

基于贝叶斯网络的复杂共病影响因素分析及风险推理

周焱然¹, 苏银霞³, 殷峰⁴, 古丽加衣娜·艾肯¹, 卢耀勤^{2,1}

1. 新疆医科大学公共卫生学院流行病与卫生统计学教研室, 新疆 乌鲁木齐 830054;

2. 乌鲁木齐市疾病预防控制中心(市卫生监督所); 3. 新疆医科大学医学工程技术学院; 4. 乌鲁木齐市第一人民医院

摘要:目的 探讨复杂共病与影响因素的相关性,并通过网络推理揭示疾病与因素之间的相互作用,以识别高危人群。

方法 基于乌鲁木齐市公共卫生监测数据库和电子病历信息库中 2016 至 2022 年纵向数据,获取研究对象复杂共病的发生情况及相关变量信息。通过最大最小爬山算法结合先验知识进行贝叶斯网络结构学习,采用贝叶斯估计法进行参数学习,利用有向无环图识别混杂因素,指导回归模型的构建。**结果** 共纳入 6 938 名研究对象,12.96% (899/6 938) 在 7 年内发生了复杂共病。筛选出的 6 个预测因子用于模型构建,模型包含 7 个节点和 10 条有向边。结果显示:年龄、性别、来源以及 BMI 与复杂共病的发生直接相关,均为复杂共病发生的父节点。基于 DAG 指导的 logistic 回归结果显示:慢性病患者年龄每增长一岁,其复杂共病患病风险将增加 8.70% [$OR = 1.087 (95\% CI: 1.077 \sim 1.098)$]; 与城市居民相比,农村居民患复杂共病的 $OR = 0.274 (95\% CI: 0.237 \sim 0.317)$; 与体重正常人群相比,肥胖人群患复杂共病的 $OR = 1.019 (95\% CI: 1.008 \sim 1.504)$ 。**结论** 贝叶斯网络能够有效识别复杂共病与影响因素之间的关系及各因素间的相互作用,从而实现对复杂共病发生风险的推理。预防和控制复杂共病,需要关注老龄化、城市环境和肥胖管理。

关键词: 复杂共病; 贝叶斯网络; 风险推理

中图分类号: R195.4 文献标志码: A 文章编号: 1003-8507(2025)02-211-09

DOI: 10.20043/j.cnki.MPM.202409101

Bayesian network – based analysis of factors influencing complex multimorbidity and risk inference

ZHOU Yi-ran*, SU Yin-xia, YIN Feng, Guligiayina·Aiken, LU Yao-qin

* Department of Epidemiology and Health Statistics, School of Public Health, Xinjiang Medical University, Urumqi, Xinjiang 830054, China

Abstract: Objective To explore the correlation between complex multimorbidity and their influencing factors and to reveal the interactions between diseases and factors using network inference methods to identify high-risk populations. **Methods** Based on longitudinal data from 2016 to 2022 in the Urumqi public health surveillance database and electronic medical record information database, this study collected information on the occurrence of complex multimorbidity and related variables. The structure of the Bayesian network was learned using the maximum-minimum hill-climbing algorithm combined with prior knowledge, and parameter learning was conducted using Bayesian estimation. Directed acyclic graphs were employed to identify confounding factors and guide the construction of regression models. **Results** A total of 6 938 participants were included in the study, of which 12.96% (899/6 938) developed complex multimorbidity over the seven-year period. After screening influencing factors, six predictors were selected for model construction, resulting in a model with 7 nodes and 10 directed edges. The results indicated that age, gender, source, and BMI were directly related to the occurrence of complex multimorbidity, all serving as parent nodes in the model. The results of logistic regression based on DAG guidelines showed that the risk of complex multimorbidity would increase by 8.70% [$OR = 1.087 (95\% CI: 1.077 \sim 1.098)$] for each year of age increase in patients with chronic diseases; the OR of complex multimorbidity for rural residents compared to urban residents = 0.274 (95% CI: 0.237-0.317); the OR of complex multimorbidity for obese people compared with the normal weight people = 1.019 (95% CI: 1.008-1.504). **Conclusion** Bayesian networks effectively identify the relationships between complex multimorbidity and influencing factors, as well as the interactions among these factors, thus enabling inference about the risk of complex multimorbidity occurrence. Prevention and control of complex multimorbidity requires attention to aging, urban

基金项目: 新疆维吾尔自治区研究生科研创新项目(XJ2024G170); 新疆维吾尔自治区自然科学基金项目(2024D01E29)

作者简介: 周焱然(1999—), 女, 硕士在读, 研究方向: 医学大数据挖掘

通信作者: 卢耀勤, E-mail: lyq_superior@163.com

environments, and obesity management.

Keywords: Complex multimorbidity; Bayesian network; Risk inference

慢性病已成为全球主要死因之一^[1],根据 2021 年数据,慢性病已导致全球超过 4 300 万人死亡^[2]。随着经济社会发展和人口老龄化,慢性病患病率逐年上升^[3],患者病种累积数目不断增加。研究显示,慢性病累多个器官系统时,患者更容易出现急诊入院、计划外住院和多重用药^[4],从而给个人、社会和医疗系统带来巨大经济负担。为反映身体各系统中常见疾病的异质性^[5],识别治疗和护理需求较高的患者,Harrison 等^[6]提出了复杂共病 (complex multimorbidity) 的概念,即同一个体同时患有三种或三种以上影响不同身体系统的慢性病。复杂共病往往伴随着更沉重的疾病负担^[7]和更恶劣的健康结局^[8],因此,关注复杂共病及其影响因素,有助于识别高危人群并减缓疾病累积。

既往研究常采用 logistic 回归^[4, 9-11]、Cox 比例风险回归^[12]、Poisson 回归等方法^[8, 13]探讨了复杂共病的影响因素,但这些研究存在一定局限性。首先,疾病的影响因素通常并非独立,这违反了模型假设^[14];其次,这些模型无法有效呈现疾病与因素、因素与因素之间的复杂网络关系,缺乏因果证据;最后,这些研究大多集中于发达国家,对于发展中国家,尤其是我国的相关探索相对不足。

在大数据分析中,贝叶斯网络 (Bayesian Network, BN) 的适用性和重要价值得到了广泛认可^[15-17],因其能直观解决不确定性建模和推理问题^[18]。BN 通过有向无环图 (Directed Acyclic Graph, DAG) 编码一组反映随机变量间概率依赖关系的条件概率分布表 (Conditional Probability Table, CPT)^[19],DAG 被认为是因果关系、中介和相互作用相关研究的标准语言^[20],DAG 与 CPT 的结合使得 BN 能以定量和定性的方式描述变量之间的因果关系^[21]。

本研究拟采用 logistic 回归筛选复杂共病影响因素,利用 BN 探索复杂共病与这些因素之间的潜在因果关系,基于 DAG 进行混杂因素识别后,再次进行 logistic 回归建模,并对结果进行综合分析。这一研究将有助于识别复杂共病的高危人群,减缓疾病的累积,并为制定相关干预措施提供理论参考。

1 对象与方法

1.1 研究对象 本研究参照李丽萍等^[22]的研究设计,从新疆省乌鲁木齐市公共卫生监测数据库和电子病历信息库中筛选出 2016 年 1 月 1 日至 2022 年 12 月 31 日期间首次诊断为纳入慢性病的患者。参与者

需前往二级及以上医疗机构就诊或体检,以参与随访。收集的数据包括人口学特征、行为习惯、家族史等基本信息及慢性病诊断信息。纳入标准为:①年龄 ≥ 15 岁;②2016 至 2022 年期间至少参与过一次随访;③随访期间首次诊断至少患有一种本研究纳入的慢性病。排除标准为:①基本信息缺失者;②慢性病诊断记录不全者。最终,共纳入 6 938 例慢性病患者作为研究对象,所有参与者均已签署知情同意书。

1.2 调查内容 通过对乌鲁木齐市公共卫生监测数据库与电子病历信息库内多个表单的关联,获取以下信息:年龄(岁)、性别(男性;女性)、来源(城市;农村)、受教育水平、婚姻状况(未婚;已婚;离异;丧偶)、每日吸烟量(支)、身体质量指数 (Body Mass Index, BMI)、糖尿病家族史(是;否)、高血压家族史(是;否)、龋齿(是;否)及近视(是;否)患病情况。其中,受教育水平被划分为文盲(年满十五周岁且不能识字^[23]),小学(受教育年限 ≤ 6 年),中学($6 <$ 受教育年限 ≤ 9 年),高等教育(受教育年限 > 9 年);BMI 以“体重(kg)/身高(m)²”进行计算,依据中国标准划分为正常($18.5 \leq \text{BMI} < 24$)、超重($24 \leq \text{BMI} < 28$)和肥胖($\text{BMI} \geq 28$)^[24]。

1.3 相关定义 本研究关注的系统及慢性病包括:①运动系统:关节炎、颈椎病、腰椎间盘突出;②消化系统:胆囊炎、胆结石、胃炎、胆管炎、肝囊肿、胆囊息肉;③呼吸系统:结核、支气管炎、肺气肿、哮喘、肺心病、支气管扩张;④泌尿系统:肾囊肿、肾结石、肾积水;⑤循环系统:高血压、冠心病、大动脉粥样硬化、血脂异常、脑卒中;⑥内分泌系统:甲状腺功能亢进、糖尿病;⑦神经系统:严重精神障碍、精神分裂。共涵盖七大系统内的 27 种常见慢性病,疾病系统划分依据国际疾病分类 ICD-10。复杂共病定义为患者同时患有三种或三种以上身体系统内的慢性病。

1.4 贝叶斯网络 BN 通常包括具有 DAG 的 BN 拓扑结构和在已知 DAG 基础上从数据中学习估计节点的 CPT。DAG 能够识别在因果推断框架下回归分析需要控制的协变量^[25],具体识别方法如图 1 所示:假设存在 DAG,其中 X 为解释变量,Y 为结局变量,其余为可能的协变量。路径的关系包括:① $X \rightarrow Y$ (直接因果关系);② $X \rightarrow C \rightarrow Y$ (链式结构);③ $X \leftarrow A \rightarrow B \rightarrow Y$ (叉式结构);④ $X \rightarrow E \leftarrow D \rightarrow Y$ (对撞结构)。在路径①和②中,所有箭头均指向同一方向,表示因果关系;在路径③和④中,箭头方向不统一,表示非因果关系,可能涉及混杂和/或碰撞器。对于路径②,X 通过中介

变量 C 影响 Y, 此时估计 X 对 Y 的总体影响时不应控制 C。对于路径③, A 作为 X 与 Y 的共同原因会引发混淆, 控制 A 或其他相关变量(B)可以消除混淆。对于路径④, 碰撞器 E 会阻塞路径, 对其进行调整可能导致新的混杂。

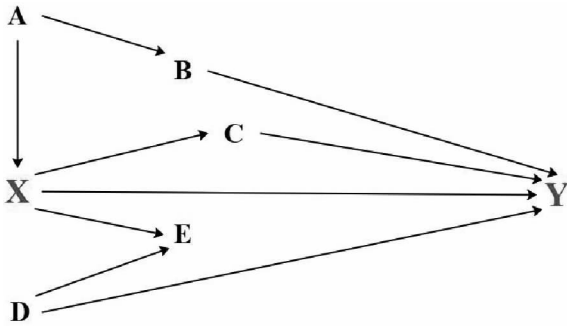


图 1 有向无环图示例

Fig. 1 Example of a directed acyclic graph

1.5 统计学分析 首先, 对于符合正态分布的定量数据, 采用均数 ± 标准差描述; 非正态分布的定量数据则使用中位数和四分位数。定性资料采用频数和构成比进行描述性统计。通过方差分析和 χ^2 检验比较复杂共病患者与非患者在基线特征上的差异。其次, 使用单因素分析和基于前向逐步回归法的多因素二分类 logistic 回归粗略评估影响因素与复杂共病发生之间的关联, 双侧检验 $\alpha = 0.05$ 。第三, 根据数据分布及 logistic 回归分析结果确定贝叶斯网络模型的

变量节点, 利用 R 4. 4. 0 中“bnlearn”包进行贝叶斯网络分析, 通过最大最小爬山算法结合先验知识进行网络结构学习, 并采用贝叶斯估计法进行参数学习。在结构学习中, 为确保网络结构稳定性, 生成 10 000 个网络, 通过 Bootstrap 法计算各边缘及指向的出现频率, 边缘保留阈值设为 85%, 指向保留阈值设为 51% [26], 对网络进行平均以生成最终结构。第四, 将 DAG 中各变量节点作为暴露因素, 识别需要调整的混杂因素, 并在调整混杂因素后进行二次 logistic 回归建模, 选用赤池信息准则 (akaike information criterion, AIC) [27] 及贝叶斯信息准则 (bayesian information criterion, BIC) [28] 比较两次建模的效果。最后, 通过 Netica 软件 (版本号: 7. 01) 实现贝叶斯网络风险推理。

2 结果

2.1 基线情况 本研究共纳入 6 938 名研究对象, 包含男性 2 972 (42. 85%) 名, 女性 3 965 (57. 15%) 名, 中位年龄为 61. 00 (52. 00, 70. 00) 岁, 在特征“日吸烟量分组”、“龋齿”及“糖尿病家族史”中存在变量聚集 (亚组频数占比 > 90%) 的情况。七年间共有 899 名慢性病患者发展为复杂共病, 其中位年龄远高于非复杂共病群体, 70. 00 (63. 00, 75. 00) vs. 59. 00 (51. 00, 69. 00)。复杂共病群体与未患人群在性别、来源、受教育水平、婚姻状况、近视、龋齿及糖尿病家族史方面差异均具有统计学意义 ($P < 0.05$) (表 1)。

表 1 复杂共病与非复杂共病患者基线特征比较

Table 1 Comparison of baseline status of patients with and without complexmultimorbidity

变量	合计 (n = 6 938)	否 (n = 6 039)	是 (n = 899)	统计量	P
年龄 (岁) [M (P ₂₅ , P ₇₅)]	61. 00 (52. 00, 70. 00)	59. 00 (51. 00, 69. 00)	70. 00 (63. 00, 75. 00)	Z = -20. 945	< 0. 001
年龄分组 (岁) [n (%)]				$\chi^2 = 367. 721$	< 0. 001
15 ~ < 25	27 (0. 389)	27 (0. 447)	0 (0. 00)		
25 ~ < 35	199 (2. 868)	198 (3. 279)	1 (0. 111)		
35 ~ < 45	481 (6. 933)	472 (7. 816)	9 (1. 001)		
45 ~ < 55	1 525 (21. 980)	1 438 (23. 812)	87 (9. 677)		
55 ~ < 65	1 795 (25. 872)	1 623 (26. 875)	172 (19. 132)		
≥ 65	2 911 (41. 957)	2 281 (37. 771)	630 (70. 078)		
性别 n (%)				$\chi^2 = 4. 625$	0. 032
男	2 973 (42. 851)	2 558 (42. 358)	415 (46. 162)		
女	3 965 (57. 149)	3 481 (57. 642)	484 (53. 838)		
来源 n (%)				$\chi^2 = 329. 869$	< 0. 001
城市	1 553 (22. 384)	1 140 (18. 877)	413 (45. 940)		
农村	5 385 (77. 616)	4 899 (81. 123)	486 (54. 060)		
受教育水平 n (%)				$\chi^2 = 52. 344$	< 0. 001
文盲	594 (8. 562)	479 (7. 932)	115 (12. 792)		
小学	3 000 (43. 240)	2 583 (42. 772)	417 (46. 385)		
中学	2 868 (41. 338)	2 581 (42. 739)	287 (31. 924)		
高等教育	476 (6. 861)	396 (6. 557)	80 (8. 899)		
婚姻状况 n (%)				$\chi^2 = 36. 531$	< 0. 001
未婚	138 (1. 989)	128 (2. 120)	10 (1. 112)		
已婚	6 183 (89. 118)	5 410 (89. 584)	773 (85. 984)		

(续表)

变量	合计(n=6 938)	否(n=6 039)	是(n=899)	统计量	P
离异	87(1.254)	81(1.341)	6(0.667)		
丧偶	530(7.639)	420(6.955)	110(12.236)		
日吸烟量分组(支)[n(%)]				-	0.147 ^a
<5	6 462(93.139)	5 606(92.830)	856(95.217)		
5~10	69(0.995)	59(0.977)	10(1.112)		
10~<15	140(2.018)	128(2.120)	12(1.335)		
15~<20	36(0.519)	34(0.563)	2(0.222)		
20~<25	197(2.839)	181(2.997)	16(1.780)		
25~<30	6(0.086)	6(0.099)	0(0.00)		
≥30	28(0.404)	25(0.414)	3(0.334)		
BMI 分组 n(%)				$\chi^2 = 2.121$	0.346
正常	2 414(34.794)	2 120(35.105)	294(32.703)		
超重	2 902(41.828)	2 510(41.563)	392(43.604)		
肥胖	1 622(23.378)	1 409(23.332)	213(23.693)		
近视 n(%)				$\chi^2 = 58.656$	<0.001
是	4 547(65.538)	3 856(63.852)	691(76.863)		
否	2 391(34.462)	2 183(36.148)	208(23.137)		
龋齿 n(%)				$\chi^2 = 25.934$	<0.001
是	196(2.825)	147(2.434)	49(5.451)		
否	6 742(97.175)	5 892(97.566)	850(94.549)		
糖尿病家族史 n(%)				$\chi^2 = 4.916$	0.027
是	337(4.857)	280(4.637)	57(6.340)		
否	6 601(95.143)	5 759(95.363)	842(93.660)		
高血压家族史 n(%)				$\chi^2 = 1.158$	0.282
是	1 223(17.628)	1 076(17.818)	147(16.352)		
否	5 715(82.372)	4 963(82.182)	752(83.648)		

注:Z: Mann-Whitney 检验; χ^2 :卡方检验;-: Fisher 精确检验;a: 模拟 P 值;M: 中位数; P_{25} : 第一四分位数; P_{75} : 第三四分位数。

2.2 复杂共病相关影响因素分析 (1)复杂共病的二分类 logistic 回归模型:由于多分类变量“年龄分组”及“日吸烟量分组”个别亚组中频数过低,在模型中转换为定量资料进行处理。不以单因素分析结果作为多因素分析变量纳入标准,以前向逐步回归法进

行探索,结果显示:年龄增长、男性、城市人口、受教育水平低下、未婚、超重及肥胖、存在龋齿、存在糖尿病家族史均会使复杂共病发生风险显著升高($P < 0.05$)(表 2)。

表 2 复杂共病相关影响因素单、多因素 logistic 回归分析结果

Table 2 Results of single and multifactorial logistic regression analysis of influential factors associated with complex multimorbidity

变量	单因素				
	β	SE	Z	P	OR(95% CI)
年龄(岁)	0.084	0.005	16.976	<0.001	1.087(1.077~1.098)
性别					
男(参照)					1.000
女	-0.154	0.072	-2.149	0.032	0.857(0.745~0.987)
来源					
城市(参照)					1.000
农村	-1.295	0.075	-17.370	<0.001	0.274(0.237~0.317)
受教育水平					
文盲(参照)					1.000
小学	-0.397	0.116	-3.407	<0.001	0.672(0.535~0.845)
中学	-0.770	0.121	-6.358	<0.001	0.463(0.365~0.587)
高等教育	-0.173	0.161	-1.075	0.283	0.841(0.614~1.153)
婚姻状况					
未婚(参照)					1.000
已婚	0.604	0.331	1.826	0.068	1.829(0.957~3.496)
离异	-0.053	0.536	-0.099	0.921	0.948(0.332~2.709)
丧偶	1.210	0.345	3.502	<0.001	3.352(1.704~6.597)

(续表)

变量	单因素				
	β	SE	Z	P	OR(95% CI)
每日吸烟量(支)	-0.028	0.010	-2.717	0.007	0.972(0.952~0.992)
BMI 分组					
正常(参照)					1.000
超重	0.119	0.083	1.438	0.150	1.126(0.958~1.324)
肥胖	0.086	0.096	0.895	0.371	1.090(0.903~1.317)
近视					
否(参照)					1.000
是	0.632	0.084	7.565	<0.001	1.881(1.597~2.215)
龋齿					
否(参照)					1.000
是	0.838	0.169	4.956	<0.001	2.311(1.659~3.218)
糖尿病家族史					
否(参照)					1.000
是	0.331	0.150	2.208	0.027	1.392(1.038~1.868)
高血压家族史					
否(参照)					1.000
是	-0.104	0.096	-1.076	0.282	0.902(0.747~1.089)

变量	多因素 ^a				
	β	SE	Z	P	OR(95% CI)
年龄(岁)	0.066	0.006	11.675	<0.001	1.068(1.057~1.080)
性别					
男(参照)					1.000
女	-0.236	0.080	-2.949	0.003	0.790(0.675~0.924)
来源					
城市(参照)					1.000
农村	-0.922	0.084	-11.002	<0.001	0.398(0.337~0.469)
受教育水平					
文盲(参照)					1.000
小学	-0.146	0.123	-1.192	0.233	0.864(0.679~1.099)
中学	-0.282	0.133	-2.123	0.034	0.754(0.581~0.979)
高等教育	-0.145	0.177	-0.818	0.413	0.865(0.611~1.224)
婚姻状况					
未婚(参照)					1.000
已婚	-0.900	0.364	-2.473	0.013	0.407(0.199~0.830)
离异	-1.428	0.563	-2.535	0.011	0.240(0.080~0.723)
丧偶	-0.857	0.386	-2.218	0.027	0.424(0.199~0.905)
每日吸烟量(支)	-0.011	0.011	-1.041	0.298	0.989(0.968~1.010)
BMI 分组					
正常(参照)					1.000
超重	0.186	0.087	2.146	0.032	1.205(1.016~1.429)
肥胖	0.276	0.102	2.711	0.007	1.318(1.080~1.610)
近视					
否(参照)					1.000
是	0.168	0.091	1.841	0.066	1.183(0.989~1.415)
龋齿					
否(参照)					1.000
是	0.650	0.181	3.589	<0.001	1.915(1.343~2.730)
糖尿病家族史					
否(参照)					1.000
是	0.540	0.167	3.238	0.001	1.715(1.237~2.378)
高血压家族史					
否(参照)					1.000
是	0.005	0.106	0.052	0.959	1.006(0.816~1.238)

注:OR 为比值比(Odds Ratio);CI 为置信区间(Confidence Interval);SE 为标准误差(Standard Error);a:AIC=4799.30,BIC=4958.56。

(2)复杂共病的贝叶斯网络模型:结合前文结果,为提高 BN 模型可靠性,剔除高度聚集变量:“日吸烟

量分组”、“龋齿”、“糖尿病家族史”及多因素 logistic 回归结果不显著变量:“近视”、“高血压家族史”,选

取剩余 6 个变量与复杂共病有关的变量构建 BN 模型。如图 2 所示,该模型包含 7 个节点和 10 个有向边,有向边表示节点间存在概率依赖,图中展示了各

节点定义及其先验概率。结果显示:年龄、性别、来源及 BMI 为复杂共病发生的父节点,受教育程度及婚姻状况与复杂共病之间间接相连。

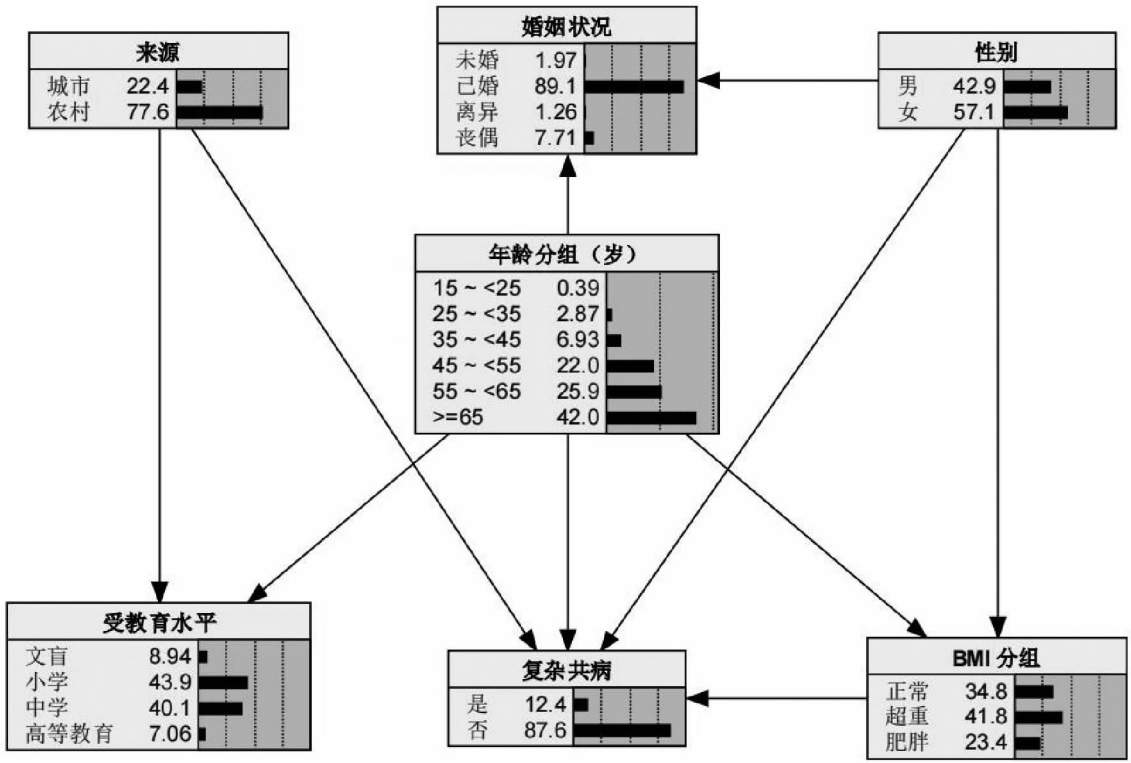


图 2 复杂共病影响因素的有向无环图及先验概率

Fig. 2 DAG and prior probability of complex multimorbidity influences

(3) 基于 DAG 的 logistic 回归分析:依据前文所述协变量识别方法分别找出各暴露因素所需控制的协变量并进行建模,整合结果见表 3:年龄增长、城市

人口、受教育水平低下、未婚、肥胖会使复杂共病发生风险显著升高($P < 0.05$)。

表 3 基于有向无环图的复杂共病相关影响因素 logistic 回归模型结果整合

Table 3 Integration of results from logistic regression modeling of influences related to complex multimorbidity based on directed acyclic graphs

变量	β	SE	Z 值	P 值	OR 值 (95% CI)
年龄(岁) ^a	0.084	0.005	16.980	<0.001	1.087 (1.077 ~ 1.098)
性别 ^b					
男(参照)					1.000
女	-14.23	0.074	-1.926	0.054	0.867 (0.751 ~ 1.003)
来源 ^c					
城市(参照)					1.000
农村	-1.295	0.075	-17.370	<0.001	0.274 (0.237 ~ 0.317)
受教育水平 ^d					
文盲(参照)					1.000
小学	-0.306	0.120	-2.544	0.011	0.736 (0.583 ~ 0.935)
中学	-0.755	0.125	-6.046	<0.001	0.471 (0.369 ~ 0.602)
高等教育	-0.578	0.167	-3.453	<0.001	0.561 (0.403 ~ 0.309)
婚姻状况 ^e					
未婚(参照)					1.000
已婚	-1.130	0.363	-3.112	0.002	0.323 (0.165 ~ 0.695)
离异	-1.601	0.562	-2.849	0.004	0.202 (0.063 ~ 0.595)

(续表)

变量	β	SE	Z 值	P 值	OR 值 (95% CI)
丧偶	-0.992	0.385	-2.577	0.010	0.371 (0.181 ~ 0.828)
BMI 分组^f					
正常 (参照)					1.000
超重	0.160	0.085	1.879	0.060	1.173 (0.994 ~ 1.386)
肥胖	0.214	0.099	2.154	0.031	1.019 (1.008 ~ 1.504)

注:OR 为比值比 (Odds Ratio);CI 为置信区间 (Confidence Interval);SE 为标准误差 (Standard Error);a:无调整,AIC = 4 953.18,BIC = 4 966.87;b:调整年龄、BMI,AIC = 4 949.94,BIC = 4 984.17;c:无调整,AIC = 5 067.49,BIC = 5 081.18;d:调整来源,AIC = 5 025.03,BIC = 5 059.25;e:调整年龄、性别,AIC = 4 946.86,BIC = 4 987.93;f:调整年龄、性别,AIC = 4 949.94,BIC = 4 984.17。

2.3 复杂共病发生的风险推理 BN 能从已知的节点状态推断出未知节点的概率,从而实现复杂共病的风险推理。将前文基于 DAG 的 logistic 回归分析中结果显著的节点变量设定为已知(即各节点某定义的先

验概率为 100%),如图 3 所示,如果一个慢性病患者特征为:年龄 ≥ 65 岁、城市人口、未婚、文盲、肥胖,则该患者发生复杂共病的概率会上升至 41.40%。

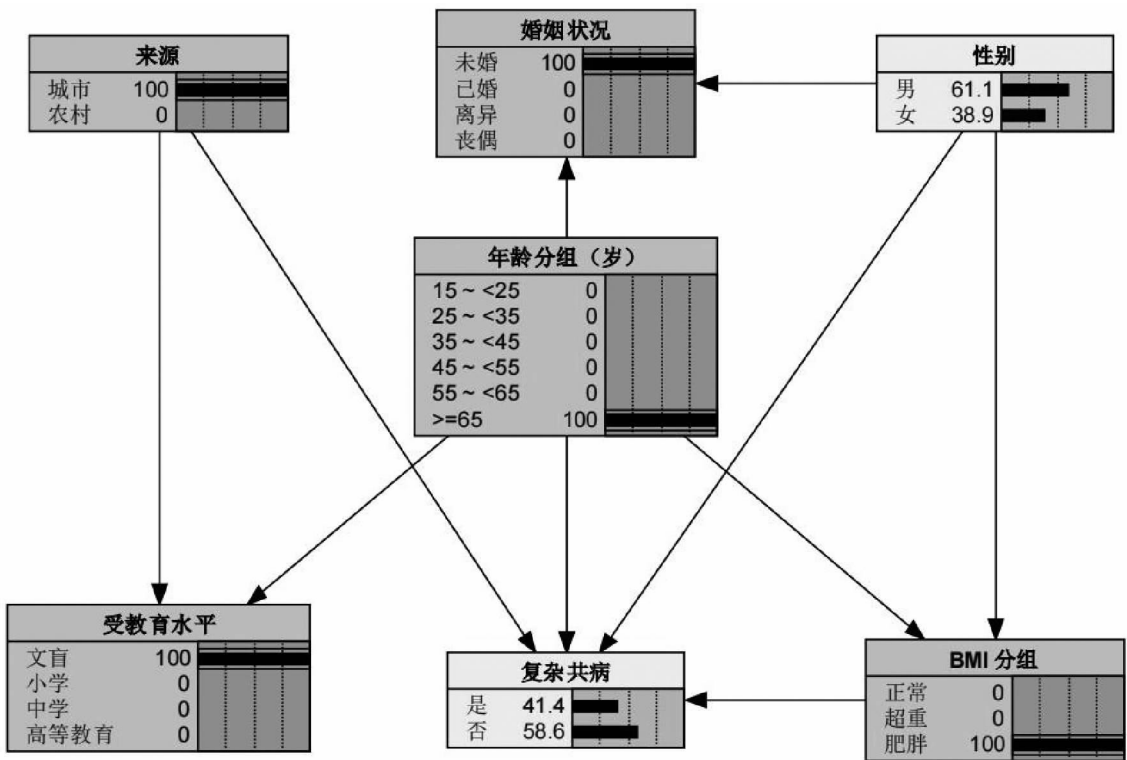


图 3 已知部分特征下的复杂共病风险推理

Fig. 3 Inference of complex multimorbidity risk under known partial characteristics

3 讨论

本研究构建了一个预测慢性病患者复杂共病风险的贝叶斯网络 (BN) 模型。首先,选择了七大系统中的 27 种常见慢性病定义复杂共病,并纳入了 11 个可能影响因素。通过 logistic 回归模型结合数据分布筛选,确定了 6 个主要预测因子:年龄、性别、来源、受教育程度、婚姻状况及 BMI 分组。基于 BN 的拓扑结构 DAG 识别了混杂因素,并在此基础上进行 logistic 回归的二次建模,结果显示年龄、来源、受教育水平、

婚姻状况和 BMI 与复杂共病显著相关。最终,将各节点中最劣势特征的先验概率设定为 100% 进行风险推理。

研究表明,年长的慢性病患者更容易发生复杂共病,这与老龄化密切相关。老龄化是多种疾病的重要风险因素^[29],随着年龄的增长,身体功能衰退使得慢性病累积风险增加。城市居民发生复杂共病的可能性高于农村居民,这可能与城市生活方式中的不良因素如不健康饮食、久坐不动等有关^[30],提示城市环境中的生活方式干预对于减少慢性病和复杂共病

的发生具有重要意义。此外,本研究显示肥胖与复杂共病的发生有显著关联。肥胖不仅与无病寿命的减少 3 ~ 8 年相关,还增加了过早死亡的风险 1.3 倍^[31]。肥胖被认为是多种慢性疾病的危险因素,包括心脏代谢疾病、消化系统疾病、呼吸系统疾病、神经系统疾病、肌肉骨骼疾病以及感染性疾病^[12]。肥胖对健康的负面影响通过多种机制作用于身体多个系统,增加了多种慢性病的风险,从而进一步增加复杂共病的发生概率,强调了肥胖管理在慢性病预防和控制中的重要性。

本研究基于 DAG 识别回归分析中需要控制的协变量,并进行了 logistic 回归的二次建模,结果显示与一次建模有所不同,这可能归因于 DAG 对控制变量的准确识别。BN 补充了 logistic 回归的不足,能揭示多因素间的复杂关系,并在已知先验概率的情况下对复杂共病进行风险推理。Logistic 回归显示年龄、来源、受教育水平、婚姻状况和 BMI 与复杂共病相关,而 BN 发现年龄、性别、来源和 BMI 直接相关,受教育程度和婚姻状况可能仅间接影响复杂共病。因此,BN 能够识别及各因素在疾病发生过程中所起的作用,能在一定程度上为复杂共病的防治提供科学依据。

该研究也存在部分局限性。首先,复杂共病的发生是多因素共同作用的结果。尽管我们根据现有研究选择了一些关键变量,仍可能存在其他未被纳入的变量与复杂共病的发生相关。其次,BN 的构建是数据驱动与先验知识结合的结果。尽管我们采用了统计方法尽可能确保模型的稳定性,但由于变量选择的局限性,模型结果可能与真实情况存在一定的偏差。

未来研究应扩大样本规模,收集更全面的因素数据,应用动态 BN 模型深入探讨因果关系,并关注不同干预措施的实际效果,以提供更精准的风险预测和干预策略,为高危人群识别、公共卫生政策制定和个体化干预提供理论支持。

利益冲突声明 本研究不存在任何利益冲突

参考文献

[1] Skou ST, Mair FS, Fortin M, et al. Multimorbidity [J]. *Nature Reviews. Disease Primers*, 2022, 8(1): 48.

[2] Ho ISS, Azcoaga - Lorenzo A, Akbari A, et al. Variation in the estimated prevalence of multimorbidity: systematic review and meta - analysis of 193 international studies [J]. *BMJ Open*, 2022, 12(4): e057017.

[3] 曹新西,徐晨婕,侯亚冰,等. 1990—2025 年我国高发慢性病的流行趋势及预测 [J]. *中国慢性病预防与控制*, 2020, 28(1): 14 - 19.

Cao XX, Xu CJ, Hou YB, et al. The epidemic trend and prediction of chronic diseases with high incidence in China from 1990 to 2025 [J]. *Chinese Journal of Prevention and Control of Chronic Diseases*, 2020, 28(1): 14 - 19. (In Chinese)

[4] Sinha A, Kerketta S, Ghosal S, et al. Multimorbidity and complex multimorbidity in India: findings from the 2017 - 2018 longitudinal ageing study in India (LASI) [J]. *International Journal of Environmental Research and Public Health*, 2022, 19(15): 9091.

[5] Nicholson K, Makovski TT, Griffith LE, et al. Multimorbidity and comorbidity revisited: refining the concepts for international health research [J]. *Journal of Clinical Epidemiology*, 2019, 105: 142 - 146.

[6] Harrison C, Britt H, Miller G, et al. Examining different measures of multimorbidity, using a large prospective cross - sectional study in Australian general practice [J]. *BMJ Open*, 2014, 4(7): e004694.

[7] Wang ZJ, Peng WJ, Li MY, et al. Association between multimorbidity patterns and disability among older People covered by long - term care insurance in Shanghai, China [J]. *BMC Public Health*, 2021, 21(1): 418.

[8] Rodrigues APDS, Batista SRR, Santos ASEA, et al. Multimorbidity and complex multimorbidity in Brazilians with severe obesity [J]. *Scientific Reports*, 2023, 13(1): 16629.

[9] Kato D, Kawachi I, Saito J, et al. Complex multimorbidity and incidence of Long - Term care needs in Japan: a prospective cohort study [J]. *International Journal of Environmental Research and Public Health*, 2021, 18(19): 10523.

[10] Petarli GB, Cattafesta M, Sant'anna MM, et al. Multimorbidity and complex multimorbidity in Brazilian rural workers [J]. *PLOS One*, 2019, 14(11): e0225416.

[11] Kato D, Kawachi I, Kondo N. Complex multimorbidity and working beyond retirement age in Japan: a prospective propensity - matched analysis [J]. *International Journal of Environmental Research and Public Health*, 2022, 19(11): 6553.

[12] Kivimäki M, Strandberg T, Pentti J, et al. Body - mass index and risk of obesity - related complex multimorbidity: an observational multicohort study [J]. *The Lancet. Diabetes & Endocrinology*, 2022, 10(4): 253 - 263.

[13] Sugiyama Y, Mutai R, Aoki T, et al. Multimorbidity and complex multimorbidity, their prevalence, and associated factors on a remote island in Japan: a cross - sectional study [J]. *BMC Primary Care*, 2022, 23(1): 258.

[14] 王齐里,宋文柱,张岩波,等. 贝叶斯网络在老年抑郁症危险因素中的应用——基于 CHARLS 数据库的实证分析 [J]. *现代预防医学*, 2023, 50(20): 3649 - 3655, 3662.

Wang QL, Song WZ, Zhang YB, et al. Applications of Bayesian network in risk factors of senile depression - empirical analysis based on CHARLS database [J]. *Modern Preventive Medicine*, 2023, 50(20): 3649 - 3655, 3662. (In Chinese)

[15] Song WZ, Gong H, Wang QL, et al. Using bayesian networks with Max - Min Hill - Climbing algorithm to detect factors related to multimorbidity [J]. *Frontiers in Cardiovascular Medicine*, 2022, 9: 984883.

[16] Ke XJ, Keenan K, Smith VA. Treatment of missing data in Bayesian network structure learning: an application to linked biomedical and social survey data [J]. *BMC Medical Research Methodology*, 2022, 22(1): 326.

[17] Zhu Z, Xing W, Hu Y, et al. Paradigm shift: Moving from symptom clusters to symptom networks [J]. *Asia - Pacific Journal of Oncology Nursing*, 2021, 9(1): 5 - 6.

[18] Sieswerda MS, Bermejo I, Geleijnse G, et al. Predicting lung cancer survival using probabilistic reclassification of TNM editions with a Bayesian network [J]. *JCO Clinical Cancer Informatics*,

- 2020, 4: 436 – 443.
- [19] Guo ZG, Constantinou AC. Approximate learning of high dimensional bayesian network structures via pruning of candidate parent Sets[J]. Entropy, 2020, 22(10): 1142.
- [20] Digitale JC, Martin JN, Glymour MM. Tutorial on directed acyclic graphs[J]. Journal of Clinical Epidemiology, 2022, 142: 264 – 267.
- [21] Moe SJ, Carriger JF, Glendell M. Increased use of bayesian network models has improved environmental risk assessments[J]. Integrated Environmental Assessment and Management, 2021, 17(1): 53 – 61.
- [22] 李丽萍, 廖婧, 高鑫源, 等. 中国共病加权指数与老年人卫生服务利用的关联性研究[J]. 中国全科医学, 2025, 28(1): 65 – 70. Li LP, Liao J, Gao XY, et al. Association between the Chinese multimorbidity – weighted index and health service utilization among the elderly in China[J]. Chinese General Practice, 2025, 28(1): 65 – 70. (In Chinese)
- [23] Petri M, Messinis L, Patrikelis P, et al. Illiteracy, neuropsychological assessment, and cognitive rehabilitation: a narrative review [J]. Advances in Experimental Medicine and Biology, 2023, 1425: 477 – 484.
- [24] Zhou BF, Cooperative Meta – Analysis Group of the Working Group on Obesity in China. Predictive values of body mass index and waist circumference for risk factors of certain related diseases in Chinese adults – – study on optimal cut – off points of body mass index and waist circumference in Chinese adults [J]. Biomedical and Environmental Sciences, 2002, 15(1): 83 – 96.
- [25] 吴孟泽. 有向无环图在构建 logistic 预测模型中的应用研究[J]. 科学技术创新, 2023, (3): 63 – 66. Wu MZ. Research on the application of directed acyclic graph in the construction of logistic prediction model [J]. Scientific and Technological Innovation, 2023, (3): 63 – 66. (In Chinese)
- [26] 李承龙, 郭海辉, 陈维, 等. 青少年黑暗三人格的网络结构: 基于高斯图和有向无环图的探索[J]. 中国临床心理学杂志, 2023, 31(6): 1491 – 1495. Li CL, Guo HH, Chen W, et al. The network structure of the dark triad in adolescents: an exploration based on Gaussian and directed acyclic graphs[J]. Chinese Journal of Clinical Psychology, 2023, 31(6): 1491 – 1495. (In Chinese)
- [27] Ikemoto K, Takahashi K, Ozawa T, et al. Akaike’s information criterion for stoichiometry inference of supramolecular complexes [J]. Angewandte Chemie (International ed. in English), 2023, 62(14): e202219059.
- [28] Selig K, Shaw P, Ankerst D. Bayesian information criterion approximations to Bayes factors for univariate and multivariate logistic regression models [J]. The International Journal of Biostatistics, 2020, 17(2): 241 – 266.
- [29] Shi JK, Guo YB, Li Z, et al. Sociodemographic and behavioral influences on multimorbidity among adult residents of northeastern China[J]. BMC Public Health, 2022, 22(1): 342.
- [30] Fleitas Alfonso L, King T, You E, et al. Theoretical explanations for socioeconomic inequalities in multimorbidity: a scoping review [J]. BMJ Open, 2022, 12(2): e055264.
- [31] Global BMI Mortality Collaboration, Diangelantonio E, Bhupathiraju S, et al. Body – mass index and all – cause mortality: individual – participant – data meta – analysis of 239 prospective studies in four continents[J]. Lancet, 2016, 388(10046): 776 – 786.

收稿日期: 2024-09-05

(上接第 197 页)

- [14] 杨凌麟, 陈宇洁, 王怡, 等. 不同海拔地区不同年龄、性别健康成人睡眠监测指标差异性及其临床意义研究[J]. 中国全科医学, 2024, 27(24): 2961 – 2968. Yang LL, Chen YJ, Wang Y, et al. Variations and clinical significance of sleep monitoring indicators among healthy adults of different Ages and genders in various altitudes[J]. Chinese General Practice, 2024, 27(24): 2961 – 2968. (In Chinese)
- [15] Li NY, Wang YF, Deng Q, et al. A multilevel physical activity intervention among Chinese rural older adults (stay active while aging): a study protocol for a clustered randomized controlled trial [J]. Frontiers in Public Health, 2022, 10: 760457.
- [16] Murcia – Lesmes D, Domínguez – López I, Laveriano – Santos EP, et al. Association between tomato consumption and blood pressure in an older population at high cardiovascular risk; observational analysis of PREDIMED trial [J]. European Journal of Preventive Cardiology, 2024, 31(8): 922 – 934.
- [17] Vaughan K, Miller WC. Validity and reliability of the Chinese translation of the Physical Activity Scale for the Elderly (PASE) [J]. Disability and Rehabilitation, 2013, 35(3): 191 – 197.
- [18] Ngai SPC, Cheung RTH, Lam PL, et al. Validation and reliability of the Physical Activity Scale for the Elderly in Chinese population [J]. Journal of Rehabilitation Medicine, 2012, 44(5): 462 – 465.
- [19] Shahid A, Wilkinson K, Marcu S, et al. STOP, THAT and One Hundred Other Sleep Scales[M]. New York: Springer, 2012.
- [20] Tsai PS, Wang SY, Wang MY, et al. Psychometric evaluation of the Chinese version of the Pittsburgh Sleep Quality Index (CPSQI) in primary insomnia and control subjects [J]. Quality of Life Research, 2005, 14(8): 1943 – 1952.
- [21] Nam GE, Han K, Kim DH, et al. Sleep duration is associated with body fat and muscle mass and waist – to – height ratio beyond conventional obesity parameters in Korean adolescent boys [J]. Journal of Sleep Research, 2017, 26(4): 444 – 452.
- [22] Kohanmoo A, Kazemi A, Zare M, et al. Gender – specific link between sleep quality and body composition components: a cross – sectional study on the elderly [J]. Scientific Reports, 2024, 14(1): 8113.
- [23] Kocovska D, Lysen TS, Dotinga A, et al. Sleep characteristics across the lifespan in 1.1 million People from the Netherlands, United Kingdom and United States: a systematic review and meta – analysis [J]. Nature Human Behaviour, 2021, 5(1): 113 – 122.
- [24] Qin QR, Jiang GQ, Mei YN, et al. Changes and predictors of sleep quality with a 2 – year follow – up in older adults according to living arrangements [J]. Archives of Gerontology and Geriatrics, 2022, 102: 104753.
- [25] Tao ZH, Feng YT, Liu J, et al. Trends and disparities in sleep quality and duration in older adults in China from 2008 to 2018: A National observational study [J]. Frontiers in Public Health, 2023, 11: 998699.

收稿日期: 2024-09-03