

# 脑卒中

## 风险评估与预警体系

沈晓明 苏朝阳 主编



非外借

全国百佳图书出版单位  
中国中医药出版社

# 脑卒中风险评估与预警体系

名誉主编 马云枝

主 编 沈晓明 苏朝阳

全国百佳图书出版单位  
中国中医药出版社  
· 北 京 ·

## 图书在版编目 (CIP) 数据

脑卒中风险评估与预警体系/沈晓明, 苏朝阳主编. —北京: 中国中医药出版社, 2024. 3

ISBN 978-7-5132-8555-1

I. ①脑… II. ①沈… ②苏… III. ①脑血管疾病-防治 IV. ①R743

中国国家版本馆 CIP 数据核字 (2023) 第 226953 号

## 中国中医药出版社出版

北京经济技术开发区科创十三街 31 号院二区 8 号楼

邮政编码 100176

传真 010-64405721

北京盛通印刷股份有限公司印刷

各地新华书店经销

开本 787×1092 1/16 印张 18 字数 438 千字

2024 年 3 月第 1 版 2024 年 3 月第 1 次印刷

书号 ISBN 978-7-5132-8555-1

定价 79.00 元

网址 [www.cptcm.com](http://www.cptcm.com)

**服务热线 010-64405510**

**购书热线 010-89535836**

**维权打假 010-64405753**

微信服务号 **zgzyycbs**

微商城网址 **<https://kdt.im/LIdUGr>**

官方微博 **<http://e.weibo.com/cptcm>**

天猫旗舰店网址 **<https://zgzyycbs.tmall.com>**

如有印装质量问题请与本社出版部联系 (010-64405510)

版权专有 侵权必究

# 《脑卒中风险评估与预警体系》

## 编委会

名誉主编 马云枝

主 编 沈晓明 苏朝阳

副主编 吴 涛 王志飞 刘理想 王佳彬 郭如月

编 委 黄艳丽 边颂博 张明悦 杨冬谊 周 怡

兰 瑞 许玉珉 朱世瑞 张 杰 陈 冲

贾晓倩 郑伯媛 王燃冰 刁邵敏

国家中医药管理局岐黄学者支持项目 (No. 国中医药人教函〔2022〕6号)

国家高技术研究发展计划 (863 计划) (No. 2007AA02Z4B2)

“缺血性中风早期康复和避免复发中医方案研究”研究成果

国家卫生健康委员会科学研究基金省部共建重点资助项目 (NO. SBGJ 202102187)

“多模态网络分析构建缺血性卒中复发风险评估模型及管理系统的研究”研究成果

国家自然科学基金资助项目 (No. 81503422)

“基于 PINK1/Parkin 通路探讨小续命汤调控急性脑缺血再灌注后线粒体自噬的分子机制研究”研究成果

国家自然科学基金资助项目 (No. 81973618)

“小续命汤调控 JAK<sub>2</sub>/STAT<sub>3</sub> 通路影响脑缺血再灌注后星形胶质细胞活化与突触可塑性的分子机制研究”研究成果

国家中医药管理局中医药行业科研专项资助项目 (No. 201507003-8)

“中医特色健康管理的缺血性中风病复发风险评估系统研发及示范应用”研究成果

国家中医药管理局全国名老中医药专家传承工作室建设资助项目 (No. 国中医药人教发〔2014〕20号) 研究成果

河南省科技创新人才计划资助项目 (No. 496060404)

“通络愈瘫胶囊对缺血性中风气虚血瘀型患者 TXB<sub>2</sub>/6-Keto-PGF<sub>1α</sub>、TNF-α 的影响”研究成果

中国高校产学研创新基金 (No. 2023HT033)

“基于贝叶斯网络时空分析构建缺血性卒中复发风险预测多模态融合模型及病-证-药管理系统的研究”研究成果

# 序

脑卒中在我国是致死、致残率高居第一位的疾病，据统计，我国每年有 150 万~200 万的新发卒中患者，给人民健康及财产安全带来了极大威胁。大数据时代的到来，是脑卒中风险评估与预警研究的重大机遇。大数据的数据处理技术为脑卒中风险评估与预警体系建立的研究提供了支撑，为脑卒中风险评估级预警体系的建立提供了广阔前景。

21 世纪是信息化高速发展的时代，信息技术在各个领域不断渗透和推陈出新，深刻地改变着传统行业，在信息爆炸的时代产生海量数据，并迎来对大数据处理和应用从理论到技术上的巨大飞跃。不仅如此，信息时代的变革与创造性思维、全局观念和共情能力的深度融合，进一步催生了“高概念时代”。在大数据全面而深入的信息支撑下，经过人文科学富有创造性、艺术性和探寻意义的思维升华，科学问题得到凝聚、解释与深化，并被赋予美感和丰富的情感。这将极大地拓展科学研究的领域和范畴，促使科学格局向着科学与人文结合、实体本体与关系本体结合、线性与非线性结合的方向发展。医疗大数据的发展将为医疗卫生行业带来巨大的变革，进一步推动循证医学的发展。

《脑卒中风险评估与预警体系》一书不仅阐述了脑卒中中西医的研究与发展，更将其与我国传统医学相结合，共同构建起针对脑卒中的风险评估级预警体系。中医学是我国独有的具有原创思维的医学科学体系，数千年来在维护人民健康、保障中华民族的繁衍生息上起着无可替代的作用，其临床优势至今仍服务于亿万群众。本书通过阐释理念与临床实例相结合，在理念和方法上反映学科前沿和趋势，在实例上更贴近临床，极具启发和参考意义。



中国中医科学院首席研究员

2023 年 12 月

# 前言

随着信息时代的来临，各类疾病的医疗记录、医疗影像、健康检查、基因测序等医疗信息都被详尽地记录下来，既有结构化的数据，又有非结构化的数据。脑卒中作为威胁人类健康的重要疾病之一，近年来国际上的相关研究鳞次栉比，成为医学研究新兴领域的热点。医疗信息海量资料的爆发式累积亟待质的提升，就像潜在的金矿正等着我们去挖掘，而大数据正是医疗信息从量变中产生质变的关键，也正是我们挖宝的重要工具之一。中国台湾健保研究资料库的开放已为台湾医疗实务界和学术界的发展带来极大的助力，这让我们看到了信息时代医疗大数据应用于脑卒中临床的前景。毫无疑问，这里具有充满想象的空间。

信息技术在各个领域的渗透和推陈出新，深刻地改变着传统行业，在信息爆炸的时代产生海量数据，并迎来对大数据处理和应用从理论到技术的巨大飞跃。不仅如此，信息时代的变革与创造性思维、全局观念和共情能力的深度融合，进一步催生了“高概念时代”。在大数据全面而深入的信息支撑下，经过人文科学富有创造性、艺术性和探寻意义的思维升华，科学问题得到凝聚、解释与深化，并被赋予美感和丰富的情感。这将极大地拓展科学研究的领域和范畴，促使科学格局向着科学与人文结合、实体本体与关系本体结合、线性与非线性结合的方向发展。

随着西方医学的发展，生物医学模式逐渐向生物-社会-心理医学模式转变，医学关注的重点从“人的病”转化为“病的人”。经典的随机对照试验（randomized controlled trial, RCT）提供“理想”环境下的干预结果信息，而我们在临床面对的不是“理想”状态下的疾病，而是置于自然、社会、心理的高维环境之中的患者，因此，经典 RCT 结果与脑卒中的临床应用还有距离。随着对单纯还原论局限性的反思，实用性随机对照研究出现了，继而强调遵循临床实际、充分考虑患者意愿的研究理念发展起来了。

中医学是我国独有的具有原创思维的医学科学体系，数千年来在维护人民健康、保障中华民族的繁衍生息上起着无可替代的作用。其临床优势至今仍服务于亿万群众，中医学是整体论引领下的生命科学和人文科学相结合的医学科学体系，是全球范围内最重要的医药卫生资源。大数据脑卒中临床研究与中医学的传统研究实践相似，中医学的辨证论治、综合疗法、个体诊疗的特点和优势有望通过对脑卒中临床大数据的研究得以充分彰显。大数据时代的到来，是中医临床研究的重大机遇。大数据的数据处理

技术为脑卒中临床研究提供了支撑，至此，临床大数据研究的理念可得以充分体现，为中医药基于临床大数据研究获得共识疗效提供了广阔的前景。

本书对脑卒中临床大数据研究实践的梳理和总结，分为理念、方法、实例、体系四部分内容。第一章论述了大数据和中医药临床研究的内在相关性，指出大数据对中医药临床研究的巨大推动作用；第二章到第五章介绍了大数据背景下开展脑卒中临床大数据研究的方法论，包括数据来源、选题、设计、数据仓库构建、质量控制、数据处理分析等；第六章和第七章分别从“药”和“病”两个角度列举了大数据临床中医药研究脑卒中的多个实例；第八章到第十章主要讲述脑卒中风险评估与预警体系的原理、框架以及构建，希望读者能由此加深对大数据临床研究及脑卒中风险评估与预警的理解，也希望抛砖引玉，与开展中医药防治中风病临床研究的同道进行交流。

本书面向的读者是中医药学临床科研人员和临床医护人员。对于前者，希望中医药研究实践能鼓舞更多同道加入脑卒中临床大数据研究中，也希望通过脑卒中风险评估与预警体系研究能给正在开展临床大数据研究的同道一些启发和灵感；对于后者，希望本书能为临床实践提供一些线索或参考。

本书尽力在理念和方法上反映学科前沿和趋势，在实例上贴近临床，但大数据临床研究数据来源多样，设计方法不同，研究的切入点也各有特色，因此书中内容必不能反映中医药临床大数据研究的全貌。另外，中医药临床大数据研究尚属起步，从研究设计到结果解读都有尚待商榷之处，书中也会出现一些遗漏。恳请读者提出宝贵的意见和建议，以便再版时修订完善。

编者

2023年12月

# 目 录

## 第一章 中医药大数据临床研究 / 1

|     |                   |    |
|-----|-------------------|----|
| 第一节 | 信息时代的大数据医学研究      | 1  |
| 一、  | 信息时代的骄子——大数据      | 1  |
| 二、  | 信息时代大数据的特点        | 2  |
| 三、  | 信息时代下的大数据与云计算     | 3  |
| 四、  | 科学研究的第四范式         | 3  |
| 五、  | 医学理念的颠覆者——大数据     | 4  |
| 第二节 | 临床大数据研究的起源与发展     | 7  |
| 一、  | 临床大数据研究的源流        | 7  |
| 二、  | 国际上临床大数据研究现状      | 7  |
| 三、  | 大数据研究国内发展与创新      | 9  |
| 四、  | 大数据在中医药领域的研究      | 9  |
| 第三节 | 临床大数据研究的核心思想      | 11 |
| 一、  | 树立临床大数据研究的理念      | 11 |
| 二、  | 建立中医药临床大数据研究的思路   | 11 |
| 第四节 | 大数据理念与中医药临床研究的结合  | 12 |
| 一、  | 大数据与中医药在理念上相通     | 12 |
| 二、  | 中医药研究的传统模式——大数据研究 | 13 |
| 三、  | 中医药临床大数据研究的广阔时代背景 | 14 |
|     | 参考文献              | 14 |

## 第二章 临床大数据研究的选题与设计 / 16

|     |                    |    |
|-----|--------------------|----|
| 第一节 | 设定研究的问题与目标         | 16 |
| 一、  | 如何提出研究问题与研究目标      | 17 |
| 二、  | 中医药研究的得力工具——HIS 数据 | 19 |
| 第二节 | 确定研究设计             | 21 |
| 一、  | 队列研究               | 21 |
| 二、  | 巢式病例对照研究           | 23 |
| 三、  | 分析处方序列             | 24 |

|                       |    |
|-----------------------|----|
| 四、其他设计类型 .....        | 24 |
| 第三节 研究方案的制订 .....     | 24 |
| 一、确定研究题目 .....        | 25 |
| 二、分析研究背景 .....        | 25 |
| 三、明确研究目的 .....        | 25 |
| 四、确立研究方法 .....        | 26 |
| 五、绘制技术路线 .....        | 28 |
| 六、预期成果 .....          | 29 |
| 第四节 临床大数据的来源与特征 ..... | 29 |
| 一、临床大数据的来源 .....      | 29 |
| 二、医疗大数据的特点 .....      | 31 |
| 参考文献 .....            | 35 |

### 第三章 临床研究数据仓库的构建 / 36

|                              |    |
|------------------------------|----|
| 第一节 数据仓库在大数据临床研究应用中的特征 ..... | 36 |
| 一、HIS 的发展变化 .....            | 36 |
| 二、HIS 的结构模块 .....            | 37 |
| 三、HIS 数据在大数据临床研究中的应用 .....   | 39 |
| 四、HIS 数据应用于大数据临床研究的问题 .....  | 40 |
| 第二节 大型 HIS 数据仓库的建立 .....     | 42 |
| 一、数据仓库在临床大数据研究中的价值 .....     | 42 |
| 二、数据仓库对 HIS 的要求 .....        | 43 |
| 三、数据仓库的建设方案 .....            | 44 |
| 四、建设数据仓库的主流与前沿技术 .....       | 45 |
| 参考文献 .....                   | 47 |

### 第四章 临床大数据的处理与分析 / 48

|                           |    |
|---------------------------|----|
| 第一节 数据预处理 .....           | 48 |
| 一、数据预处理的必要性 .....         | 48 |
| 二、如何进行数据预处理 .....         | 49 |
| 三、数据清理 .....              | 51 |
| 四、数据选择 .....              | 56 |
| 五、现代医院关于脑卒中的数据集成与变换 ..... | 57 |
| 六、数据归约 .....              | 59 |

|     |                       |     |
|-----|-----------------------|-----|
|     | 七、领域数据预处理 .....       | 64  |
| 第二节 | 统计分析 .....            | 68  |
|     | 一、定性数据的分析 .....       | 68  |
|     | 二、定量数据的分析 .....       | 79  |
|     | 三、统计图表 .....          | 91  |
|     | 四、混杂因素的控制 .....       | 96  |
| 第三节 | 数据挖掘.....             | 115 |
|     | 一、数据挖掘的基本步骤.....      | 116 |
|     | 二、数据挖掘的主要任务与基本方法..... | 117 |
|     | 三、文本数据挖掘.....         | 122 |
|     | 四、时间序列数据挖掘.....       | 125 |
|     | 五、复杂网络社区发现.....       | 126 |
|     | 六、数据可视化技术.....        | 132 |
|     | 参考文献.....             | 133 |

## 第五章 临床大数据研究的质量控制 / 135

|     |                     |     |
|-----|---------------------|-----|
| 第一节 | 专属数据库的质量控制.....     | 135 |
|     | 一、数据管理计划.....       | 135 |
|     | 二、设计数据采集工具.....     | 136 |
|     | 三、建立专有数据库.....      | 137 |
|     | 四、源数据的现场核查.....     | 138 |
|     | 五、整体数据核查.....       | 138 |
|     | 六、数据库锁定.....        | 139 |
|     | 七、数据管理文件归档.....     | 140 |
|     | 八、数据保密及受试者隐私保护..... | 140 |
|     | 九、数据采集的质量控制.....    | 141 |
| 第二节 | 实施过程的质量控制.....      | 141 |
|     | 一、四级检查的程序.....      | 142 |
|     | 二、四级检查的内容.....      | 143 |
|     | 三、人员培训.....         | 149 |
|     | 四、制定标准操作规程.....     | 150 |

## 第六章 临床中成药研究实例 / 152

|     |                |     |
|-----|----------------|-----|
| 第一节 | 中成药临床应用分析..... | 152 |
|-----|----------------|-----|

|                                       |     |
|---------------------------------------|-----|
| 一、脑栓通胶囊对缺血性中风患者随访半年生存质量指标评分变化的影响..... | 152 |
| 二、苦碟子注射液治疗缺血性中风急性期上市后再评价.....         | 155 |
| 三、灯盏细辛注射液与灯盏生脉胶囊治疗缺血性中风上市后临床再评价.....  | 157 |
| 四、舒郁颗粒治疗中风后抑郁症评分分析.....               | 160 |
| 五、天智颗粒治疗阴虚阳亢型缺血性中风患者症状评分分析.....       | 162 |
| 六、安宫牛黄制剂治疗老年急性脑梗死的临床观察.....           | 165 |
| 第二节 基于实验室指标的脑卒中中成药安全性评价.....          | 170 |
| 一、黄芪注射液治疗缺血性中风的临床观察.....              | 170 |
| 二、通络愈瘫胶囊对30例缺血性中风气虚血瘀型患者血清影响的分析.....  | 172 |
| 三、通络愈瘫胶囊对缺血性中风气虚血瘀型患者血清影响的分析.....     | 175 |
| 四、通络愈瘫胶囊治疗急性缺血性中风的血流变学影响分析.....       | 178 |
| 五、血栓通联合脑得生浓缩丸治疗急性缺血性脑卒中的临床研究.....     | 181 |
| 第三节 中医药联合治疗脑卒中的研究与评价.....             | 185 |
| 一、中药结合针刺治疗对椎-基底动脉供血不足患者血管内皮功能的影响..... | 185 |
| 二、中西医结合治疗脑梗死后抑郁症用药方案分析.....           | 188 |
| 参考文献.....                             | 191 |

## 第七章 临床脑卒中大数据病症研究实例 / 192

|                            |     |
|----------------------------|-----|
| 第一节 疾病临床治疗方案及其实效评估.....    | 192 |
| 一、缺血性中风病患者临床用药特征分析.....    | 192 |
| 二、缺血性中风病急性期中西药物群组模块分析..... | 200 |
| 第二节 疾病发病及转归的时空因素影响.....    | 209 |
| 一、中医学整体观念对疾病的认识.....       | 209 |
| 二、气象因素影响出血性中风发病的分析.....    | 210 |
| 参考文献.....                  | 216 |

## 第八章 中风病的危险因素 / 217

|                    |     |
|--------------------|-----|
| 第一节 危险因素定义及分类..... | 217 |
| 一、危险因素定义.....      | 217 |
| 二、危险因素的分类.....     | 217 |
| 第二节 危险因素研究进展.....  | 217 |

|                  |     |
|------------------|-----|
| 一、可预防性危险因素·····  | 217 |
| 二、不可预防性危险因素····· | 221 |
| 参考文献·····        | 223 |

## 第九章 中医特色的中风病复发风险评估与早期预警系统的构建 / 225

|                        |     |
|------------------------|-----|
| 第一节 中医风险预警系统框架·····    | 225 |
| 一、中医“治未病”理论·····       | 225 |
| 二、构建中风复发风险预警系统框架·····  | 226 |
| 第二节 中医风险预警系统的宏观理念····· | 239 |
| 一、中医“整体观念”·····        | 239 |
| 二、风险预警系统的宏观理念·····     | 250 |
| 参考文献·····              | 256 |

## 第十章 基于风险评估模式的中医健康管理 / 257

|                            |     |
|----------------------------|-----|
| 第一节 中风病健康管理的概念、特点与模式·····  | 257 |
| 一、概念·····                  | 257 |
| 二、中风病健康管理的特点·····          | 258 |
| 三、中医特色的中风病健康管理模式·····      | 259 |
| 第二节 风险评估模式下的社区健康管理模式·····  | 262 |
| 一、概述·····                  | 262 |
| 二、社区健康管理的价值考量·····         | 264 |
| 三、社区中风病高危人群健康管理现状·····     | 264 |
| 四、基于中医药特色创建的中风病社区健康管理····· | 265 |
| 第三节 大数据时代的中医特色健康管理·····    | 269 |
| 一、互联网及云计算远程平台技术的发展·····    | 269 |
| 二、健康医疗大数据时代·····           | 272 |
| 参考文献·····                  | 274 |

# 第一章 中医药大数据临床研究

## 第一节 信息时代的大数据医学研究

### 一、信息时代的骄子——大数据

在过去的数年中，信息技术在社会、经济、生活等各个领域不断渗透和推陈出新。在移动计算、物联网、云计算等一系列新兴技术的支持下，社交媒体、众包、虚拟服务等新型应用模式持续拓展着人类创造和利用信息的范围和形式。当今信息技术的发展及创新正使各个行业发生改变，推动信息时代进入大数据的新纪元。同时也推动了脑卒中临床大数据研究的发展，为脑卒中的临床研究带来了广阔的前景。

2010年2月，《The Economist》杂志发表了《The Data Deluge》的封面文章。文章指出，当今世界上的信息数量正快速递增，随着这股数据洪流不断增加，存储这些数据，提取并分析有用信息将变得更困难。商业、政府、科学以及人们的日常生活，都已经显现数据泛滥的前兆。处理数据泛滥的最好方法就是让更多数据被用到正确的地方，但这个过程可能会十分漫长。毕竟，人类学习处理数据洪流、找到如何管理它们的过程才刚刚开始。

2011年6月，麦肯锡咨询公司发布了《大数据：下一个竞争、创新和生产力的前沿领域》研究报告。麦肯锡在研究报告中指出，数据正渗透到当今每一个行业和业务职能领域，成为重要的生产因素。各行各业海量数据的挖掘和运用，预示着新一波生产率增长和消费者盈余浪潮的到来，大数据时代已经降临。

2012年3月，美国政府宣布投资2亿美元发起“大数据研究和发展倡议”，致力于提高从大型复杂数据集中提取信息和知识的能力，并服务于能源、健康、金融和信息技术等领域的高科技企业。2012年4月，英国、美国、德国、芬兰和澳大利亚研究者联合推出“世界大数据周”活动，旨在促使政府制定战略性的大数据措施。联合国也在2012年5月发布了《大数据促发展：挑战与机遇》白皮书，指出大数据对于联合国和各国政府来说是一个历史性的机遇，人们如今可以使用极为丰富的数据资源，来对社会经济进行前所未有的实时分析，帮助政府更好地响应社会和经济运行。

越来越多的政府、企业等机构开始意识到数据正在成为最重要的资产，数据分析能力正在成为核心竞争力。大数据时代对政府管理转型来说是一个历史性机遇，对于企业来说，对海量数据的运用将成为未来竞争和增长的基础。同时，大数据也已引起学术界的广泛研究兴趣，2008年和2011年，《Nature》与《Science》杂志分别出版专刊 *Big Data: Science in the Petabyte Era* 和 *Dealing With Data*，从互联网技术、互联网经济学、超级计算、环境科学、生物医药等多个方面讨论了大数据处理和应用。

## 二、信息时代大数据的特点

大数据指的是无法使用传统流程或工具处理或分析的大量数据的集合。大数据既是数据量的一个激增，同时也是数据复杂性的提升。大数据与过去的海量数据有所区别，其基本特征可以用3个“V”开头的英文关键词来描述，即体量（volume）大、类型（variety）多、速度（velocity）快。

大数据的第一特征是数据体量巨大。大数据的数据存储量的计量单位从太字节（TB）量级跃升到拍字节（PB）量级。当前，典型个人计算机硬盘的容量为TB量级，而一些大企业的数据量已经接近艾字节（EB）量级。1E字节=1152921504606846976字节，约相当于一般个人计算机硬盘容量的100万倍。如今，传感器是生成数据的主要来源，2010年生成了1250亿千兆字节的数据，超过了宇宙中所有星星的数量。

大数据的第二特征是数据类型繁多，包括结构化的数据表和半结构化的网页以及非结构化的文本、图像、视频、地理位置等。物联网、云计算、移动互联网、车联网、手机、平板电脑以及遍布地球各个角落的各种各样的传感器，无一不是数据来源或者承载的方式。这些多类型多来源的数据对数据处理能力提出了更高的要求。

大数据的第三个特征是数据增长与处理速度快。数据源增加、数据通信的吞吐量提高、数据生成设备的计算能力提高，使得数据产生和更新的速度非常快。传统数据仓库、商务智能应用都采用的是批处理方式，但对于大数据，必须进行实时数据流处理。

产业界对大数据特征的定义普遍采用上述“3V”特征来描述，不过也有人认为除了“3V”特征，还应该增加1个“V”，即价值（value），它是大数据处理与分析的最终意义，即获得洞察力和价值。日本野村综合研究所认为“所谓大数据，是一个综合型概念，它包括因具3V（volume/variety/velocity）特征而难以进行管理的数据，对这些数据进行存储、处理、分析的技术，以及能够通过分析这些数据获得实用意义和观点的人才和组织”。这实际上是在广义层面上为大数据给出了一个定义，如图1-1所示。

所谓“存储、处理、分析的技术”，指的是用于大规模数据分布式处理的框架Hadoop，具备良好扩展性的NoSQL数据库，以及机器学习和统计分析等。所谓“能够通过分析这些数据获得实用意义和观点的人才和组织”，指的是目前在世界各地十分紧俏的“数据科学家”，各行各业能与“数据科学家”对接的多学科人才，以及能够对大数据进行有效运用的组织。

维克托·迈尔·舍恩伯格认为大数据有三个主要的特点，分别是全体性、混杂性和相关性。第一，是全体性，即收集和分析有关研究问题更多、更全面的数据，数据量的绝对数字并不重要，重要的是有多少数据和研究的现象相关，通过与研究问题有关的更多、更全面的数据可以看到很多细节，这些细节在以前通过随机抽样方式获取少量样本数据的条件下是得

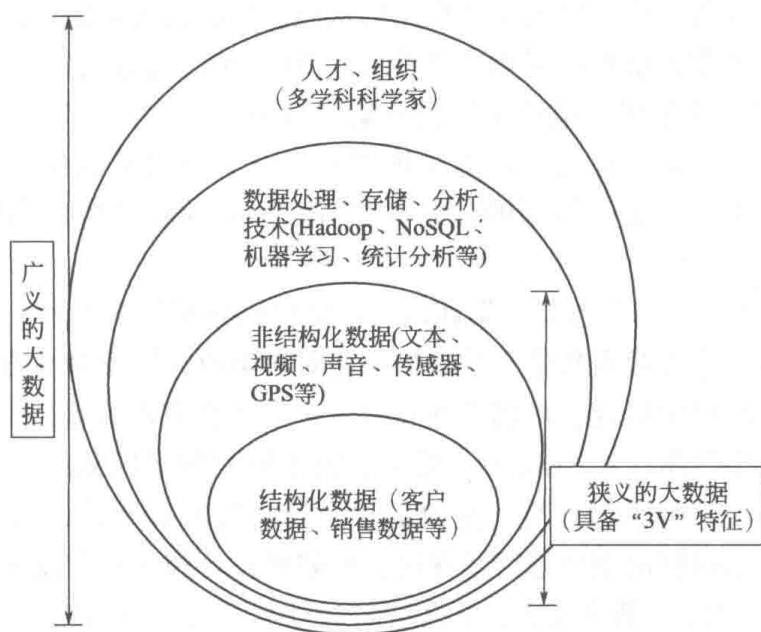


图 1-1 大数据的范畴

不到的。第二，是混杂性，即接受混杂的数据，在小数据时代人们总试图收集一些非常干净的、高质量的数据，花费很多金钱和精力来确定这些数据是好数据、高质量的数据，可是在大数据时代，就不再去追求特别的精确性。当微观上失去了精确性，宏观上却能获得准确性。第三，是相关性，因为大数据的混杂性特点，要求人们从小数据时代寻求因果关系转向大数据时代发现相关关系。

### 三、信息时代下的大数据与云计算

大数据的爆发性增长和互联网技术的飞速发展催生了云计算。大数据巨大的数据量使得传统的单机运算无法进行，云计算通过将计算分布在大量的分布式计算机而非本地计算机或远程服务器中从而使数据得以利用，这样的服务基于互联网使普通用户受益，使得无法接触高性能计算的用户也可以享受每秒百万亿次的计算能力。

所谓的云计算，从广义上讲，它是一种动态的、易扩展的，且通常是通过互联网提供虚拟化资源的计算方式。从狭义上讲，它是指 IT 基础设施的交付和使用模式，通过网络以按需、易扩展的方式获得所需的资源（硬件、平台、软件）。提供资源的网络被称为“云”。云计算是分布式处理（distributed computing）、并行处理（parallel computing）和网格计算（grid computing）的发展，通过利用非本地或远程服务器（集群）的分布式计算机为互联网用户提供服务（计算、存储、软硬件等服务），从而有效地提高了对软硬件资源的利用效率，使用户通过云计算享受高性能并行计算所带来的便利。

### 四、科学研究的第四范式

随着大数据及相关信息技术的发展，科学研究的知识基础发生了革命性的变化。通过各类观察、感知、计算、仿真、模拟、传播等，科研领域的大数据正迅速产生、广泛传播和有效保存，正在逐渐成为科学研究的新基础和有力工具。尽管科学界一直在与数据打交道，但

大数据的洪流也在改变着人们对数据及其作用的认识。当数据海量化、泛在化、开放化、网络化和计算化时，它的作用就发生了根本性变化。2007年，已故的图灵奖得主吉姆·格雷（Jim Gray）在他最后一次演讲中描绘了数据密集型科研“第四范式”（the fourth paradigm）的愿景。将大数据科研从第三范式（计算机模拟）中分离出来单独作为一种科研范式，是因为其研究方式不同于基于数学模型的传统研究方式。这为脑卒中的临床研究提供了新的方向。

2009年10月，微软公司发布了《第四范式：数据密集型科学发现》一书的英文版。这是国际上第一本系统描述大数据现象、深刻揭示其对科学研究的革命性影响的著作，对大数据时代如何理解和组织科学研究、科研管理和科研服务具有重要的意义。该书扩展了吉姆·格雷的思想，基于数字科研（e-Science）提出了科学研究的第四范式，即以大数据为基础的数据密集型科学研究，从地球环境、健康医疗、科学的基础架构以及学术交流等四个方面，对数据密集型科学研究的愿景进行了探讨，就如何充分利用科学发展的第四范式提供了深刻见解。2012年11月，《第四范式：数据密集型科学发现》一书的中文版出版。

e-Science这一术语最早由英国科学家于2000年提出，用以概括在信息化基础设施支持下所开展的科学研究活动所需要的一系列工具和技术。如今，随着大数据时代的到来，科学发展正在迈入一个新阶段，科研的方法也从之前的实验型科研（experimental science）、理论型科研（theoretical science）、计算型科研（computational science）推进到第四范式——数据密集型科研（data-intensive science）。科学研究第四范式，将带来科学研究的革命。当科研人员可以方便地从宏观到微观、从自然到社会获得海量实时的观察和（或）实验数据，当这些海量数据普遍地可网络获取、可计算、可开放关联，当对这些数据进行分析、更新、扩展的方法和技术成为科学家和公众的普惠性工具时，知识成为可以被及时更新、广泛连接、灵活计算的活的生命体，可个性化地、动态地、交互地、智能化地嵌入到人们的研究、学习、管理和生活中，许多激动人心的潜力将被进一步开发，许多未知的领域和方向将呈现在人们面前。通过第四范式——数据密集型科研将更加有利于临床医学的研究推动和发展。

## 五、医学理念的颠覆者——大数据

### （一）群体模式向个体化模式的转变

从某种意义上说包括脑病在内的医学体系建立于流行病学基础之上，至少是建立于流行病学理念的基础之上。流行病学是研究特定人群中疾病、健康状况的分布及其决定因素，并研究防治疾病及促进健康的策略和措施的科学。2004年世界卫生组织（World Health Organization, WHO）对临床流行病学给予了极高评价，指出这门学科从群体层面和定量研究的方法出发，在推动全球卫生研究、创造最佳的研究成果、推进人类健康事业方面做出了突出贡献，在推动医学领域发展中起到了举足轻重的作用。

脑卒中的流行病学基础是概率论，关注在人群中占主体地位的人。因此，标准的临床医学试验有严格的纳入和排除标准，将非主体的人群，如老年人、儿童、妊娠或哺乳期妇女、肝肾功能障碍者，以及患有某些疾病的人排除在外，并且在纳入的人群中取95%或99%的可置信区间，假设可置信区间之外的个体表现出来的差异都是偶然的，可以不予考虑。这种