主编. —北京: 科学出版社, 2017.12

ISBN 978-7-03-055474-1

I. ①传… II. ①宋… ②郝… ③曾… III. ①传染病—信息学—概论 IV. ①R51-39

中国版本图书馆 CIP 数据核字 (2017) 第 281357 号

责任编辑: 丁慧颖 杨小玲 / 责任校对: 何艳萍

责任印制: 肖 兴 / 封面设计: 陈 敬

科学出版社 出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

中国科学院印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2017 年 12 月第 一 版 开本: 787 × 1092 1/16

2017 年 12 月第一次印刷 印张: 9 1/2 插页 4

字数: 200 000

定价: 68.00 元

(如有印装质量问题, 我社负责调换)

《传染病信息学概论》编写人员

主 编 宋宏彬 郝荣章 曾大军

副主编 王立贵 钱 全 曹志冬

主 审 孙岩松 马 慧 蒲 卫

编 委 (按姓氏汉语拼音排序)

曹志冬 郝荣章 贾雷立 李 浩 李 鹏

李沛翰 李振军 刘鸿博 马 慧 彭友松

蒲 卫 钱 全 邱少富 宋宏彬 孙岩松

王姣姣 王立贵 王全意 谢 靖 杨 朗

杨超杰 曾大军 张金萍 赵荣涛 祝丙华

前 言

传染病信息学是管理和分析传染病相关数据的一个新兴多学科交叉研究领域, 主要采用信息计算的定量化和可视化方法研究传染病流行的发生、发展、演化规律, 是构建传染病疫情及症候群的实时监测与预警体系、挖掘传染病疫情的时空传播规律、揭示内在传播扩散机制、评估公共卫生事件应急响应措施的有效性、提供数据和模型驱动的传染病防控科学决策支持的有效方法。目前传染病信息学在国外已有了长足的发展, 而在国内传染病信息学从理论到应用才刚刚起步。

因此, 国内迫切需要一本系统介绍传染病信息学基本理论和技术、应用实例的专业书籍来满足公共卫生研究与卫生管理人员的需求, 为他们深入掌握和应用传染病信息学提供一个切入点, 也为他们如何更好地搜集数据、更有效地分析数据, 乃至如何做出更科学的防控决策提供有益参考。

本书共 9 章, 第 1 章和第 2 章阐述传染病信息学基本理论和基本技术; 第 3 ~ 5 章阐述传染病时空数据分析算法和模型; 第 6 章和第 7 章介绍传染病信息学在监测预警领域的应用; 第 8 章介绍传染病信息学在生物安全领域的拓展; 第 9 章分析传染病信息学面临的主要挑战并展望未来发展方向。各个章节有机结合, 从理论到实际应用全方位地介绍传染病信息学技术在传染病防控中的应用。

本书所介绍的内容既有疾控专业人员和卫生决策者所需具备的传染病信息学基本知识, 也有传染病流行病学研究者需关注的前沿研究方向和未来发展展望; 既有理论与技术的概述, 也有大量实际应用案例介绍与分析, 使读者能更容易理解和掌握传染病信息学基本理论、技术与发展趋势。本书也为预防医学、生物统计学、兽医学、信息系统学、计算机科学、公共管理和政策等多个不同学科的在校本科生和研究生提供传染病信息学的概念、技术和实践相关知识。

本书作者均为从事传染病信息学研究及传染病预防与控制工作的专业人员, 对传染病信息学的历史脉络、前沿理论和实践应用有着深刻的认识和见解, 在此, 对参与本书编写的所有人员表示感谢。本书的部分内容来源于国家重大传染病科技专项 (2017ZX10303401, 2013ZX10004218) 和国家重点研发计划 (2016YFC1200700) 的研究成果, 此外, 本书的出版得到了国家科学技术学术著作出版基金的资助, 在此一并表示衷心感谢!

由于本书作者受各自领域学识水平所限, 文中难免存在一些不足、疏漏之处, 在此恳请读者不吝指正!

编 者

2017 年 8 月 1 日

目 录

第 1 章	概论.....	1
1.1	导言	1
1.2	概述	3
第 2 章	传染病信息学技术基础.....	9
2.1	传染病的流行强度	9
2.2	传染病流行的三种基本特征	10
2.3	传染病信息学技术方法的框架结构	12
2.4	传染病信息学的时空数据分析方法	16
2.5	传染病信息学分析的操作流程	18
第 3 章	传染病监测数据的空间分析.....	23
3.1	疫区范围的几何特征分析	23
3.2	空间相关性分析	26
3.3	点模式分析	31
3.4	空间回归分析	33
第 4 章	传染病动力学模型	37
4.1	经典 SIR 模型	37
4.2	SEIR 模型的改进	45
4.3	案例分析: 2003 年广州市 SARS 流行	50
第 5 章	基于复杂网络的传染病传播模型.....	57
5.1	复杂网络模型	57
5.2	基于复杂网络模型的传染病防控研究	63
5.3	小结	70

第 6 章	传统传染病监测预警技术	74
6.1	传染病监测预警探测算法	74
6.2	相关案例: ESSENCE II	82
6.3	小结	87
第 7 章	新型传染病监测预警技术	90
7.1	基于基因组学的传染病溯源与预测预警技术	90
7.2	基于互联网大数据的传染病监测预警技术	103
第 8 章	全球生物安全与生物监测	119
8.1	导言	119
8.2	生物安全威胁及其对国家安全的影响	120
8.3	全球生物安全监测策略	124
8.4	当前生物监测面临的挑战	128
8.5	建立高效全球生物监测系统	130
8.6	全球生物监测系统的展望	132
8.7	小结	133
第 9 章	传染病信息学面临的挑战与发展方向	137
9.1	传染病信息学 1.0 与传染病信息学 2.0	139
9.2	传染病信息学 1.0 面临的挑战	140
9.3	传染病信息学 2.0 的发展方向	142

彩图

第 1 章 概 论

1.1 导言

传染病流行，古时人们称之为瘟疫，既神秘又古老。几百年前，人们还认为传染病流行是一种天罚，科学文明降临以后，才逐渐认识到传染病是由病原微生物引起的能在人与人、动物与动物或人与动物之间相互传播的疾病。

人类在传染病面前是非常脆弱的，历史上人类曾多次遭遇灭顶之灾。据不完全统计，有史证以来，传染病造成的死亡人数远远超过战争、饥饿、地震、洪水等灾难造成的死亡人数之总和。例如，1918 年的西班牙流感在短时期内迅速杀死的人数超过 5000 万^[1]；中世纪，鼠疫横行使得欧洲人口减少近 1/3；死于天花的人口数量更是以亿计，庞大的印加帝国因天花流行而灭绝。当前，科技文明虽高度发达，然而，艾滋病^[2]、病毒性肝炎^[3]、鼠疫^[4, 5]、严重急性呼吸综合征（SARS）^[6-8]、甲型 H1N1^[9-11] 等重大传染病仍不时肆虐人间，给人类带来了巨大的生存危机。

进入 21 世纪以后，SARS、人禽流感、甲型 H1N1 等重大传染病流行接连出现，频率和强度呈不断增强趋势，达到空前的历史水平。究其原因，生态破坏、环境污染、城市化加剧、人口高度集中、生产生活方式改变、药物泛滥等问题或许是传染病频发的重要影响因素^[12, 13]。此外，交通工具的高度发达和全球经济的一体化，使得世界成为“地球村”，24 小时以内，人们可以抵达地球上任一角落，这意味着任一区域性传染病都有可能短时期内发展成为全球性的大流行^[14, 15]。无论是传染病出现频率的增强，还是传染病传播流行能力的极大提升，都使人类社会面临着越来越严峻的传染病流行的重大威胁。

自古以来，人们就在积极探索应对传染病流行的有效途径，根据对传染病流行现象的长期观察，不断积累应对传染病的历史经验。人们发现，及时将被感染者隔离起来，焚烧其衣物和居室能够有效遏制传染病的进一步流行。如果将接触过传染病患者的人也一并隔离，效果则更加显著。几百年前，限于知识水平的不足，人们不清楚传染病因何而起，也不知道它的致病机制，更不明白它是如何成批地杀死正常人群，无从知晓其幽灵般地突然出现有何规律性或早期征兆。然而，基于隔离等朴素的观点，人们还是寻找到了一条行之有效的自我保护途径——只要保持与传染病患者一定的距离就

能够保护自己免受“上帝的惩罚”。

直到病原微生物的发现,人类对抗传染病的历史才开始步入科学与理性的范畴,经验防控逐步发展到科学防控。分子生物学、病原学、预防医学等学科得到蓬勃发展,并已成为应对传染病的强大工具。

新发突发传染病流行时,最根本性的方法是快速研制特异性药物迅速救治传染病患者,亦或提前在人群中普遍接种疫苗以建立抵御感染的屏障^[16]。然而,传染病流行具有出现突然、暴发性增长、传染人群庞大、演变复杂等特点,基于病原学等方法研制特效药的方式往往需要一个较长的时间周期,且受到科技水平的局限,实践中很难满足传染病防控的高时效性需求。同样,研制疫苗并在易感人群中普遍接种需要较长时间周期,实践应用受到很大局限。

历史经验表明,暴发性传染病流行时,及时有效的应对方法是快速识别并隔离传染源、切断传播途径以保护未受感染的正常人群等措施^[17],它能最大限度地减缓疫情带来的不利影响。这就需要从流行病学角度来分析传染病流行的特征、流行规律、演化模式、影响要素及防控措施对疫情的影响等,使疫情防控的决策方法更为科学、更有针对性,使有限的防控资源取得最大成效。

在与传染病斗争了几千年后,人们切实认识到单纯依靠临床医学来提高个体治疗并不能最终达到消灭疾病、提高人群整体健康水平的目的。长期实践表明,基于宏观的角度,探究传染病在人群中的传播、发生和发展,即流行病学,才是真正有效应对新发突发传染病流行的手段。

流行病学的兴起和发展与数学方法的不断完善密切相关,它是采用数学方法从群体水平上研究疫情传播规律及环境与健康的关系。随着对疾病认识的不断深入,更多更精巧的数学方法被应用于研究传染病的流行,流行病学逐渐从一门实用科学发展成为方法论科学,其原理和方法不断向精确化和系统化方向发展,应用范围日益扩大,成为既古老又时新的科学研究。

21 世纪是网络信息化的时代,信息、计算机与网络科学的兴起正在改变世界,深刻影响各个学科的发展,改变了流行病学研究的传统模式。①疫情相关数据的收集逐步从传统的流行病学调查和病例诊断报告,扩展到基于 Web 社交媒体(web social media)的数据收集与疫情监测^[18-20];②地理信息科学的发展,使得传统采用的基于样本独立的经典统计方法逐渐发展到基于空间相关性假设的空间统计分析方法^[21-26];③流行病学不断与其他学科交叉融合,发展出新的学科,如传染病信息学^[27, 28]、空间流行病学^[25, 29, 30]等;④全球化使疫情防控面临空前挑战,世界各国联防联控已成为不可替代的疫情应对方式,开放模式的研究方法已成为重要方式^[14, 15];⑤疫情出现和演化的不确定性急剧增强,异构特性对疫情整体发展的影响越来越大,仿真模拟和随机性建模方法将是非常有效和可靠的研究手段^[31-34]。

如何适应新形势的发展需求,充分利用现代科学发展的新方法和新技术,增强应对重大传染病威胁的能力和水平已成为迫切需求。

1.2 概述

1.2.1 传染病信息学基本定义

传染病信息学是采用信息计算的定量化方法对于传染病流行的发生、发展、演化规律及人为干预措施之于传染病的防控效果进行系统、客观研究的一门学科。传染病信息学能够为传染病监测预警系统及突发重大传染病的科学防控提供支撑。

传染病信息学是一门多学科交叉的流行病学研究分支,它主要利用现代信息技术手段,例如,时空分析技术、概率模型、动力学建模和复杂网络建模等预警分析技术,构建传染病疫情及症候群的实时监测与预警体系,挖掘传染病疫情的时空传播规律,揭示内在传播扩散机制,评估公共卫生事件应急响应措施的有效性,为新发突发重大传染病疫情的应急响应与防控提供科学化的技术手段。

1.2.2 传染病信息学的研究方式

传染病流行中出现的各类现象及人与传染病互动的过程中,传染病信息学需要尽可能性地保持科学性。这就是说,传染病信息学并不是将自己的研究建立在无确凿证据的经验知识,而是依赖于科学证据,这些证据主要通过对传染病流行及人与传染病互动中的各类现象进行观察获得。传染病流行是一个高度复杂且动态变化的复杂巨系统,不同时间、不同区域、不同对象人群中暴发传染病流行的特征与传播模式会有很大差异,任何应对方式的制订都需要因时因地因人的根据所获取的科学数据来做出科学判断。

自然科学研究中,已有非常成熟的对客观事物进行客观研究的科学手段。例如,物理学研究中,只有当实验结果能够在重复实验中进行重现,实验方案及其结果方可被接受。

然而,对于传染病流行这一特殊领域,由于研究对象是人,因此,无法做重复性的科学实验。显然,我们不可能故意在人群中制造某种传染病以获取在特定监控手段下的科学数据,更不可能一而再,再而三的重复这一实验过程。对于具有特定时空域和特定人群中的传染病暴发和流行,均是独特的科学样本,具有一次性,不可重复。

因此,传染病信息学的研究手段往往不是基于控制实验,而是基于观察。在传染病流行现象及人与传染病互动的过程中,会产生大量数据。例如,人被感染后,会显

现症状,便会形成症状特征观察数据;显现症状的时间记录,便会形成发病时间数据;病例所在家庭住址记录,便形成了家庭地址数据;病人购买相关药物,便会形成药品销售记录数据;病人不能去上学,便会形成学校缺勤记录数据;病人在互联网上自诉症状以求医问药,则会形成网络文本记录数据。针对这些传染病相关数据进行一定规则下的系统性的数据收集和整理,便形成了当前形形色色的传染病监测系统,它为传染病信息学研究提供必不可少的数据基础。

基于各类传染病监测系统获取的监测数据,对传染病流行进行科学性的判断和逻辑推理,是传染病信息学的核心工作,它是从数据表象通往理性认识及由此形成科学决策的必由之路。从这个角度来看,任何能够实现对传染病监测数据进行科学判读从而实现对传染病流行问题的理性认识的方法和技术,均属于传染病信息学的理论范畴。这是一个相对广义的界定,它并不局限于某一假设、某一定律、某一方法准则、某一技术范式。例如,基于传染病数据的经典统计分析方法是传染病信息学研究,因为它能提供对传染病在时间、空间和人群间分布的宏观特性的科学认识,而这一科学认识是传染病监测数据所无法直接提供的;基于传染病监测数据的传播动力学建模分析也是传染病信息学研究,因为它能提供传染病在人群中传播扩散的动力学特性及传染病流行的宏观趋势,这些科学认识也是传染病监测数据所无法直接提供的;基于地理信息系统(Geographic Information System, GIS)的传染病数据可视化技术也是传染病信息学研究,因为它以一种非常直观的方式将监测数据转化为一种易于人类视觉判断的整体感知,这也是传染病监测数据所无法直接提供的。

利用传染病信息学方法,能够显著提升传染病早期预测、预警分析的能力,增强时效性与准确度,实现从传染病流行病学特征描述到内在机制过程演化规律挖掘的突破与升华。

1.2.3 发展传染病信息学的目的与意义

当前社会,人口不断向城市高度集中,群居混居十分普遍,发达的交通网络更使得人群的活动交往达到空前的广度和深度,急剧增强了新发突发传染病流行的重大威胁。我国是世界上人口最多的国家,人口聚集度很高,跨国界、跨地区的人口流动非常频繁,急性传染病一旦在人群中暴发便将造成极为严重的后果。

最近10年,我国接连遭遇2003年SARS流行、2005年人禽流感流行、2009年甲型H1N1流行等重大疫情,它们不仅严重威胁公众生命安全,而且事关社会稳定与国家安全,制约我国经济可持续发展和人民健康水平的提高,如何有效应对重大疫情带来的公共卫生危机已成为我国政府面临的严峻挑战。

2003年SARS流行是中华人民共和国成立以来我国面临的最大一次传染病威胁,

我国公共健康体系面临前所未有的艰巨挑战。事实证明,在应对重大传染病流行以保障我国人民生命健康方面,仍有相当长的路要走。SARS 事件平息以后,中国公共卫生部门不断总结经验教训,反思如何更科学有效地应对突发重大疫情的威胁。于是,我国先后启动了“国家救灾防病报告管理信息系统”和“传染病疫情网络直报系统”,军队系统也同步建立了“全军传染病疫情信息监测系统”。国家、省市及军队体系建立了面向重大传染病疫情事件的应急指挥系统,这些举措使我国的传染病疫情监测与防控水平得到了显著提升。目前,我国已积累了传染病疫情和媒介生物方面的海量数据,取得了举世瞩目的成就。然而,疫情数据的科学利用仍然严重不足,主要体现:①疫情数据的分散、孤立、标准不统一、非结构化、时空尺度不一致等现象仍然非常普遍;②影响传染病流行的自然因素和社会因素的作用机制的挖掘不充分;③现有的监测预警系统往往只考虑传染病病例在时空上的分布变化,缺乏对主导传染病流行的自然因素和社会因素的考虑,使得预测预警的时效性较为滞后,缺乏前瞻性的预判;④疫情信息来源较为单一,监测预警效果不理想,无法实时地捕捉到疫情动态变化的早期征兆;⑤疫情数据积累很多,但相应的科学分析方法、技术和工具相对落后,难以满足疫情应对处置的现实需求。以上不足,严重制约了我国对传染病疫情的实时掌控和应急防控的能力,极大地限制了制定防控策略的科学化水平。综合运用现代多学科交叉的信息技术方法,发展一套基于病原、自然环境、社会环境及网络舆情环境等多源信息整合的体系化的传染病信息学理论与方法,不仅十分必要,而且非常迫切。

目前,我国已基本建设完成一套覆盖 47 000 家哨点医院的传染病网络直报系统,已基本实现了对覆盖我国全境的 39 类法定传染病的实时监测。目前基于此套监测系统构建与之配套的传染病疫情监测预警分析系统,主要根据不同时空区域的传染病病例数量进行异常模式探测。尽管目前正在投入运营的传染病监测预警方法能够达到一定程度的监测预警效果,但也存在明显的不足,主要体现:①预警时间节点要晚于公共卫生部门上报病例的时间节点,因此,预警时间节点较为滞后,时效性仍然不足;②传染病疫情流行是病原体与自然环境和社会环境相互作用的结果,自然环境和社会环境是“成因”,不同时空区域出现的病例是“结果”,现有的预警体系主要针对“结果”,不对影响疫情的自然环境和社会环境“成因”进行监测与分析,因此,无法对大范围、广尺度、宏观层面的疫情动态趋势进行前瞻性预测,无法在疫情流行结果出现之前即迅速、灵敏地预判出传染病流行的变化动向与未来发展趋势。因此,有必要发展适应我国国情的、与现有的监测预警体系互补的、从传染病疫情“结果”监测扩展到“成因”监测的具有前瞻性预测效果的传染病信息学预测预警方法,这对我国的公共卫生防控有重要意义。

传染病流行是一个演变复杂、实时动态变化、多成因关联、涉及面广泛的复杂巨系统,实时快速地捕捉到传染病疫情中的各种异征兆、变化和动向是极其困难的,并

且是关键难点,直接关系到预测预警的实施效果。现有的监测预警系统虽然能够监测医院上报的病例信息,然而,仍有大量反映疫情实时动向的公共卫生信息却并未进入医院监测的范畴,就医的群体只是传播源的一部分,而且只有当严重影响个人正常生活的症状出现时才会选择就医,使得预警的时效性变差。近年来,互联网已跨入 Web 社会媒体时代,社交化、即时化、个性化已成为不可逆转的发展趋势,博客、微博客、社交网络、移动网络等创新的网络信息传播模式不断革新,正在急剧改变人们的生活方式,人们乐于在网络空间中分享各自的地理空间信息,积极播报身边发生的各类事件,并注明时空地理位置信息。这将给公共卫生监测体系带来新的机遇和挑战。Web 社会媒体中的时空信息拥有极其巨大的应用价值,在突发公共卫生事件的应急响应中尤其重要。任何一个区域和地区性的疫情暴发,往往会成为网络舆情热点,大量民众会自发地把发生在自己身边的疫情相关的第一手资料共享到网络上,对这些网络空间中的公共卫生信息进行实时监测与分析,能够非常灵敏地捕捉到疫情暴发的各类实时信息及其动态变化,这一新兴手段的监测模式将给公共卫生事件监测和应急防控带来重大革新。例如,基于 Web 社会媒体数据的在线监测与分析,HealthMap (<http://www.healthmap.org/zh/>)实时追踪了 2009 年甲型 H1N1 全球大流行,成为世界卫生组织(WHO)的重要参考信息。我国拥有近 5 亿网民,在我国人口中的普及极广。Web 社会媒体与移动网络的广泛普及,使得每个人都可以成为一个智能传感器,每个人在社会和自然环境中自主移动,参与各项活动,并通过感觉器官对周围环境中发生的变化进行全面的感知、解读与集成,最后利用信息技术手段,在网络平台上分享感知得到的各类信息与知识。亿万网民群体则构成了一个极其庞大的社会传感器网络,对其提供的海量时空信息和知识进行加工处理,将能够有效监测传染病疫情的当前发展态势,实时评估各类公共健康风险,实现对传染病流行趋势进行快速及时地预测与预警。

目前,基于 Web 社会媒体对传染病疫情进行监测的新方法和新技术已成为国际前沿热点,以美国为首的西方国家及 WHO 正在全力构建网络空间中的疫情监测体系。我国虽有学者在这方面开展了一些前瞻性理论研究,但缺乏成套的技术体系,更无成熟实用的技术平台。面向未来,我国新一代的传染病预测预警体系中必须加入基于 Web 社会媒体的新的监测预警手段,与传统的监测预警方法形成互补,综合起来共同构成更为完善的传染病预测预警体系。本书中的传染病信息学方法将在基于 Web 社会媒体的传染病监测预警方面开展尝试性工作。

参 考 文 献

- [1] Mills CE, Robins JM, Lipsitch M. Transmissibility of 1918 pandemic influenza. *Nature*, 2004, 432(7019): 904-906.
- [2] Cohen J, Pape J. GLOBAL HEALTH Haiti's Quake Shifts Clinic's Focus From AIDS to Aid. *Science*,

- 2010, 327(5965): 509.
- [3] Zaki SA, Asif S, Dadge D, et al. Co-existence of viral hepatitis with malaria. *Journal of Postgraduate Medicine*, 2009, 55(3): 233.
- [4] Taubes G. Prosperity's Plague. *Science*, 2009, 325(5938): 256-260.
- [5] Stenseth NC. Plague and the end of antiquity. *Science*, 2008, 321(5890): 773-774.
- [6] Anderson RM, Fraser C, Ghani AC, et al. Epidemiology, transmission dynamics and control of SARS: the 2002-2003 epidemic. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 2004, 359(1447): 1091-1105.
- [7] 王劲峰, 孟斌, 郑晓瑛, 等. 北京市 2003 年 SARS 疫情的多维分布及其影响因素分析. *中华流行病学杂志*, 2005, 26(3): 164-168.
- [8] Anderson LJ. Twenty-first century plague—The story of SARS. *Science*, 2005, 310(5747): 444-445.
- [9] Yang Y, Sugimoto JD, Halloran ME, et al. The Transmissibility and Control of Pandemic Influenza A (H1N1) Virus. *Science*, 2009, 326(5953): 729-733.
- [10] Fraser C, Donnelly CA, Cauchemez S, et al. Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings. *Science*, 2009, 324(5934): 1557-1561.
- [11] Cohen J. A Race Against Time to Vaccinate Against Novel H1N1 Virus. *Science*, 2009, 325(5946): 1328-1329.
- [12] Howlett R. Infectious disease—Past and future of an old foe. *Nature*, 2004, 432(7019): 817.
- [13] Michor F, Nowak MA. Immunology and evolution of infectious disease. *Nature*, 2002, 420(6917): 741-742.
- [14] Colizza V, Alain B, Marc B, et al. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, 2006, 103(7): 2015-2020.
- [15] Chan Y, Xu MRH. SARS: Epidemiology. *Respirology*, 2003, 8: S9-S14.
- [16] Villegas P, Hamoud M, Purvis LB, et al. Infectious Bursal Disease Subunit Vaccination. *Avian Diseases*, 2008, 52(4): 670-674.
- [17] Godoy MG, Aedo A, Kibenge MJT, et al. First detection, isolation and molecular characterization of infectious salmon anaemia virus associated with clinical disease in farmed Atlantic salmon (*Salmo salar*) in Chile. *Bmc Veterinary Research*, 2008, 4(1): 28.
- [18] Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data. *Nature*, 2009, 457(7232): 1012-1014.
- [19] Freifeld CC, Mandl KD, Ras BY, et al. HealthMap: Global infectious disease monitoring through automated classification and visualization of Internet media reports. *Journal of the American Medical Informatics Association*, 2008, 15(2): 150-157.
- [20] Collier N, Doan S, Kawazoe A, et al. BioCaster: detecting public health rumors with a Web-based text mining system. *Bioinformatics*, 2008, 24(24): 2940-2941.
- [21] Gerbier G, Bacro JN, Pouillot R, et al. A point pattern model of the spread of foot-and-mouth disease. *Preventive Veterinary Medicine*, 2002, 56(1): 33-49.
- [22] Meng B, Wang J, Liu J, et al. Understanding the spatial diffusion process of severe acute respiratory

- syndrome in Beijing. *Public Health*, 2005, 119(12): 1080-1087.
- [23] Haining RP. *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University, 2003.
- [24] 王劲峰. 空间分析. 北京: 科学出版社, 2006: 490.
- [25] Elliott PD, Wartenberg. Spatial epidemiology: Current approaches and future challenges. *Environmental Health Perspectives*, 2004, 112(9): 998-1006.
- [26] Wakefield J. Disease mapping and spatial regression with count data. *Biostatistics*, 2007, 8(2): 158-183.
- [27] Zollner I, Klarz R. Bridging from Medical Informatics, Biometry and Epidemiology to Medical Engineering. *Methods of Information in Medicine*, 2009, 48(5): 397-398.
- [28] Hota B, Jones RC, Schwartz DN. Informatics and infectious diseases: What is the connection and efficacy of information technology tools for therapy and health care epidemiology? *American Journal of Infection Control*, 2008, 36(3): S47-S56.
- [29] Ryan L. Spatial Epidemiology Some Pitfalls and Opportunities. *Epidemiology*, 2009, 20(2): 242-244.
- [30] Norman SA. Spatial Epidemiology and GIS in Marine Mammal Conservation Medicine and Disease Research. *Ecohealth*, 2008, 5(3): 257-267.
- [31] Eubank S, Kumar VSA, Marathe MV, et al. Structure of social contact networks and their impact on epidemics. *Discrete Methods in Epidemiology*, 2006, 70: 181-214.
- [32] Shirley M, Steve DF, Rushton P. The impacts of network topology on disease spread. *Ecological Complexity*, 2005, 2(3): 287-299.
- [33] Camitz M, Liljeros F. The effect of travel restrictions on the spread of a moderately contagious disease. *Bmc Medicine*, 2006, 4(1): 1-10.
- [34] Cauchemez S, Boelle PY, Thomas G, et al. Estimating in real time the efficacy of measures to control emerging communicable diseases. *American Journal of Epidemiology*, 2006, 164(6): 591-597.

第2章 传染病信息学技术基础

传染病是由病原体引起的能在生物体（人与其他生物）之间相互传播的疾病，与其他类型疾病最大的不同是，传染病具有传染性和流行性特征，且生物体被感染后往往会产生一定的免疫性。传染病通过各种传播途径在生物群落中传播、扩散与蔓延，随时间和空间不断演变，这一复杂过程看似随机和不可预测，然则富含科学内涵，内蕴大量控制传染病流行的机制和机理问题。目前，限于科技发展水平，突发新发传染病流行时，往往无法在短时间内掌握其根本性的机制和机理，因而，难以从分子生物学和病原学角度快速彻底地解决传染病流行对人类社会的重大威胁。作为认识和掌握科学规律的一种最重要手段，数学方法已成为推动人类文明发展的关键引擎，其在传染病流行研究中发挥着至关重要的作用^[1-9]。

马克思曾说“任何一门科学成熟的标志是数学的充分使用”。利用数学建模手段来定量研究传染病的流行特征及其发展规律，以实现科学指导下的有效防控，是目前国际社会应对传染病威胁的共识。目前，基于不同学科研究的视角，针对传染病流行规律和控制措施评价，已有大量数学建模的理论、方法和技术。尽管如此，在我国面临新发突发传染病流行时，传染病应对防控的决策仍主要依靠非量化的专家知识判断，在实践中缺乏时效性高、可靠性强的量化技术支持。有多种原因导致这种情形出现，如疫情数据非电子化、收集信息内容不全面、信息模糊等，流行病学调查数据的非结构化存储，缺失、遗漏、错误和不完备信息过多，这些因素都有可能大大限制疫情现场防控的科学决策。

为提高传染病疫情防控的科学化水平，一方面需要大力提高数据监测系统的能力和水平，另一方面，则需要对传染病流行中的特征、模式和规律有较为清晰的定量认知。

传染病信息学作为一门系统性的利用数学建模方法分析传染病流行中各类现象与规律的学科分支，是揭示传染病的流行特征与传播模式的重要方法，本章概要阐述传染病信息学所涉及的技术方法及体系研究的框架结构。它主要针对不同类型、不同完备程度的传染病调查数据，建立传染病流行的量化研究体系。

2.1 传染病的流行强度^[10]

传染病之所以被称为传染病，决定因素并不是传染病患者在总体人群中所占的比

例或传染病的致命程度，而是传染病的传染性。由于传染病的传染性，它会导致“一生十，十生百”的复合增长模式。平均情况下，若一个传播源在感染期内能够引起超过一个人被感染，则当一个传播源进入易感人群中后，它所引起的感染人数将随时间而呈指数增长。对于传染病而言，即使初始感染人数非常少，仍然可以造成灾难性损失，灾难性后果取决于再生速度。

大多数情况下，传染病在人群中引起感染的比例保持一个较低水平的稳定状态，此时的传染病的传播状态被称为非流行状态，也被称为散发（sporadic）。在散发情况下，观测病例会保持在历史同期的一般水平进行波动，且传染病例间不存在明显的关联关系。

在同一较短时期内某一局部范围，当传染病引发大量感染病例时，便被认为出现传染病暴发（outbreak）。当传染病引发区域人群感染的比例超过预期，并造成很大影响时，便被认为进入传染病流行（epidemic），它是与疾病散发相对的用于表示传染病传播强度的指标。例如，当某一学校里集中出现大量传染病性的病例时，可以认为学校内出现传染病暴发。若传染病“跑出”学校，传入社会，并在城市内引起远超历史同期的发病水平时，城市内的疫情便被认为是传染病流行。

当传染病迅速蔓延，疾病传播开始越过省界、国界和洲界，在全球多个不同区域引发流行时，传染病传播状态便进入大流行状态（pandemic）。传染病大流行往往造成巨大损失，影响深远，具有毁灭性威胁。历史上，曾多次出现传染病大流行，例如，中世纪欧洲的黑死病大流行（black death pandemic），第一次世界大战后期的西班牙流感大流行（Spanish flu pandemic）。最近一次大流行是2009年的甲型H1N1流感大流行[influenza A (2009) pandemic]。

当传染病在某一局限区域或特定人群中扎根，定期在此区域和人群中引起流行时，便被称为地方性传染病（endemic），例如，许多热带地区出现的麻疹流行。

2.2 传染病流行的三种基本特征

基于流行病学调查数据的横断面研究（cross-sectional study）是传染病研究中最常用的方法，它是在某一特定时间对某一定范围内的人群，以个人为单位收集和描述人群的特征以及疾病或健康状况。横断面研究最直接的目标是通过对于某一地区或人群的调查，获得某种疾病在时间、空间和人群间的分布特征（简称三间分布特征），从而发现高危人群或发现有关的病因线索，为传染病的防治提供依据。

2.2.1 时间分布特征

传染病传播是一个随时间动态变化的过程，其变化形式包括暴发、季节性、周期