

## 综述

## 浅谈流行病学数据库的意义及应用

军事医学科学院 计算中心 周士波

随着电子计算机技术的迅速发展,它对流行病学的研究方法也起到了促进作用。首先是作为现代流行病学两大支柱之一的统计学有了惊人的发展,其次是数据库方法的引入,为流行病学的基本研究方法调查分析:收集、整理、分析数据提供了一种有力工具。本文仅从建立流行病学数据库的意义和方法谈一点初步的认识,以期推动流行病学数据库方法的发展。

**一、建立流行病学数据库的意义:**流行病学的特点是研究人群中疾病的分布及影响分布的因子,借以探索致病的原因及提供为疾病预防、治疗所用的方法和措施。为了实现此目的,我们要观察人群中疾病的分布和进程,检验从上述研究中提出的各种假说,研究改变某个认为有害的环境因子后对人群的影响或者衡量采取的某些预防性、治疗性和改良性措施在人群中的效应。由此可知,流行病学的研究对象是人群,不仅包括患病者而且包括未患病者和有关环境因子。流行病学研究采用的基本方法是调查、分析:对选定的人群进行有计划有目的的调查,获取大量的数据,再用统计方法进行分析,探索各种因素与疾病的定量关系建立数学模型<sup>[1~3]</sup>。也就是说,在流行病学的研究中将有大量的数据需要归纳、整理、分析。电子计算机技术的发展,特别是数据库方法的兴起和完善,成为流行病学研究的一个有力工具。即我们可以将调查所得建成一个数据库,从库中取得各种信息进行统计分析,以期建立一个理想的数学模型来反映疾病与因素之间的关系。这种方法,我们称之为建立流行病学数据库的方法。

流行病学数据库的建立将为流行病学的研究带来许多益处。首先它可以节约大量的人力和时间,提高工作效率和质量。数据库建成之后,可由计算机程序调用所需数据进行各种方式的处理。例如在疾病的初步分析中,常常要在某些特定条件下计算某一因子的百分率,以表的形式输出以便进行比较。用手工方法从一大堆数据中选取有关因子分类、整理、计算再制成表格是一件十分费时、费力的工作,而用数据库

方法完成上述任务则十分容易而且准确。

流行病学数据库的建立特别有助于进行前瞻性调查分析。因为前瞻性调查能直接估计因素与发病的联系以及联系的程度,而且所得的结论比较可靠,所以是流行病学研究中的一种有效方法。但是在前瞻性调查中花费时间长,积累的数据多,用人工处理这些数据比较困难。而用数据库方法将每次调查数据都存入库中,应用计算机程序进行分析,就为前瞻性调查提供了有力的工具。

流行病学数据库的建立能促进理论流行病学的发展。因为流行病学的研究已从定性描述向定量分析发展,研究某一因素对疾病流行的影响时需要做出定量的分析,而且为了研究各种因素同疾病之间的定量关系需要建立数学模型。而流行病学数据库和现代统计学方法的结合,将有效地发挥电子计算机的效力,为建立合理、可靠的流行病学数学模型提供有力帮助。

仅从以上几点,可以看出建立流行病学数据库对于推动流行病学的研究将具有重要意义。

**二、建立流行病学数据库的方法:**随着计算机技术的发展,提供了许多建立数据库的方法和手段。根据流行病学数据库的特点,归纳一下各种建库的方法,一般可分为二类:一类是建立数据文件型数据库的方法,一类是应用数据库管理系统建立数据库的方法<sup>[4]</sup>。

所谓数据文件型数据库就是把数据存放在一个文件中。根据数据存取方式不同,文件分为顺序文件、关系文件、索引文件等等。各种计算机高级语言都支持建立数据文件,有的计算机还提供专门建立数据文件型数据库的软件。如果在流行病学的研究中,我们采用数量化方法把调查所得的各种致病因素都用数值表示时,建立数据文件型数据库的方法是比较方便的。而且用各种统计分析程序,从该数据库中调用数据进行分析也十分方便。

当在流行病学研究中遇到因素比较复杂、涉及关系比较多时,我们可以应用数据库管理系统来建立流



行病学数据库。用此方法建立的库一般为关系数据库，它的特点是：允许存入大量数据，允许考虑因子之间存在的复杂关系，允许多个用户同时共享该数据库。这就为多个流行病学专家协同工作提供条件，每人可将自己调查所得及时存入库中，每人也可随时取得别人调查的新结果。我们设想在关系数据库中，是否可以存入与多种疾病有关因子的数据，以便对多个疾病进行流行病学的研究。

**三、一个应用实例：**在对某单位人员病毒性肝炎流行影响因子的分析中，我们试用了建立流行病学数据库的方法。我们在超级小型机 VAX-11/780 上用 Fortran 77 语言设计了一个流行病学信息处理系统 EPM，该系统具有建立数据库、输出各种报表及进行多因素分析三大功能〔5〕。

在 EPM 系统中包含一个建立数据文件型数据库的程序，它采用人机对话的方式将调查收集的数据输入到文件中去，即建成数据库。它不要求事先对数据整理、分类，可以由计算机程序自动分类排序。它还适用于随访调查，在原库的基础上很容易建成随访数据库。而且此库中数据可由计算机程序自动调用，进行各种分析处理。

在 EPM 系统中还有 25 个程序用来完成各种“率”的计算，并能输出打印成报表。这些程序可以根据不同要求，自动地在数据库中寻找、比较有关信息，在此基础上计算“率”。与原来的手工方式比较，EPM 系统具有迅速、准确二大优点。

此外 EPM 系统还具有多因素分析功能，可以对库中数据进行多因素逐步回归分析和 Logistic 回归分析，寻找与疾病发生关系最密切的因子。

我们用 EPM 系统建立的病毒性肝炎影响因素数据库共包含 2 053 个记录近 4 万个数据。根据流行病学分析需要输出了不同单位城乡人口 HBV、HAV 感染率的比较、不同籍贯中城乡人口感染率的比较等 25 张报表。最后用 Logistic 回归对数据进行分析得到了比较满意的结果。

**四、结语：**通过应用实践，我们感到采用建立流行病学数据库的方法，将推进流行病学研究的深入发展。在提高效率、节省劳力和时间、加强定量分析、建立数学模型等方面，流行病学数据库将成为流行病学研究的有力工具。

#### 参 考 文 献

1. 上海第一医学院，武汉医学院. 流行病学. 第一版. 北京：人民卫生出版社，1981：1~82.
2. Alderson M. 流行病学导论. 第一版，北京：人民卫生出版社，1982：1~11.
3. 重松逸造. 流行病学方法论. 第一版. 北京：人民卫生出版社，1984：1~25.
4. 冯玉才. 数据库系统基础. 第一版. 武汉：华中工学院出版社，1984：1~54.
5. 周士波，马静. 流行病学信息处理系统 EPM，军事医学科学院院刊 1987；11(6)：491.

## 一起由一例绦虫病现患引致猪囊虫病爆发流行调查

内蒙古乌审旗卫生防疫站 程 辉

1979至1980年，内蒙古乌审旗达布察克镇发生一起猪囊虫病爆发流行。为了查清本病的流行情况，找出传染源、为驱虫灭囊工作提供科学依据，近几年来我们深入该镇进行调查，现将调查结果报告如下。

**流行情况：**调查结果表明，猪囊虫病过去在该镇虽偶有发生，但均呈局部地段个例散发流行。本次猪囊虫病爆发流行波及全镇6个发病点，发现囊虫猪44头（其中78年2头、79年19头，80年23）头，占该镇历年发病头数的72%。本次流行自78年3月始至79年9月，历时18个月。传染源未能及时发现、且该镇猪大多散放街头是造成本次爆发流行的主要因素，流行强度为该镇历史资料所未见。

**流行因素：**1978年3月至1980年6月，一外来绦虫病现患付某流串该镇做木活，每到一处至少食宿半月以上，据当地居民反映付某个人卫生习惯差、随地便溺。1979年9月付某因食欲不振、恶心、腹痛、节片便史入院就诊，经本旗医院驱虫（驱出绦虫约2米多）治疗后于1980年6月返回原籍。这次调查证实，凡付某食宿做活之处，猪均有不同程度被感染，且无一处幸免。此外，我们在对该镇人群调查中亦未发现与本次爆发流行的明显线索。

**防治措施：**经现患治疗，猪全部圈养、圈厕定期消杀、流行终止。据笔者调查观察，从1981年至今该镇尚未发现囊虫猪。