

主编◎李立明

大型人群队列研究 调查适宜技术



人民卫生出版社
PEOPLE'S MEDICAL PUBLISHING HOUSE

大型人群队列研究 调查适宜技术

主 编 李立明

编 者 (按姓氏笔画排序)

王 碧 琦 北京大学公共卫生学院

吕 筠 北京大学公共卫生学院

朱 硕 斌 北京大学公共卫生学院

刘 青 青 北京大学公共卫生学院

江 莱 北京大学公共卫生学院

孙点剑一 北京大学公共卫生学院

张 凤 北京大学公共卫生学院

周 斌 北京大学公共卫生学院

孟 若 谷 北京大学公共卫生学院

秦 晨 曦 北京大学公共卫生学院

郭 彧 中国医学科学院/北京协和医学院

谢 春 艳 北京大学公共卫生学院

廖 春 晓 北京大学公共卫生学院

秘 书 王胜锋 北京大学公共卫生学院

人民卫生出版社

图书在版编目(CIP)数据

大型人群队列研究调查适宜技术/李立明主编. —北京:
人民卫生出版社, 2014

ISBN 978-7-117-19727-4

I. ①大… II. ①李… III. ①流行病学-调查研究-中国
IV. ①R181.2

中国版本图书馆 CIP 数据核字(2014)第 241351 号

人卫社官网	www.pmph.com	出版物查询, 在线购书
人卫医学网	www.ipmph.com	医学考试辅导, 医学数据 库服务, 医学教育资源, 大众健康资讯

版权所有, 侵权必究!

大型人群队列研究调查适宜技术

主 编: 李立明

出版发行: 人民卫生出版社(中继线 010-59780011)

地 址: 北京市朝阳区潘家园南里 19 号

邮 编: 100021

E - mail: pmph@pmph.com

购书热线: 010-59787592 010-59787584 010-65264830

印 刷: 北京汇林印务有限公司

经 销: 新华书店

开 本: 787×1092 1/16 印张: 19

字 数: 474 千字

版 次: 2014 年 12 月第 1 版 2014 年 12 月第 1 版第 1 次印刷

标准书号: ISBN 978-7-117-19727-4/R · 19728

定 价: 59.00 元

打击盗版举报电话: 010-59787491 E-mail: WQ@pmph.com

(凡属印装质量问题请与本社市场营销中心联系退换)

前言

人群队列研究是国际上公认的研究环境因素和遗传危险因素与疾病发生结局之间关联的群体流行病学研究方法,是解决现代医学一些迫切问题的重要研究手段,也是转化医学研究的重要基础。目前,全球队列研究数量不断增加且规模愈来愈大,计划开展人数达 50 万的超大型人群队列研究就有 8 个。2011 年,《医学科技发展“十二五”规划》重点任务明确指出要建立我国的大型人群队列,目前国内已有中国成人慢性病前瞻性研究项目(CKB)构建的 51 万成人队列、复旦大学构建的 20 万泰州成人队列等大型队列问世,为后续开展各种专题性研究提供了实践经验和基础平台。

为贯彻落实我国《科学技术进步法》,科技部、卫生部等部委于 2010 年联合筹建国家人口与健康科学数据共享平台,旨在充分发挥人口健康领域科技资源的作用,提升我国人口健康领域的科技创新能力。就大型队列而言,尽管目前数量不断增多,但可比性低、共享性差,样本资源库和信息库建设不够规范等问题普遍存在,直接造成了重复投资和严重的资源浪费。在此背景下,有必要对我国大型队列研究整体设计和实施各个环节中的标准进行规范化研究,以便为国家层面上的标准统一、数据规范和资源共享奠定基础。

本书以国家科技支撑计划项目《区域人口健康大型队列适宜技术的标准化与应用》为依托,着眼于大数据时代及超大型队列研究不断发展的趋势,对大型队列应用中的各类信息和样本收集的规范化方式及流程进行了系统的梳理总结,并针对性地推荐了一些适用范围广、效果好、可行高的调查技术,以期为读者提供一本最终可获取高质量研究数据的案头工具书。大型队列研发的技术规范标准是慢性病防治综合技术和策略研究的技术支撑体系,一定程度上也是实现国家“疾病防治重心前移”战略目标的基础工作。“无以规矩,不成方圆”,作为流行病学工作者,希望《大型人群队列研究调查适宜技术》的出版,能够对推动开展我国大型人群队列研究,加强项目质量控制,以为形成科学可比的本土化科学证据作出积极的贡献。

全书内容分为四大部分 7 章 32 节:第一章介绍了大型人群队列调查的产生背景、开展现状及发展趋势;第二章简要介绍了队列中暴露测量的影响因素及注意事项;第三章至第六章按照不同暴露的测量方式,分别介绍了采用问卷调查、体格检查、器械检查、标本采集检测等各种方式评估不同暴露信息的可选操作种类,并加以对比和针对性推荐;第七章主要结合研究经费时常受限的原因,告知读者如何从研究目的出发,选择必要且充分的研究指标。整体而言,本书在写作过程中,尽量从读者的角度出发,考虑可读性、可参考性,同时努力及时地收集最新最权威的相关证据,力求实用高效。

本书写作团队以北京大学公共卫生学院流行病与卫生统计学系的十余名硕士、博士研究生为主力,历时近两年的时间,他们年轻且富有活力的创意和视野是本书顺利面世的基础。但因水平有限,难免会有疏漏,恳请读者批评指正。

衷心感谢吕筠老师和博士研究生孙点剑一同学前期所做的大量组织协调工作和全书写作框架的设计,为本书高效规范的写作奠定了基础。同时,本书特别邀请了北京大学公共卫生学院的詹思延教授、曹卫华教授、吴涛副教授、李劲副教授和陶秋山老师,北京协和医学院的潘慧教授,吉林大学公共卫生学院的于雅琴教授、寇长贵副教授,中国疾病预防控制中心施晓明教授、姜垣教授审阅了部分相关章节,他们丰富的专业知识和严谨的科学态度为书稿增色不少。本书学术秘书王胜锋老师的辛勤工作和多方协调使本书及时出版成为可能,在此一并致谢。

李立明

2014年7月

目 录

第一章 大型人群队列研究概述	1
一、前言	1
二、为什么构建 Mega cohort	1
三、Mega Cohort 现状简介	13
四、机遇与挑战	20
五、总结	21
第二章 暴露测量的原则	26
一、测量对象	26
二、测量方法及工具	27
三、测量误差	29
第三章 问卷法	34
第一节 评价方法	34
一、定义及分类	34
二、问卷的流行病学应用	35
三、选择问卷	36
四、问卷的开发	38
五、问卷的选用标准	41
第二节 吸烟调查	42
一、吸烟调查的分类	42
二、吸烟问卷调查的相关概念	44
三、吸烟调查问卷的设计经验	44
四、问卷调查的局限性	51
五、总结	51
第三节 体力活动	53
一、定义及测量目的	53
二、问卷的选择	55
三、问卷的设计思路	59
四、问卷的评价与选用	62

五、小结	64
第四节 膳食	72
一、概述	72
二、膳食调查方法	81
三、总结	92
第五节 饮酒	94
一、概述	94
二、调查问卷设置及举例	97
三、饮酒行为异常筛查问卷	108
四、总结	113
第六节 心理	118
一、综合心理测评量表	118
二、人格性格量表	127
三、抑郁量表	142
四、焦虑量表	155
五、小结	163
第七节 生活质量	169
一、简介	169
二、常用量表	169
三、总结	178
第八节 疾病史和家族史	182
一、定义及调查目的	182
二、调查内容及方式	184
三、小结	186
四、附录	187
第九节 用药史及干预史	191
一、定义	191
二、调查目的	191
三、调查内容	192
四、调查方式	193
五、总结	197
第十节 女性史	198
一、定义	198
二、调查目的	198

三、调查内容	199
四、调查方式	201
五、总结	205
第四章 体格检查	207
第一节 血压和心率	207
一、概述	207
二、水银血压计	208
三、电子血压计	211
第二节 体重和体脂百分含量	213
一、概述	213
二、体重(电子体重计)	213
三、体脂百分含量(皮褶厚度法)	214
四、体重及体脂百分含量(生物电抗阻法)	215
第三节 身高和坐高	217
一、身高的测量	217
二、坐高的测量	218
第四节 腰围和臀围	219
第五节 肺功能和肺一氧化碳浓度测定	220
一、肺功能测定(肺活量)	220
二、肺一氧化碳浓度测定	222
第六节 体质测试	223
第七节 被动吸烟	228
第八节 总结	229
第五章 影像学检查	233
第一节 胸部 X 线片	233
一、简介	233
二、前期准备及注意事项	233
第二节 心电图	234
一、简介	234
二、前期准备及注意事项	234
第三节 肝胆超声	235
一、肝脏超声	235
二、胆囊超声	236
第四节 颈动脉超声	236

一、简介	236
二、前期准备及注意事项	236
三、检查步骤	236
四、颈动脉狭窄诊断标准	237
第五节 骨密度——双能 X 线吸收法	238
一、简介	238
二、前期准备及注意事项	239
三、检查步骤	239
四、诊断标准	240
五、总结	240
第六章 生物样本处理与指标检测	242
第一节 血液样本采集与处理	242
一、简介及意义	242
二、血液样本的采集	242
三、样本的编码与识别	249
四、血液样本的处理	250
五、血液样本的保存与运输	254
六、计算机系统与样本的接收与贮存	255
七、总结	255
第二节 尿液样本采集与处理	256
一、简介及意义	256
二、采尿前期工作准备	257
三、随机中段尿采集法	257
四、尿液样本的临时保存与运送	258
五、标本的处理	259
六、现场化验项目	260
七、小结	261
第三节 粪便样本采集与处理	262
一、简介及意义	262
二、粪便样本的采集、保存、运输及注意事项	263
三、小结	266
第四节 唾液标本采集及处理	267
一、简介及意义	267
二、唾液样本的采集	268

三、唾液样本的保存与运输	272
四、小结	272
第五节 头发与指甲的采集与处理	274
一、意义	274
二、头发与指甲样本的采集	275
三、样本的保存与运输	276
四、样本的处理	277
第六节 常见生化指标列举	279
第七章 大型人群队列研究指标确定原则	290
一、建立大型人群队列的意义及特点	290
二、大型人群队列研究指标选择应遵循的原则	291

第一章

大型人群队列研究概述

>>>

一、前言

- 1854 年,61 例死亡病例分布,John Snow 发现了被污染的水源造成了霍乱的暴发¹。
- 1948 年,107 例肺结核患者,Austin Bradford Hill 赋予了“随机对照临床试验(randomized clinical trials,RCT)”真正的内涵²。
- 1950 年,709 例肺癌病例分析³,Richard Doll 与 Austin Bradford Hill 提出了吸烟导致肺癌的研究假设,随即启动了 British Doctor Study 队列,并证实了此关联⁴。
- 1958 年,5209 名小镇居民的募集,美国 Framingham Heart Study(FHS)提出了“危险因素”(risk factors)的概念^{5,6},并建立了“弗明汉风险评分”(Framingham Risk Score)的 10 年心血管疾病预测模型⁷。
- 1976 年,122 000 封回收调查问卷,美国的 Nurses' Health Study(NHS)开启了针对职业人群健康全球最大规模的队列研究,时至今日,自填式问卷调查的方式仍得以在 NHS 中被以最大化地利用和发展⁸。
- 1992 年,521 468 名欧洲 10 国的居民,23 个研究中心的横向合并形成了 EPIC(The European Prospective Investigation into Cancer and Nutrition)研究,短期横向合并的方式也真正开启了慢性病流行病学队列研究多国合作的先河⁹。

上述一个又一个流行病学里程碑式印记的出现,多伴随着研究样本数量级的递增。21 世纪的生物医学研究,“百万级”或“类百万级”的研究正在蓬勃发展:英国 Million Women Study 研究,于 2001 年完成了 1,300,000 名研究对象的纳入;我国的 China Kadoorie Biobank 研究,于 2007 年完成了 51,3211 名研究对象的纳入以及 51,3116 份血液样本的采集;法国、瑞典、美国等国也纷纷于近些年启动了以 50 万人群规模为目标的队列研究。截至 2014 年,据不完全统计,50 万及以上人群规模的队列研究已完成基线纳入的有 5 个,仍在进行的有 4 个,募集研究对象累计将超过 538 万¹⁰。

这样超大规模的人群队列研究,被形象地称为“超大规模队列(Mega cohort)”,其产生和发展震动了世界。Mega cohort 究竟是什么?为何出现?将如何发展?21 世纪生物医学领域是否会因为这样的“航母”级研究平台的出现而发生改变?在不远的未来,这些核心问题将被不断提出、质疑、解决和发展。

二、为什么构建 Mega cohort

原因 1:获得足够的罕见疾病新发病例

如表 1 所示,以 2005 年美国发病率数据为例,一个随访 5 年样本量为 50 万的队列的最直观结果是,在结束时能获得如表中所示的 13 种人类复杂性疾病病例共计 8 万多人。以

“帕金森氏病或精神分裂症”为例，一个随访 10 年样本量为 5000 的队列在结束时理论上仅能发现 4 例，但一个随访 10 年样本量为 50 万的队列研究能累积 424 例(如表 1.1)。通过超大规模样本量、积累够足量病例，此时不仅可以通过传统队列研究设计开展病因学研究，还能通过衍生病例对照研究设计(如巢式病例对照研究、病例队列研究、两阶段病例对照研究^[20]等)的方式开展进一步分析。

表 1.1 前瞻性队列研究发病估计¹¹

发病数 (率%)/ 10 万人/年	疾病	不同样本量(S)队列在不同观察时间点(年)的估计发病数								
		S1=5000			S2=50 000			S3=500 000		
		5 年	10 年	20 年	5 年	10 年	20 年	5 年	10 年	20 年
10(0.01)	帕金森氏病,精神分裂症	2	4	7	23	42	74	228	424	737
50(0.05)	结肠直肠癌,肾衰竭	11	21	37	114	212	367	1141	2118	3672
100(0.10)	乳腺癌,髌部骨折	23	42	73	228	423	731	2279	4227	7313
200(0.20)	糖尿病,卒中,心力衰竭	45	84	145	455	842	1450	4550	8418	14 503
500(0.50)	心肌梗死,癌症	113	208	354	1131	2078	3535	11 309	20 780	35 354
3000(3.00)	白内障,高血压	646	1123	1734	6464	11 231	17 339	64 644	112 315	173 391

注:a:假定每年的损耗率(失访率)为 3%;b:来源于美国发病率患病率数据库(2005 年)

如果说 Mega cohort 的构建初衷是为了在提高研究效能(多结局)的同时,在罕见疾病研究中获取足够的病例(即 EPIC 研究膳食与癌症的关系),那么如今的 Mega cohort 的构建目的则已经远远超出了这一范畴。

原因 2:人类复杂性疾病“基因-环境交互作用”病因研究的客观要求

现代病因理论认为“大部分人类常见疾病是基于基因、环境、基因-基因、环境-环境、基因-环境的交互作用而产生的”^{12,13},由此而产生了“人类复杂性疾病”的这一概念。人类复杂性疾病以慢性疾病、心理精神类疾病为主,潜隐期长,共享一定的危险因素(如吸烟、肥胖、少体力活动等),对个体而言具有病程长、预后差、致残高的生物学特点,对社会而言具有负担重、影响广、耗资大的社会经济学特点。作为 21 世纪威胁人类健康的首要原因,如何解开人类复杂性疾病的病因链,将注定成为 21 世纪生物学领域的研究核心。

解开人类复杂性疾病的“基因-环境交互作用”的谜团,以生物样本库为基础的前瞻性人群 Mega cohort 研究设计将是唯一途径。针对不同发病率的疾病(表 1.1),假定其发生是基于一定的基因-环境交互作用,通过 QUANTO 软件(版本 1.2.4)可以计算出所需的理论病例数¹⁴,按照 1:1 匹配的病例对照设计方法,笔者估算出不同样本量队列研究所需的最短随访年数如表 1.2 所示。

表 1.2 前瞻性队列研究时间估计(以探索基因-环境交互作用的巢式病例对照研究设计研究)

基因型 频率	环境暴 露率	所需的匹配 病例对照数 (对数)	不同样本量队列获得以下疾病所需病例数的最小年数(年)								
			帕金森氏病,精神分裂症 (发病率为 0.01%)			糖尿病,卒中,心力衰竭 (发病率为 0.20%)			白内障,高血压 (发病率为 3.00%)		
			20 万	50 万	100 万	20 万	50 万	100 万	20 万	50 万	100 万
基因-环境交互作用=1.5											
0.05	0.05	44 389	—	—	—	—	—	41	10	3	2

续表

基因型 频率	环境暴 露率	所需的匹配 病例对照数 (对数)	不同样本量队列获得以下疾病所需病例数的最小年数(年)								
			帕金森氏病,精神分裂症 (发病率为 0.01%)			糖尿病,卒中,心力衰竭 (发病率为 0.20%)			白内障,高血压 (发病率为 3.00%)		
			20 万	50 万	100 万	20 万	50 万	100 万	20 万	50 万	100 万
0.15	0.05	22 272	—	—	—	—	41	15	5	2	1
0.45	0.05	28 007	—	—	—	—	—	20	6	2	1
0.05	0.15	17 754	—	—	—	—	28	11	4	2	1
0.15	0.15	8743	—	—	—	40	11	5	2	1	1
0.45	0.15	10 665	—	—	—	—	14	6	2	1	1
0.05	0.45	11 184	—	—	—	—	15	7	2	1	1
0.15	0.45	5258	—	—	—	18	6	3	1	1	1
0.45	0.45	5870	—	—	—	21	7	4	2	1	1
基因-环境交互作用=3.0											
0.05	0.05	5157	—	—	—	18	6	3	1	1	1
0.15	0.05	2973	—	—	—	9	4	2	1	1	1
0.45	0.05	4699	—	—	—	16	6	3	1	1	1
0.05	0.15	2196	—	—	38	7	3	2	1	1	1
0.15	0.15	1173	—	—	15	4	2	1	1	1	1
0.45	0.15	1710	—	—	25	5	2	1	1	1	1
0.05	0.45	1642	—	—	24	5	2	1	1	1	1
0.15	0.45	746	—	21	9	2	1	1	1	1	1
0.45	0.45	829	—	24	10	3	1	1	1	1	1

注:“—”表示为了获得该类疾病所需的匹配病例对照数,在相应样本量队列中随访观察的时间>50年。(表2的假定前提为:研究开始时队列中不存在任何患者,显性等位易感基因频率为10%,环境暴露频率为10%,把握度定为80%,I类错误的概率为0.0001,遗传和环境边际效应为1.5,队列年失访率为3%。)

假定某基因型频率与某环境暴露率均为0.15,其基因-环境交互作用对于疾病的效应值(如比值比)为3.0(乘法模型),可以计算出理论上需要累计病例至少1173例。依照这一最低病例数,通过现有不同疾病年发病率的估计,则计算出不同样本量的队列研究需要随访的最短时间:

1)以发病率最高(约为3.00%)的“白内障,高血压”为例,样本量为20万、50万和100万的队列均可在1年内获取足够病例;

2)以发病率居中(约为0.20%)的“糖尿病,卒中,心力衰竭”为例,样本量为20万、50万和100万的队列获取足够病例需分别随访4年、2年和1年;

3)以发病率最小(约为0.01%)的“帕金森氏病,精神分裂症”为例,100万的队列需要随访15年,而样本量为20万和50万的队列即使随访50年也不能获取足够病例。

由此可见,相比以往较为单一的基于环境因素或遗传因素的流行病学病因研究而言,为识别人类复杂性疾病的“基因-环境交互作用”,队列研究的样本量需达到数以万计、十万计,

甚至是百万计的规模。

原因 3：“大数据”时代的必然趋势

“Big data is like teenage sex: everyone talks about it, nobody really knows how to do it. Everyone thinks everyone else is doing it, so everyone claims they are doing it. . .”

——Dan Ariely

杜克大学行为经济学家 Dan Ariely 在 Facebook 上的这句话从 2013 年 1 月 6 日发表以来,已不知在社交网络及主流媒体上被疯狂转发了多少次,形象地比喻了“大数据(Big Data)”概念在当下的混沌状态——“人人谈论,人人无策,人人推诿,人人冒名”。

从“机器学习”的角度理解,通过海量数据建模,数据从一个端口进去,结果从一个端口出来,中间的“黑匣子”在当前“实用主义”至上的社会似乎变得并不重要,但这绝非是“大数据”的本质诉求¹⁵。网络媒体监测(internet-based media surveillance)是公共卫生领域中率先应用大数据理念的先驱之一,其典型范例为 Google 公司于 2008 年开发的 Google Flu Trends(GFT)。GFT 以流感相关检索词的搜索量构建模型,来预测流感活动水平,构建之初,其运行结果不但和美国 CDC 的历史监测数据高度吻合,而且似乎能更早预测出未来的流感暴发,这当即给了固守传统监测的公共卫生学家迎头一击¹⁶。虽然随后有研究显示,GFT 的预测结果高估了实际的流感活动水平¹⁸,但在“时效性”和“精确性”的对垒中,“利用网络检索进行数据挖掘以预测”的方法,尽管在信噪比(signal-to-noise ratio)处理和选择偏倚的控制上存在诸多障碍,对于传统的监测仍然无疑是一种有效补充¹⁸。

“大数据”谈的不仅是数据量(Volume),还包括时效性(Velocity)、多样性(Variety)和精确性(Veracity),而“大数据”分析将凸显“总体取代样本、效率优于精确、相关重于因果”的三大特征¹⁵。瞬息万变的互联网时代,收发电子邮件、网络交易记录、多媒体文件上传下载等,任何一次点击就能产生数据,海量结构化或非结构化数据的产生也许能推动被互联网渗透下各行各业的革新和发展¹⁹。

Mega cohort 的建立对于生物医学研究领域来讲,似乎已经承载着“大数据”试验田的使命。假定一个随访 20 年人群规模为 50 万的队列研究,需收集的信息包括全基因组数据(30 亿个碱基对),重复收集表观遗传组学数据(illumina HumanMethylation 450K chip)及生化检验项目 5 次,以及 10 次问卷调查及体检测量,其数据量预估如表 1.3 所示。

表 1.3 50 万人队列研究随访 20 年的数据量预估

类别	单位数据量	测量次数	队列研究个体数据量	队列研究总数据量
全基因组	3 000 000 000	1	3 000 000 000	1.5×10^{15}
表观遗传组学	450 000	5	2 250 000	1.125×10^{12}
检验项目	500	5	2,500	1.25×10^9
问卷项目	500	10	5,000	2.5×10^9
体检项目	100	10	1,000	5×10^8

● 加法模型计算次数: $500\ 000 \times (3\ 000\ 000\ 000 + 2\ 250\ 000 + 2500 + 5000 + 1000) \approx 1.5 \times 10^{15}$

● 乘法模型计算次数: $500\ 000 \times 3\ 000\ 000\ 000 \times 2\ 250\ 000 \times 2500 \times 5000 \times 1000 \approx 4.2 \times 10^{31}$

● 指数模型计算次数: $500\ 000 \times 3\ 000\ 000\ 000 \times 2\ 250\ 000 \times 5000 \times 5000 \times 1000 = \infty$

从上述三种模型的简单计算来看, Mega cohort 能产生的数据价值已不言而喻²⁰。作为 21 世纪重要的研究平台和数据平台之一, Mega cohort 的到来将会进一步从时间顺序、暴露交互、病因网络等方面, 为明晰人类复杂性疾病的病因链而提供有力证据:

1) 拆分“外暴露-内暴露-暴露时间窗-生物有效计量-健康效应-临床前期标志物-疾病发生”时间链上的每一个环节²¹⁻²⁴;

2) 大量相关数据(控制变量或协变量)的收集和调整, 有利于控制混杂^{25,26}, 进一步理清病因网络的各个环节;

3) 个体化疾病预测、诊断、治疗及预防, 将会因为更为个体化的海量数据和参数而真正成为现实^{27,28}。

此外, 从实际操作和管理层面来说, 建立在 Mega Cohort 研究平台下的数据收集、清理、存储、检索、管理、共享、传输、分析、可视化等, 可能比完成研究本身更为重要, 其管理模式、技术创新及人才培养等方面的发展, 也将进一步服务于“大数据”。

原因 4: 组学技术的成熟

By the combinations of high-throughput analytical technologies in the fields of transcriptomics, proteomics, and metabolomics, we are now able to gain comprehensive and quantitative snapshots of the intracellular processes. Dynamic intracellular activities and their regulations can be elucidated by systematic observation of these multi-omics data.

——Kazuharu Arakawa and Masaru Tomita

生物组学发展已经跨越了学科和领域的界限, 从传统的基因组学到如今蓬勃发展的表观遗传组学、转录组学、蛋白组学和代谢组学, 组学测量技术的发展已宛如“纳米级相机”, 让人类可以捕捉到细胞内的反应过程²⁹ (当前生物组学名目汇总如表 1.4 所示)。

组学检测技术的发展, 使得 Mega cohort 在运用多组学测量技术方面的试探和应用成为可能。当前 Mega cohort 的建立越来越多的关注生物样本的采集、测量、储存和管理, 将生物银行(Biobank)的理念纳入到被研究对象的基线纳入及后期的随访环节。

对于疾病个体的组学探索, 在过往 30 年中蓬勃发展, 但如今却也已经到达一定的瓶颈。然而, 远端病因、外暴露、内暴露等如何作用于基因组、表观遗传组、转录组、蛋白组, 从而影响到代谢组, 产生生物学效应, 未来多组学发展只会越来越向“暴露”端倾斜, 如黄曲霉毒素加合物与肝癌、多环芳烃化合物与肺癌、芳香胺与膀胱癌等内暴露与疾病的关系如表 1.5 所示。对此, Mega cohort 的优势不仅在于丰富多样的生物样本资源及相对完善的个体信息, 更在于标化资料的收集过程和一定的人群代表性, 使得多组学分析的结果具有可比性和普适性²¹。

表 1.4 生物组学列表

	组学名称	核心	所属学科范畴
1	过敏原组学 (Allergenomics)	过敏原蛋白质组 (Proteomics of allergens)	
2	医学 (Medicine)	由临床试验所得的人类干预试验数据, 及其所扩展的临床结果和科学发表 (Human interventional trials data from clinical trial registries extended with trial results and links to resulting publications)	

	组学名称	核心	所属学科范畴
3	交互作用物组学 (Interactomics)	交互作用 (All interactions)	
4	生理组学 (Physiomics)	有机体生理 (Physiology of an organism)	
5	植物化学物组学 (Phytochemomics)	植物化学物 (Phytochemicals)	农业、食品、药品及化妆品学 (Agricultural, food, medicine and cosmetic sciences)
6	脂质组学 (Lipidomics)	脂质 (Lipids)	生物化学 (Biochemistry)
7	脏器组学 (Organomics)	脏器交互作用 (Organ interactions)	细胞信号及组织工程学 (Cell Signaling and Tissue Engineering)
8	细胞组学 (Cytomics)	有机体细胞组 (Cellular systems of an organism)	细胞学 (Cytology)
9	基因组学 (Genomics)	基因 (Genes)	遗传学 (Genetics)
10	表型组学 (Phenomics)	表型 (Phenotypes)	遗传学 (Genetics)
11	糖组学 (Glycomics)	糖类 (Glycans)	糖生物学 (Glycobiology)
12	干扰素组学 (Interferomics)	干扰素 (Interferons)	免疫学 (Immunology)
13	离子组学 (Ionomics)	无机生物分子 (Inorganic biomolecules)	分子生物学 (Molecular Biology)
14	激酶组学 (Kinomics)	激酶 (Kinases)	分子生物学 (Molecular Biology)
15	环境基因组学 (Metagenomics)	环境样品中的遗传物质 (Genetic material found in an environmental sample)	分子生物学 (Molecular Biology)
16	蛋白组学 (Proteomics)	蛋白质 (Proteins)	分子生物学 (Molecular Biology)
17	调控组学 (Regulomics)	影响和调控基因表达的转录因子及其他分子 (Transcription factors and other molecules involved in the regulation of gene expression)	分子生物学 (Molecular Biology)

续表

	组学名称	核心	所属学科范畴
18	转录组学 (Transcriptomics)	信使 RNA (mRNA transcripts)	分子生物学 (Molecular Biology)
19	表观遗传组学 (Epigenomics)	表观遗传学修饰 (Epigenetic modifications)	分子遗传学 (Molecular Genetics)
20	暴露组学 (Exposomics)	包括宫内环境在内的所有个体的环境暴露 (An individual's environmental exposures, including in the prenatal environment)	分子遗传学 (Molecular Genetics)
21	外显子组学 (Exomics)	外显子 (Exons in a genome)	分子遗传学 (Molecular Genetics)
22	开放阅读框组学 (ORFeomics)	开放阅读框 (Open reading frames, ORFs)	分子遗传学 (Molecular Genetics)
23	连接组学 (Connectomics)	脑的结构和功能连接 (Structural and functional brain connectivity at different spatiotemporal scales)	神经学 (Neuroscience)
24	食物组学 (Foodomics)	以消费者健康为核心,通过整合应用组学技术所涉及到的食品营养的生物活性、质量、安全性、可追溯性等 (Food and Nutrition issues related to bioactivity, quality, safety and traceability of foods through the application and integration of omics techs to improve consumer's well-being, health, and confidence)	营养学 (Nutrition)
25	职业暴露组学 (Exposome)	职业暴露及职业健康 (Composite occupational exposures and occupational health problems)	职业安全与健康 (Occupational safety and health)
26	药物基因组学 (Pharmacogenomics)	基因组对于药物的影响 (The effect of changes on the genome on pharmacology)	遗传学 (Genomics)
27	遗传药理学 (Pharmacogenetics)	单核苷酸多态性对于药物动力及药效的影响 (SNPs and their effect on pharmacokinetics and pharmacodynamics)	遗传学 (Genomics)
28	肥胖组学 (Obesidomics)	肥胖相关蛋白质 (Obesity related proteins)	蛋白组学 (Proteomics)
29	分泌蛋白组学 (Secretomics)	分泌蛋白 (Secreted proteins)	蛋白组学 (Proteomics)