Internal Ligament of Conservation Medicine: Big Data

Xiaolong Wang

Center of Conservation Medicine and Ecological Safety, Northeast Forestry University, Harbin 150040,
China

保护医学的内在纽带：大数据

王晓龙

东北林业大学保护医学与生态安全研究中心，哈尔滨 150040，中国

Abstract

Conservation medicine is an emerging interdisciplinary, taking the common health of human-animal-environment as its purpose. But since its come into being, the realistic vector of the epistemology of conservation medicine risk analysis, the approach of depth fusion for related disciplines and other key problems has not been solving yet. In this article the fact that data communication is the material base for internal integration of conservation medicine has been revealed by the analysis of the quantitative standards of discipline integration and the key nodes of discipline development of conservation medicine; By the analysis of the epistemology and demands of the conservation medicine risk analysis, that the big data is the most realistic carrier and internal driving force was illustrated, thus clearly pointed out that for the first time the ecological risk analysis based on the big data is the ultimate goal of conservation medicine.

Keywords: Conservation medicine; Big data; Risk analysis

摘要

保护医学是一门新兴的交叉学科，以实现人-动物-环境共同健康为宗旨，但自其产生以来，关于保护医学风险认识论的现实载体和学科深度融合的途径等关键问题一直没有得到解决。本文通过对保护医学学科融合的定量标准

和学科发展的关键节点分析，揭示了数据交流是保护医学内在整合的物质基础；通过对保护医学风险认识论和风险分析需求分析，阐释了大数据是保护医学最现实的载体和内驱动力，由此第一次清晰地指出了基于大数据的生态风险分析是保护医学的终极目标。

关键词：保护医学；大数据；风险分析

1. 前言

保护医学是解决人类所遇到的前所未有的生态难题的一门科学，是生命与健康技术的立体、交错式融合，健康数据交流是保护医学得以实现的物质基础和桥梁。

自 2002 年产生以来，即以诠释系列性、多学科、广泛可变空间中健康动态平衡为宗旨，囊括了人与公众健康、流行病学、兽医学、毒理学、生态学和保护生物学技术，融汇了个体健康、群体健康及景观健康等学科门类^[1]。

人们既惊喜又疑惑，惊喜的是我们终于有了一个探索生态健康的终极手段，疑惑的是人-动物-环境健康到底该如何契合到一起？

十多年来，该领域的实践尚未给出答案^[2-4]。

然而，随着大数据时代的到来，使我们看到了弥合学科间鸿沟的可能^[5,6]，也极大地拓展了保护医学的发展空间，即在数据采

集与交流、数据分析与存储等诸多方面实现各学科间的合作,对于进一步促进保护医学学科的完整与独立意义非凡。

2. 数据交流是保护医学的内涵

2.1 数据交流是保护医学学科融合的定量标准

保护医学研究的范畴是人的健康、动物健康(以新发传染病为代表)和环境健康的交集,但是各学科传统技术,以及数据采集和解读基本上是独立的,对于其他相关学科而言意义不大,甚至不存在,比如广州珠江口水域汞污染事件^[7,8]的数据,并不能给人医和兽医专业人员提供直接的执业参考,就是因为这种非数据转换大大降低了同一相关数据在不同风险分析主体间的确认,虽然事件过程中疾控中心对于公众食用生蚝安全问题给予了关注,但这并非基于数据的风险分析。

因此,保护医学相关各学科数据采集和解读的独立性是影响学科深度融合的关键问题之一,破解这一难题的技术思考即是数据层面的交流的实现。

2.2 数据交流是保护医学发展的关键节点

如何让临床流行病学数据、环境流行病学数据及新发传染病流行病学数据,甚至病原学数据、动物生物学数据、人体生理学数据等有机的统一起来是当前我们必须面对的问题。

上述大数据间的通道不打通,就无法实现真正意义上的保护医学风险评估,无法对环境健康、人畜健康进行数字化量化评估。而实质上,上述各大数据究其本质都是对生态健康的不同侧面的解读,打通壁垒,实现数据共享的物质基础正源于此。

一般来讲,我们由此可以得出一个结论,一个区域的环境大数据恶化,必然带来人和动物健康的恶化,以及传染病的爆发和流行概率的增大,只不过更多的时候,这种变化相对于现有检测技术的精度而言过于轻微,或干脆被忽略的结果,譬如内蒙古阿拉善地区的排污事件^[9],污染是明确的,而在鸟类的死因上却存在争议,而产生这种结果的原因就是环境数据不能转化为疾病驱动数据,因此一般监测机构或公众无法直观快速建立起二者之间的联系,管理部门也难以决断罢了。

此过程中沙漠生态从未被人提及,其实沙漠也有其独特而脆弱的生态系统,也是很多“不起眼”的物种的多样性宝库,这种程度的污染,我们有理由相信在一定时间尺度内是对沙漠生态的不可逆性破坏,污染数据与生态健康间数据链条断裂掩盖了事实,可见数据的共享与交流是多么的重要。

3. 基于大数据的风险分析是保护医学的终极目标

3.1 大数据是保护医学风险认识论最现实的载体

传统上,环境健康研究、人畜健康研究都是小样本研究,尽管在流行病学层面科学家基于各种目的,如提高流行病学检测精度、获得毒株等,大大提高了样本采集数量,但几千、上万的数量也难以被称为大数据!抽样终究是抽样,要的就是以管窥豹事半功倍的效果;大数据就是大数据,要做的就是全样本分析^[10]。

大数据对于采集的数据通过训练集与测试集加以区分,再依据逻辑与算法加工大数据、产生大数据,这些海量的大数据填充了每一个未经探索和求证过的领域,使人类认知的界限被最大化,多年的归纳和演绎手段被全样本分析和海量计算得出的结果所代替,体现在热敏图上的颗粒更加精细,空间尺度往往以米为标准,甚至可以做到更精确;此时,还有谁会去在意几十个样本的结论呢?

近年,基于Meta分析的统计研究表明,在医学、生态学等学科,很多同类的不同小样本研究结果需要统计评估其科学价值^[11],这又给传统分析方法带来了限制与烦恼。

3.2 基于大数据的风险分析是保护医学内驱动力

保护医学各相关学科融合的理论基础是WHO、OIE提出的:“同一世界,同一健康”,现行的环境毒理学、生态毒理学的研究终点是毒物的毒性阈值,看似合理,但是细究起来,却并无多少生态意义,甚至针对某种受试生物体获得的阈值,在环境低水平暴露情况下也缺乏指导意义,人类社会健康发展需要的是基因水平、组织水平、个体水平、种群(群体)水平和生态环境水平的健康反馈,环境保护的成

效不能由污染物环境残留量识别，必须以物种、环境的综合健康加以表征。

同样，新发传染病的管理不能单纯以控制为手段，应充分发挥生态稳定性的作用，使人类不经意释放的病原回到其自然疫源地，以生态位、种群适应力等理论关系使其健康、稳定存在，切记病原微生物亦是生物多样性的一份子，且是占绝大多数的类群！“消灭”的理念当休矣。

3.3 大数据为保护医学风险分析提供了直接工具

当前上述风险识别向风险表征的过渡往往通过拟合函数或其他数学模型解决^[12]，虽是量化手段，但存在着两个无法回避的问题，即精度和普遍性问题；且由于通过大量实验来验证和修正函数并不现实，使得拟合函数这一硬伤无法去除。

人的健康、动物健康（以新发传染病为代表）和环境健康正如保护医学金字塔的三条边，其共同的顶点即是风险分析，在最底层，即是他们原本的学科属性，相距甚远，而消弭他们之间隔阂的一定是基于大数据的风险分析，只有数据能够超越一切界限，只有数据能不断验证人类新的判断。

4 大数据存储是实时动态风险分析的前提

大数据存储是海量数据保有和超级计算的需要。

实时动态风险分析的产生经历了定性分析、定量分析、动态分析和实时动态分析四个阶段，较前两个阶段，动态分析更需要数据库的支撑，动态模型产生于海量预警数据，因为保护医学风险分析是事前预警，不是事后跟踪，不能通过近期发生的事件描述即时的风险，只有累积大量历史数据库，以及衍生大量数据库才能驱动动态模型运转；同样的，实时动态预警是对后台数据库的智能细分，达到提高预警时间尺度精度的目的，数据存储的需求更高。

存储也是一个硬件问题，关乎数据的可靠性、保密性和商业价值等。

超级计算，是军事、航天等多领域共同关注的超级话题，单就其单位时间运算次数就足

以令我们重视，保护医学风险预警的实时动态目标的实现，也依赖于计算速度；生态风险的复杂性，以及保护医学综合了多学科技术的兼容性问题，给算法的数据辨识和处理提出了空前大量的要求，如何在最短的时间内实现预警预报，关乎预警的即时性，也就是保护医学风险预警存在合理性问题。

5. 结论

大数据产业的发展给我们提供了借鉴，确保必须硬件设施、优化逻辑与算法、开发存储技术等是我们必须经历的，在这条道路上，每个节点都会再一次爆发出惊人的变革，带给我们超出学科发展的机遇和启示，大数据产业的井喷式发展，使我们对这些机遇满心期待！

保护医学风险分析也是大数据之源，我们利用大数据，同时也创造大数据，当我们脚步轻盈地追赶大数据的时候，也不可随意松开与我们握了多年传统领域的手，毕竟，逻辑与算法出自于对各学科问题的专业理解，海量的大数据需要专业人员的解读。

参考文献

- [1] A.A. Aguirre, R.S. Ostfeld, G.M. Tabor, C. House, M.C. Pearl. Conservation Medicine, Oxford University Press, 2002
- [2] 孙晓冬, 王海银. 新发传染病流行现状及防治策略. 上海预防医学杂志. 2009, 21(9): 461-465.
- [3] D. McCready, J.B. Williams. Risk assessment for a chemical spill into a river. Journal of Risk Analysis and Crisis Response, 2013, 3(3): 116-126.
- [4] J. Rovira, V. Linares, M. Bellés, M. Nadal, J.L. Domingo. Airborne levels of uranium in the surroundings of various industrial facilities: human health risks. Journal of Risk Analysis and Crisis Response. 2011, 1(1):42-47.
- [5] <http://www.ciotimes.com/bi/sjck/100072.html>
- [6] <http://www.36dsj.com/archives/29992>
- [7] <http://www.wasu.cn/Play/show/id/1025302>
- [8] 郑习健. 珠江广州河段底泥中汞铜铅的污染及其与有机质硫化物积累的关系. 热带亚热带土壤科学. 1994, 3(3): 152-

137.

[9]<http://finance.sina.com.cn/chanjing/cyxw/20150427/163722054894.shtml>

[10]盛杨燕, 周涛译. 大数据时代: 生活、工作与思维的大变革. 浙江人民出版社. 2013

[11]<http://scitech.people.com.cn/n/2014/0827/c1057-25552430.html>

[12]朱琳主译. 环境毒理学. 高等教育出版社. 2007