基于AI+大数据的重庆市健康预期寿命测算的方法学探索与实证研究*

郑　艳[1]　石任重[1]　王　宇[1]　黄　鹤[2]　杨　姗[2]　阮晓雯[3]　刘小利[1△]

【提　要】 **目的**　基于世界卫生组织全球疾病负担研究实践经验，探索一种客观、可靠、便捷的健康预期寿命测算方法。**方法**　运用深度自然语言处理、实体对齐、语义算法等人工智能技术，构建全球疾病负担个人健康状态与ICD-10编码对照映射体系，开展健康预期寿命测算并进行实证研究。**结果**　2022年重庆市0岁组健康预期寿命为66.47岁，脑血管疾病、癌症和伤害是导致寿命折损的三大主要疾病类别。**结论**　基于AI+大数据技术构建的健康预期寿命测算方法相对客观、准确，能够为各地开展健康预期寿命测算工作提供参考。

【关键词】 健康预期寿命　全球疾病负担　人工智能　大数据

【中图分类号】 R195.3　　【文献标志码】 A　　**DOI** 10.11783/j.issn.1002-3674.2026.05.021

健康预期寿命将预期寿命与健康水平相结合，是衡量各国国民寿命长度和质量的重要综合性指标。2030年我国人均预期寿命将达到79.0岁，人均健康预期寿命显著提高[1]。这是我国首次明确地将“健康预期寿命”作为政策指标提出，但并没有明确提出健康预期寿命目标值。自1971年美国学者首次提出健康预期寿命的概念以来，健康预期寿命的概念内涵、测算框架及测算方法即成为国内外现代人口健康、社会保障等领域的研究热点[2-6]。但总体说来，学界关于健康预期寿命的基础性研究依然薄弱，尤其是对健康预期寿命的测量方法尚未形成清晰、统一的认识[7]。本文基于目前的相关研究经验[8-10]，运用深度自然语言处理、实体对齐等人工智能和大数据技术，以期探索出一种相对客观、可靠、便捷的健康预期寿命测算方式。

材料与方法

1. 数据来源

基础数据来源主要包括：①人口数据。主要包括2022年全市常住及死亡人口性别、年龄结构等基本信息。其中，常住人口数据来源于重庆市人口与家庭信息系统；死亡人口数据来源于全国疾病监测系统重庆市级平台死因监测数据集。②电子病历数据。主要包括2022年全市医疗卫生机构门诊、住院病历信息，涉及原始数据量1.32亿条。其中，门诊病历数据包括医院、基层医疗卫生机构2类机构，分别来源于重庆市医改监测系统（覆盖全市243家医院）、重庆市基层卫生综合管理系统（覆盖全市1060家卫生院/社区卫生服务中心）。住院病历数据包括医院、基层医疗卫生机构2类机构，均来源于卫生统计网络直报系统（覆盖全市857家医院、1060家卫生院/社区卫生服务中心）。

2. 数据处理

以身份证号、就诊流水号、就诊日期、入院时间、疾病诊断名称、疾病诊断编码等字段作为关键字段开展数据清理工作。对疾病编码、疾病诊断名称皆为空值的记录视为无效数据丢弃处理。对身份证号相同、就诊日期在3日内、疾病诊断名称相同的门诊记录，以第一次就诊记录为准。对以门诊途径入院的住院病历数据，在门诊病历中通过身份证号、就诊日期等字段筛选、剔除。经过多轮去重、清洗，构建本次测算所用的“门诊+住院”基础数据集。

数据提取后，基于患者性别、年龄、疾病诊断名称等重要相关字段，对异常数据进行再清理。针对新生儿与父母共用身份证号的异常病历记录，以疾病诊断名称为依据进行分析处理。疾病诊断ICD-10编码在P00~P96范围内或疾病名称含“新生儿”、“先天性”关键词等情况，年龄则按照“出生年份”处理（即0岁）。疾病诊断可以判断为成人疾病，年龄则按照“提取的出生年份”进行处理。疾病诊断名称无法判断为新生儿或成人的记录（仅1307条，占原始数据量之比不足0.001%），在测算过程中作为垃圾数据丢弃处理。针对非0岁患者的出生年份与提取出生年份不一致的异常记录，以“提取出生年份”作为患者年龄。针对“提取出生年份”提取值格式错误的异常记录，以“出生年份”字段值作为患者年龄。针对年龄出现极值以及年龄与疾病名称逻辑不符的异常记录，在测算过程中则作为垃圾数据丢弃处理。经过处理，本次测算采用的有效门诊数据量8365.6万条，占2022年全市医疗卫生机构总诊疗人次数的42.5%。有效住院病历数据712.7万条，占2022年全市医疗卫生机构出院人数的

* 基金项目：国家卫生健康委员会卫生健康统计工作高质量发展揭榜攻关项目（国卫办规划函〔2022〕336号）

1.重庆市卫生健康统计信息中心（401120）
2.重庆市卫生健康委员会
3.平安科技（深圳）有限公司
△通信作者：刘小利，E-mail：714239564@qq.com

97. 9%。

3. 方法

(1)测算流程

①测算预期寿命。基于人口结构数据,应用人口学、统计学等学科方法,测算0岁组、分性别各岁组的死亡概率、期望寿命。②测算现患率。利用全市常住人口及其性别、年龄结构和电子病历数据,运用AI与大数据技术,测算各年龄段人群的患病率。其中,全市门诊患病规模及疾病谱基于本次提取的有效门诊数据进行估计。③测算健康预期寿命。基于现患率数据,利用GBD2004个人健康状态伤残权重系数工具,计算伤残率、疾病伤残损失人年数等系列指标。最后结合期望寿命指标测算出健康预期寿命(见图1)。

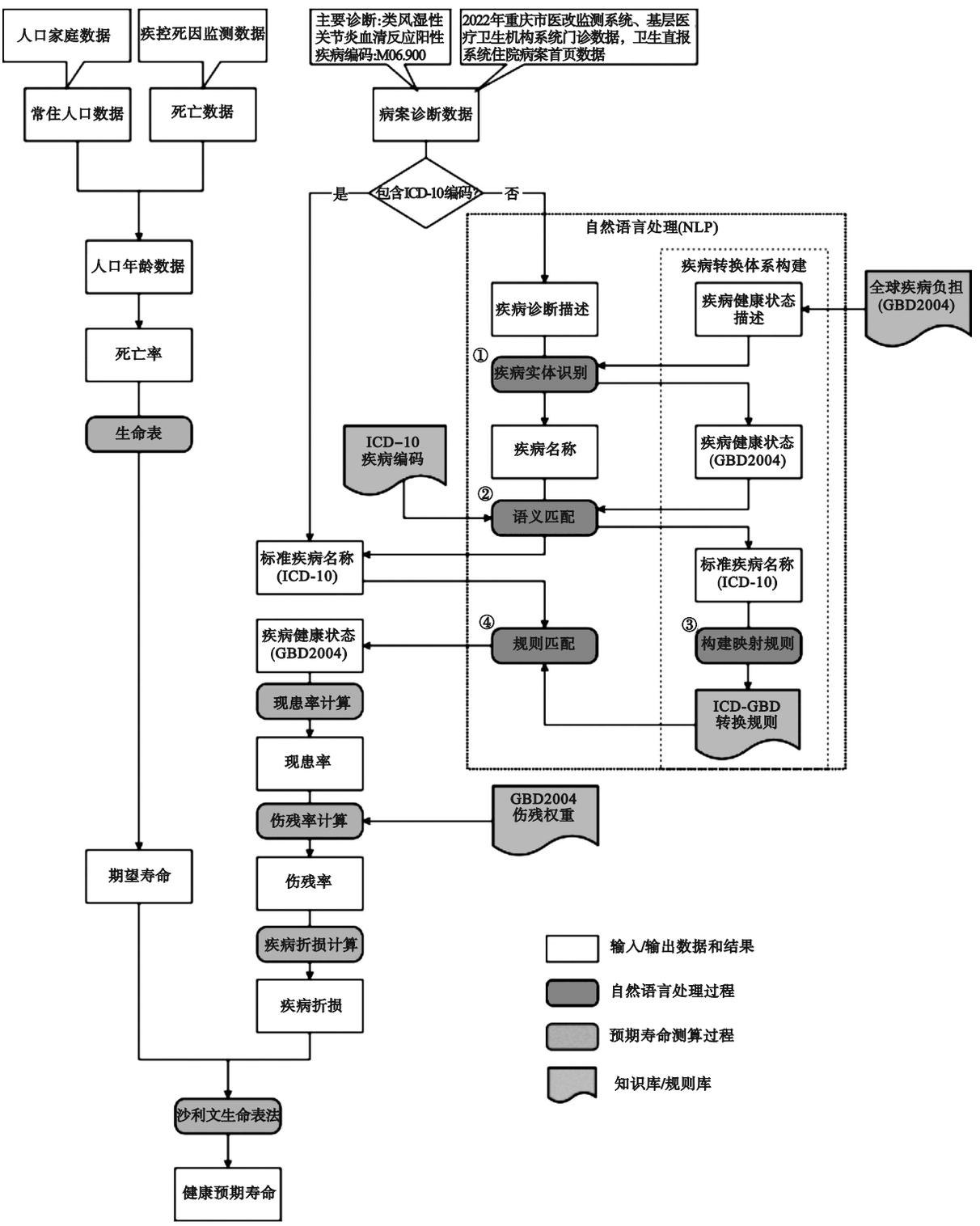


图1 基于“AI+大数据”的健康预期寿命测算流程

（2）全球疾病负担个人健康状态与国际疾病分类ICD-10编码对照体系的构建

由于电子病历门诊记录中患者主诉等数据集、GBD2004个体健康状态均以自由文本形式描述，且现阶段国内外缺乏成熟的GBD2004个人健康状态与ICD-10编码对照体系。针对上述难点，本研究创造性地使用人工智能技术设计了一套通用的疾病信息识别与转换规则，进而利用GBD2004伤残失能权重系数工具开展相关指标测算。首先，通过自然语言处理、实体对齐等人工智能技术提取电子病历记录疾病实体信息。然后，运用语义匹配算法构建国际疾病分类ICD-10代码体系与GBD2004个人健康状态转换规则。最后，将电子病历诊断提取、转换的ICD-10疾病名称与GBD2004个体健康状态进行映射匹配，由此获取相应健康状态的伤残失能权重系数，进而测算各年龄组人群特定疾病的寿命损失年。针对少数无法直接由技术方法建立ICD-10转换的GBD2004个体健康状态，则由临床医学、疾病编码专家进行补充、完善。最终，本文构建了一套覆盖11226个ICD-10疾病编码与254种GBD2004个体健康状态的对照映射表。

（3）健康预期寿命计算方法

健康预期寿命：$DALE_x = e_x - \sum_{k=1}^{K} DLE_{xk}$[11]

其中，e_x为期望寿命，指x岁组人口尚能存活的平均年数；DLE_{xk}为疾病损失年，x岁组人口由于疾病k损失的生命年数。

结　果

1. 健康预期寿命测算结果

2022年重庆市0岁组人口因疾病、伤害导致的疾病折损年为12.46年。其中，男性组损失10.44年，较女性组少4.46年。60岁组人口健康寿命损失7.05年。其中，男性组损失5.92年，较女性组少2.57年。

重庆市0岁组健康预期寿命为66.47岁，男、女性健康预期寿命分别为65.89岁、67.02岁。全市60岁组男性、女性的健康预期寿命分别为14.24岁、15.38岁，见表1。

表1　2022年重庆市健康预期寿命测算结果

年龄（岁）	疾病折损年（年）			健康预期寿命（岁）		
	全人口	男性	女性	全人口	男性	女性
0	12.46	10.44	14.90	66.47	65.89	67.02
1	12.27	10.24	14.73	65.84	65.25	66.37
5	11.06	8.93	13.62	63.12	62.65	63.54
10	10.34	8.19	12.93	58.89	58.44	59.28
15	10.08	7.93	12.65	54.23	53.80	54.63
20	9.88	7.76	12.42	49.54	49.10	49.94
25	9.66	7.60	12.14	44.86	44.38	45.28
30	9.39	7.40	11.77	40.26	39.76	40.73
35	9.10	7.20	11.40	35.68	35.14	36.17
40	8.78	6.97	11.00	31.21	30.67	31.69
45	8.43	6.72	10.53	26.71	26.12	27.25
50	8.09	6.51	10.04	22.40	21.80	22.96
55	7.63	6.25	9.35	18.40	17.77	18.98
60	7.05	5.92	8.49	14.85	14.24	15.38
65	6.32	5.47	7.41	11.33	10.68	11.91
70	5.64	5.08	6.38	8.06	7.39	8.66
75	4.77	4.50	5.18	5.40	4.66	6.04
80	3.79	3.77	3.91	3.67	2.96	4.24
85	2.79	2.98	2.70	2.25	1.58	2.75

2. 导致健康寿命损失的主要疾病类别

2022年重庆市常住人口疾病折损年均值为7.10年。导致疾病折损年最多的前5类疾病依次是脑血管疾病（2.91年）、癌症（0.84年）、伤害（0.47年）、下呼吸道感染（0.36年）、肺结核（0.35年）。分性别看，男性组疾病折损年为7.68年，较女性组多1.05年。导致男、女性疾病折损最严重的前3类疾病依次为脑血管疾病、癌症、伤害，见表2。

表2　2022年重庆市TOP10疾病导致的疾病折损年

顺位	全人口		男性		女性	
	疾病名称	疾病折损年	疾病名称	疾病折损年	疾病名称	疾病折损年
—	合计	7.10	合计	7.68	合计	6.63
1	脑血管疾病	2.91	脑血管疾病	3.23	脑血管疾病	2.66
2	癌症	0.84	癌症	0.94	癌症	0.76
3	伤害	0.47	伤害	0.49	伤害	0.46
4	下呼吸道感染	0.36	下呼吸道感染	0.41	肺结核	0.32
5	肺结核	0.35	肺结核	0.39	下呼吸道感染	0.31
6	骨关节炎	0.23	慢性阻塞性肺疾病	0.23	骨关节炎	0.27
7	上呼吸道感染	0.21	上呼吸道感染	0.20	上呼吸道感染	0.22
8	慢性阻塞性肺疾病	0.16	骨关节炎	0.18	黄斑变性	0.16
9	黄斑变性	0.14	缺血性心脏病	0.14	缺血性心脏病	0.11
10	缺血性心脏病	0.12	黄斑变性	0.12	慢性阻塞性肺疾病	0.10

讨 论

与近年相关的测算结果比较，从纵向上看，本次测算结果低于乔晓春运用2020年重庆市老年人健康自评数据进行测算的结果[12]，但不存在量级差异，可能与近两年疫情对居民健康状况的影响有关。从横向上看，本次测算结果低于石雷对上海市60岁老年人健康预期寿命的测算结果[13]，高于刘红亮对2021年甘肃省60岁老年人口健康预期寿命测算结果[14]，从经济发展水平、医疗资源配置和服务能力供给水平等角度来看，本次测算结果也具有一定的合理性。分性别看，各岁组男性的健康预期寿命均低于女性，这与胡如英等人的测算结果一致[15-16]，可能与男、女身体素质、生理结构不同，男性所承担的社会压力较大，男性生活方式如吸烟、饮酒的比例等均高于女性等因素有关。

本文运用自然语言处理等AI技术和大数据技术，构建全球疾病负担个体健康状态与国际疾病编码ICD-10编码体系的对照映射表，并应用于健康预期寿命测算工作，在国内属于创新性应用。基于电子病历、死因监测数据集及常住人口等多维大数据开展测算工作，与传统上采用健康自评数据的方法相比，可以节约测算工作成本，降低抽样偏差与健康自评的主观性造成的结果偏差，测算结果也相对客观、可靠[17]。但本测算方法还存在一定研究局限，如部分疾病ICD-10编码未能匹配到准确的个体健康状态，健康体检数据、潜在医疗需求等数据尚未纳入基础数据源，在一定程度上会影响测算结果的准确性，在后续研究中有待进一步完善、验证。

参 考 文 献

[1] 新华社. 中共中央国务院印发《“健康中国2030”规划纲要》[EB/OL]. (2016-10-25)[2025-07-28]. https://www.gov.cn/zhengce/2016-10/25/content_5124174.htm.

[2] World Health Organization. The World Health Report 2000. Health Systems: Improving Performance[R]. Geneva,2000.

[3] 蔡玥,吴士勇,陈育德. 健康预期寿命指标的国际应用概述[J]. 中华医学杂志,2023,103(4):229-234.

[4] Oortwijn W, Mathijssen J, Lankhuizen M, et al. Evaluating the Uptake of the Healthy Life Years Indicator[R]. Brussels: Rand Europe,2006.

[5] 刘娜娜,黄烈雨,刘拓,等. 基于Citespace的我国健康预期寿命研究热点和发展脉络分析[J]. 疾病监测,2024,39(11):1509-1516.

[6] 陈鹤. 1990—2013年中国健康调整预期寿命变化的分解研究[J]. 人口研究,2020,44(1):28-38.

[7] 黄匡时. 健康预期寿命的基础性和前瞻性研究[J]. 保险理论与实践,2018(3):43-81.

[8] World Health Organization. The global burden of disease: 2004 update[EB/OL]. (2004-03-02)[2026-01-29]. https://www.who.int/publications/i/item/9789241563710.

[9] 苗鑫蕾,张翔,孟群. 沙利文法中健康期望寿命的测算指标[J]. 中国卫生统计,2019,36 (6):906-908.

[10] 展元元,韩耀风,方亚. 健康预期寿命的测量指标与测算方法的应用及比较[J]. 中国卫生统计, 2020,37 (6):805-808.

[11] 王心旺,方积乾. 健康-疾病负担测量与医疗保险精算方法研究[J]. 中山大学学报论丛,2003,23(6):1-10.

[12] 乔晓春. 全国及各省份老年健康预期寿命变化及差异比较[J]. 人口与经济,2023(5):1-20.

[13] 石雷,罗娟,高凯. 上海市老年人健康预期寿命测算[J]. 医学与社会,2020,33(3):52-54.

[14] 刘红亮,王雁,王玲,等. 老龄人口健康预期寿命测算研究[J]. 西北人口,2024 ,45 (4):88-98.

[15] 胡如英,潘劲,俞敏. 浙江省15岁以上居民自评健康期望寿命[J]. 疾病监测,2010,25(8):665-667.

[16] 肖长春,朋文佳,李晓,等. 2009年合肥市居民期望寿命分析[J]. 中国卫生统计,2012,29(1):84-86.

[17] 乔晓春. 健康寿命研究的介绍与评述[J]. 人口与发展,2009,15(2):53-66.

（责任编辑：张 悦）