

综合性医院住院者年龄分布特征的探讨

宇传华¹ 余松林¹ 徐勇勇²

【提要】 目的 探讨医院住院者的疾病种类、住院费用、住院天数的年龄分布特征。方法 采用共计 55 万余住院者的综合性医院病案首页数据库数据, 以入院日期与出生日期来计算每一住院者的实足年龄, 按出院第一诊断的 ICD-9 编码进行疾病分类。结果 住院者年龄分布曲线存在一个低谷和四个高峰: 除 20~49 岁年龄组外, 其余年龄女性住院者比率均低于男性; 肿瘤、循环系统疾病等住院者人数随年龄变化具有特定的分布特征。结论 病案首页数据库数据蕴藏着丰富的信息, 是疾病统计的宝贵资源。用它来探讨住院者年龄分布特征具有一定的合理性。

【关键词】 病案首页数据 数据挖掘 年龄分布特征 ICD-9

人的一生中难免生病住院, 随着年龄的变化, 在生命的长河中哪个年龄段需要住院就医的可能性较大? 对于不同疾病、不同性别等, 哪个年龄段的住院者会更多? 人均住院费用与人均住院天数如何随年龄的改变而改变? 带着这些问题, 本研究对 1995~2002 年之间 31 所综合性医院的 55 万余住院者病案首页数据进行了分析。

数据来源与研究方法

1. 数据的来源

本研究采用了四个住院者病案首页数据库, 广东省数据来自广州市 2 所综合性医院, 省级与区级医院各 1 所, 住院者为 1996~2000 年所收治的患者; 中西部地区数据来自中西部两大军区所属的 26 所医院, 其中 3 所为军区总医院, 10 所为中心医院, 13 所为驻军医院, 住院者为 1996 年全年所收治的患者; 江苏省数据来自该省 1 所综合性市级医院, 住院者为 1995~1998 年所收治的患者; 陕西省数据来自该省某医学院校两所教学医院, 住院者为 2001~2002 年所收治的患者。具体所收治的患者人数见表 1。

表 1 四个病案首页数据库的住院者人数

性别	广东	中西部	江苏	陕西	合计
男	61250	108125	20021	123359	312755
女	62534	69968	14474	92263	239239
合计	123784	178093	34495	215622	551994

2. 年龄的计算

按以下公式计算住院者的实足年龄 (岁):

$$\text{年龄} = \text{int}[(\text{入院日期} - \text{出生日期}) / 365.25]$$

其中, int 表示对括号内数据取整数, 一年按 365.25 天计算。0 岁表示从出生到不满 1 周岁, 依此类推。按每 1 岁清点住院者人数, 某岁住院人数与总住院人数之比值与 100 的乘积称为住院者比率 (%)。

3. 疾病的分类

按病案首页数据库中出院第一诊断的 ICD-9 编码进行疾病分类, 具体分类方法见表 2。

4. 统计软件

采用 SAS 软件编程计算与分析有关数据。

结果与分析

1. 四个数据库住院者的年龄分布

对于四个病案首页数据库数据, 在同一张图中各绘制一条住院者的年龄分布曲线, 获

¹华中科技大学同济医学院公共卫生学院流行病学与卫生统计学系 (430030)

²第四军医大学预防医学系卫生统计学教研室 (710032)

得的图形见图 1 所示。尽管四条曲线的分布各有所不同, 但其总的趋势具有类似的规律, 即曲线存在一个低谷和四个高峰。低谷出现在 10~19 岁年龄组, 这一年龄组的住院者相对较少, 反映了此年龄段是人生中的最“健康”时期。与广东和江苏病案首页数据库相比, 中西部和陕西的低谷年龄略有左移, 这可能与住院者身份(如军人)、地域经济等因素有关。第一个高峰出现在 0 岁, 在广东、中西部、江苏和陕西数据中, 该年龄的住院者比率分别为 4.85%、2.00%、3.90% 和 3.22%。第二个高峰是 20~29 岁年龄段, 造成这一高峰的主要原因是女性妊娠分娩。这一高峰的位置及变化幅度对于不同的数据库略有所不同, 特别是中西部和陕西数据库, 这是因为这些医院的服务对象主要是军人, 所以该年龄段有两个峰值, 前一个峰为男性军人住院者所致, 后一个峰为女性妊娠分娩住院者所致, 如果按性别进行分层分析, 可清晰看到这一结果。第三个高峰在 45 岁左右, 该高峰较小, 导致该峰的可能原因是: 该年龄段人群负担较重, 且处于生理更年期阶段。第四个高峰在 60~69 岁年龄组, 此年龄处于人生的衰老阶段, 致使该年龄组住院者比率又有所回升。

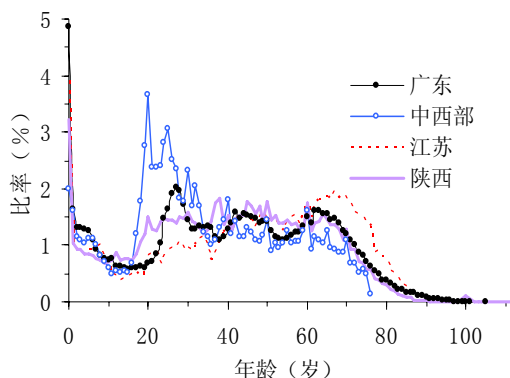


图 1 四个数据库的住院者年龄分布

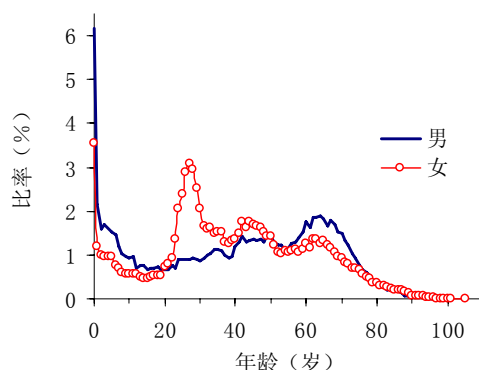


图 2 广东数据的不同性别住院者年龄分布

2. 不同性别住院者的年龄分布

由于男女之间生理上的差别, 不同性别的住院者比率可能有所不同。由于广东数据的住院者男女之比几乎为 1 (见表 1), 且其曲线位于几条曲线之间 (见图 1), 所以下面以广东数据来观察不同性别间的年龄分布。由图 2 可见, 男女之间住院者比率有较大差别。第一, 男性曲线只有明显的两个高峰 (即 0 岁组与 60~69 岁年龄组), 且其低谷不够明显; 而女性曲线共有四个峰。0 岁组与 27 岁组的高峰较明显, 45 岁左右与 60~65 岁年龄组的高峰不够明显。其低谷位于 10~19 岁。第二, 在 20 岁之前男性的住院者比率高于女性, 20~49 岁女性住院者比率高于男性, 50 岁开始男性的住院者比率又高于女性。

1993 年国家卫生服务调查^[1]以 0 岁到不满 5 岁分为第一组, 5 岁到不满 10 岁分为第二组, 其余每 10 岁分为一组, 60 岁以上作为最后一个年龄组, 得出女性住院者年龄分布曲线呈 W 形, 男性呈 V 形。如果本研究按卫生服务调查的年龄分组方法来绘制年龄分布曲线将获得类似结果。图 1、图 2 均采用了每 1 岁为一个年龄组, 所以本研究得出的结果相比之下更加精细。

3. 不同年龄的疾病顺位

为了反映不同年龄的疾病分布情况, 以病案首页数据库中出院者的出院第一个 ICD-9 诊断编码, 按 ICD 章节范围 (见表 2 第二列) 对每个住院者进行分类, 然后清点每一类的人数, 以人数最多的疾病排列第(1)位, 其次排列第(2)位, ……, 人数最少的疾病排列在第(10)位, 此称为疾病顺位。具体的疾病分类方法及四个病案首页数据库的疾病顺位结果见表 2。由表 2 可见, 四个数据库排在前五位的疾病是: 肿瘤、循环系统疾病、呼吸系统疾病、消化系统疾病、损伤和中毒。

表 2 ICD-9 分类以及四个数据库的疾病顺位结果

ICD-9 (按章节) 分类	ICD 范围	代码	顺位			
			广东	中西部	江苏	陕西
传染病与寄生虫病	001-139	1	(13)	(6)	(11)	(12)
肿瘤	140-239	2	(2)	(5)	(6)	(1)
内分泌营养代谢及免疫疾患	240-279	3	(12)	(12)	(9)	(13)
血液和造血器官疾病	280-289	4	(16)	(16)	(14)	(16)
精神疾病	290-319	5	(18)	(15)	(17)	(18)
神经系统和感觉器官疾病	320-389	6	(9)	(8)	(5)	(7)
循环系统疾病	390-459	7	(1)	(4)	(2)	(2)
呼吸系统疾病	460-519	8	(3)	(3)	(3)	(6)
消化系统疾病	520-579	9	(6)	(2)	(1)	(3)
泌尿生殖系统疾病	580-629	10	(8)	(7)	(7)	(8)
妊娠分娩和产褥期并发症	630-676	11	(5)	(10)	(18)	(9)
皮肤和皮下组织疾病	680-709	12	(15)	(11)	(16)	(14)
肌肉骨骼系统和结缔组织病	710-739	13	(11)	(9)	(10)	(11)
先天异常	740-759	14	(7)	(14)	(13)	(10)
起源于围生期的若干情况	760-779	15	(14)	(18)	(15)	(17)
症状体征和不明情况	780-799	16	(17)	(13)	(12)	(15)
损伤和中毒	800-999	17	(10)	(1)	(4)	(4)
V 编码	V01-V82	18	(4)	(17)	(8)	(5)

为了反映各数据库不同年龄的疾病情况, 将 0 岁到不满 5 岁分为一组, 5 岁到不满 10 岁分为一组, 其余每 10 岁分为一组, 然后对每一数据库的每一年龄组进行疾病顺位。每一数据库每一年龄组前五位的疾病代码结果见表 3。由表 3 可见, 婴幼儿与儿童少年(19 岁以下)住院的主要疾病是呼吸系统疾病和先天异常。青壮年(20~39 岁)住院的主要疾病是损伤和中毒、妊娠分娩与产褥期并发症。按性别分层后, 该年龄组男性住院主要疾病是损伤和中毒, 而女性住院主要“疾病”是妊娠分娩与产褥期并发症。中老年人(40 岁以上)住院的主要疾病是肿瘤、循环系统疾病和消化系统疾病。

在表 3 的基础上, 再按性别男、女对疾病进行顺位, 得到的结果基本上与表 3 相似, 差别较大的是 20~39 岁年龄组, 女性住院者主要以妊娠分娩与产褥期并发症为主, 而男性主要以损伤和中毒为主。

表 3 四个数据库不同年龄的前五位疾病

年龄组	广东	中西部	江苏	陕西
0~	14,8,15,6,9	8,17,9,1,14	8,9,15,14,17	8,14,18,15,17
5~	14,8,18,7,2	8,17,1,14,9	8,9,4,1,6	14,8,17,6,9
10~	14,18,7,2,8	17,9,8,1,6	8,9,17,6,10	17,14,6,9,8
20~	11,10,17,9,2	17,9,8,11,1	9,8,6,2,10	11,17,9,10,8
30~	11,7,2,10,9	17,9,8,10,2	9,17,8,6,2	17,2,9,7,11
40~	2,7,18,10,9	9,17,2,7,8	9,17,2,7,6	2,7,9,17,18
50~	7,2,18,9,10	7,9,2,17,8	9,7,6,2,8	2,7,9,18,6
60~	7,2,18,9,8	7,9,2,8,17	7,9,6,2,8	7,2,9,18,6
70 及以上	7,8,2,9,6	7,8,9,2,10	7,8,6,9,2	7,2,9,8,6

注: 每一格子中的 5 个数字分别代表表 2 第 3 列的代码所定义的相应疾病

4. 不同疾病住院者人数随年龄变化的特征

选择顺位排列在前的五类疾病(肿瘤、循环系统疾病、呼吸系统疾病、消化系统疾病、损伤和中毒), 来反映不同疾病住院者人数随年龄变化的特征。对每一类疾病按每一岁清点住院者人数, 以年龄为横坐标, 住院者人数为纵坐标绘制曲线。以广东病案首页数据库数据为例, 其五类疾病住院者人数随年龄变化的曲线见图 3 所示。随着年龄的增加, 肿瘤在 40~49 岁之间住院者人数有一个高峰, 50 岁以后逐渐下降, 在 55 岁左右趋于平稳, 60~69 岁

又有一个小的高峰。循环系统疾病的住院者人数在 50 岁前与肿瘤相似, 50 岁后超过任何一类疾病, 排在第一位; 65 岁左右的循环系统疾病住院者人数达到最高峰。呼吸系统疾病住院者比率有两个高峰, 一个是 0 岁婴儿, 另一个是 70 岁老年人, 后者相对较平坦。消化系统疾病的变化趋势与肿瘤类似, 但在 30~69 岁年龄段的增加幅度相对较小。损伤和中毒的高峰在 20~29 岁年龄组, 30 岁以后的变化幅度有逐年递减的趋势。对于其他数据库, 不同疾病住院者人数随年龄变化也有类似规律。

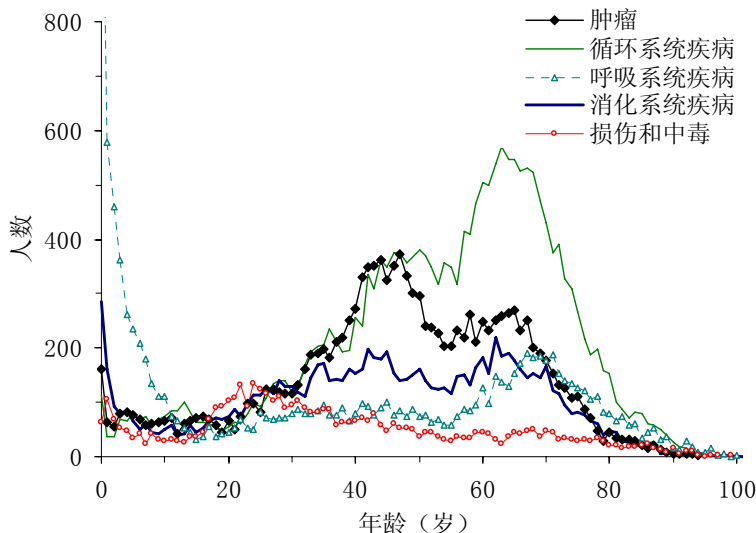


图 3 广东病案首页数据库五类疾病住院者人数随年龄变化的曲线

5. 不同年龄的人均住院费用与住院天数

人均住院费用受诸多因素, 如地域经济、消费水平, 医院诊治年份、诊治水平、病人病情轻重等的影响。表 4 列出了四个数据库不同年龄的人均住院费用, 因为通常情况下住院费用呈正偏态分布, 所以表中给出了人均住院费用的算术均数和中位数两指标。总的来说, 由于中西部和江苏的医院规模相对较小, 加之诊治年份相对靠前, 他们的人均住院费用相对较低。对于 0~19 岁, 20~59 岁两年龄组, 人均住院费用随年龄的增加而增加; 50~69 岁年龄组的人均住院费用维持在人生中的最高水平; 70 岁以后的人均住院费用又有所减少。由于 20~29 岁年龄组妊娠分娩者较多, 其人均住院费用相对较低, 由此导致了 20~29 岁年龄组的人均住院费用反而低于 10~19 岁年龄组的现象。具体数据参见表 4。

表 4 四个数据库不同年龄的人均住院费用 (元)

年龄组	广东		中西部		江苏		陕西	
	均数	中位数	均数	中位数	均数	中位数	均数	中位数
0~	8438	3196	771	450	1773	1460	4044	1362
5~	10956	4599	1322	687	1887	1462	6021	3128
10~	10974	4867	1467	789	2189	1409	6575	3141
20~	7758	3708	1687	862	3055	1693	5488	2421
30~	10432	5238	2286	1245	3769	2153	7821	4084
40~	12651	6394	2672	1588	4440	2664	8988	4589
50~	13921	7373	2972	1796	4858	3215	9972	4951
60~	13570	7864	2945	1755	5630	3906	10135	5093
70 及以上	11789	7799	2883	1715	5794	4112	9413	4428

住院天数的变化趋势基本上与住院费用相同, 即对于 0~19 岁, 20~59 岁两年龄组, 人均住院天数随年龄的增加而增加, 60 岁后人均住院天数随年龄的增加有减少, 但其变化

趋势不够明显。尽管中西部和江苏的医院规模相对较小, 人均住院费用较低, 但其人均住院天数反而较高, 具体结果参见表 5。

表 5 四个数据不同年龄的平均住院天数

年龄组	广东		中西部		江苏		陕西	
	均数	中位数	均数	中位数	均数	中位数	均数	中位数
0~	11	9	11	8	12	11	9	7
5~	12	9	15	11	13	12	13	11
10~	14	10	20	14	13	11	15	11
20~	12	8	20	14	14	11	13	9
30~	14	10	21	15	15	12	15	10
40~	15	12	22	16	17	14	15	11
50~	17	13	22	17	18	15	15	11
60~	18	14	23	18	21	16	16	12
70 及以上	19	15	22	17	20	16	15	11

讨 论

自从 1991 年全国县级以上医院全部使用 ICD-9 编码建立病案首页数据库以来, 我国各医院 10 余年来积累了大量的医院诊治病人的住院者病案首页数据库, 这些数据真实、可靠, 蕴藏着大量的有利于疾病防治的有用信息, 是我国疾病统计不可缺少的宝贵资源。然而, 这些数据除用来产生医院日常报表外, 较少得到更充分的利用。本文试图利用这些病案首页数据, 从中挖掘出与住院者年龄有关的信息, 为疾病的防治提供参考。

本研究采用了四个不同的数据库, 尽管其住院者来源地和患者特征有较大差异, 但获得的住院者年龄分布特征有许多相似之处。此外, 对这些数据进行分析得到的部分结果与 1993 年国家卫生服务调查结果^[1]相吻合。说明采用住院者病案首页数据来反映综合性医院的年龄与疾病间的关系具有一定的合理性。

为了进一步深入研究, 可从如下几个方面对本研究作出改进。第一, 本文仅按 ICD-9 的章节将疾病分成 18 类, 为了具体指导某一疾病的防治, 也可采取 ICD-9 将疾病分类成单一病种 (如高血压), 然后分析单病种的住院者比率随年龄变化的规律。第二, 不同年龄的住院者人数, 与各年龄人口数有着密切的关系, 因条件限制, 本研究没有考虑后者, 如果结合医院所在地区的年龄组人口数来反映年龄的分布特征, 得出的结果将更具有说服力。

参 考 文 献

1. <http://www.moh.gov.cn/statistics/ronhs/ronhs16.htm>. 1993 年国家卫生服务调查分析报告.

Exploring the Characteristics of Age Distribution of Inpatients Discharged from General Hospitals Yu Chuanhua, Yu Songlin, et al. Department of Epidemiology and Health Statistics, Public Health School, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei 430030, China

【Abstract】 Objective Explore the characteristics of age distribution of inpatients. **Methods** Calculate chronological age by the difference between admitted hospital date and birth date using more than 550,000 inpatients from medical record database for general hospitals. Classify the disease using ICD-9 codes of the first diagnosis of leaving hospital. **Results** There are a valley and four peaks for the curve of age distribution of inpatients. Except the age 20-49, the frequency of inpatients of the other age for female is less than for male. Many diseases such as tumour, the disease of circulatory system, have themselves characteristics of age distribution. **Conclusion** The data of medical record database contains plentiful information, it is a precious resource for disease statistics, and it seems to be reasonable using it to explore the characteristics of age distribution of inpatients.

【Key words】 Medical record database Data mining Characteristics of age distribution ICD-9