



# 团 体 标 准

T/CAS 1152—2025

## 医学观察性研究有向无环图的构建指南

Guidance for constructing directed acyclic graph in  
observational study

2025-11-27 发布

2025-11-27 实施

中国标准化协会 发布

中国标准化协会（CAS）是组织开展国内、国际标准化活动的全国性社会团体。制定中国标准化协会标准（以下简称：标协标准），满足市场需要，增加标准的有效供给，是中国标准化协会的工作内容之一。国内外相关组织和个人均可提出制修订标协标准的建议并参与有关工作。

标协标准按《中国标准化协会团体标准管理办法》进行制定和管理。

在本文件实施过程中，如发现需要修改或补充之处，请将意见和有关资料寄给中国标准化协会，以便修订时参考。

本文件版权为中国标准化协会所有，除了用于国家法律或事先得到中国标准化协会的许可外，不得以任何形式或任何手段复制、再版或使用本文件及其章节，包括电子版、影印件，或发布在互联网及内部网络等。

中国标准化协会地址：北京市海淀区增光路 33 号中国标协写字楼  
邮政编码：100048 电话：010-88416788 传真：010-68486206  
网址：www.china-cas.org 电子信箱：cas@china-cas.org

## 目 次

前言.....	III
引言.....	IV
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 缩略语.....	2
5 构建原则.....	2
6 构建流程.....	3
7 应用策略.....	6
参考文献.....	7

中国标准化协会标准  
管理及服务平台

## 前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第 1 部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件起草单位：北京大学、北京大学人民医院、四川大学华西医院、中国医学科学院医学信息研究所、中国医学科学院、首都医科大学附属北京儿童医院、北京大学第一医院、首都医科大学附属北京天坛医院、复旦大学、深圳理工大学。

本文件主要起草人：杜建、刘慧鑫、史轩宇、李舍予、陈松景、张敬晨、严若华、张斐斐、朱之恺、王安心、朱政、杨羽、杨智荣、邓光辉。

中国标准化协会标准  
管理及服务平台

## 引 言

在基于医学观察性研究数据进行因果推断时，未经充分调整的混杂，或错误调整中介变量与碰撞变量，均会严重干扰对暴露（自变量）与结局（因变量）之间真实因果关系的估计。有向无环图是一种用于描述变量间潜在因果关系的图形工具，广泛应用于医学观察性研究。通过构建有向无环图，研究者可以清晰地识别混杂变量、中介变量及碰撞变量，以避免不充分的调整和过度调整带来的偏倚。

然而，现有有向无环图的构建往往依赖研究者的个人经验或零散证据，缺乏一套科学、透明及可操作的统一流程。近年来，尽管流行病学方法学专家多次在国际著名期刊中呼吁在观察性研究中应用有向无环图进行因果推断，但关于如何科学地收集和整合文献证据，进而构建标准化、可复现的有向无环图，学界尚未达成共识。

为此，本文件提出一套有向无环图构建的规范化流程，帮助临床研究者在医学观察性研究中构建并报告标准化、可循证的有向无环图，以指导基于观察性研究数据制定更科学的统计分析策略。

中国标准化协会标准  
管理及服务平台

# 医学观察性研究有向无环图的构建指南

## 1 范围

本文件确立了医学观察性研究有向无环图的构建原则，描述了构建流程和应用方法。

本文件涉及的有向无环图本身并非用于在观察性数据中估计暴露对结局因果关系的统计学方法，而是作为概念工具，用于揭示变量间关系及潜在非因果路径，为判断偏倚（尤其是混杂偏倚）并为后续数据分析中变量纳入的决策提供依据。

本文件适用于医学观察性研究中有向无环图的构建及应用。

## 2 规范性引用文件

本文件没有规范性引用文件。

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

**有向无环图** directed acyclic graph; DAG

由节点和带有箭头的有向边所组成、形成有向路径且不包含闭环的图。

注1：有向无环图用于描述变量之间的因果结构，有助于描述潜在的偏倚来源，确认研究中的关键变量和最小充分调整集，从而更清晰地描述特定暴露或治疗对健康结局的因果效应。

注2：节点表示变量，箭头表示变量间因果关系的方向（原因→结果），一个箭头代表一个变量对另一变量的直接因果效应。

注3：若箭头从变量X出发指向变量Y（ $X \rightarrow Y$ ），如果中间无其他变量，则X是Y的父节点，Y是X的子节点。

### 3.2

**观察性研究** observational study

根据特定研究问题，不施加主动干预的、以自然人群或临床人群为对象的、探索暴露或治疗与结局因果关系的研究。

### 3.3

**真实世界数据** real-world data; RWD

来源于日常所收集的各种与患者健康状况和/或诊疗及保健有关的数据。

注：并非所有的真实世界数据经分析后都能成为真实世界证据，只有满足适用性的真实世界数据才有可能产生

真实世界证据。

### 3.4

#### 混杂变量 confounder

同时影响暴露和结局、但不在二者中间路径上的变量。

注：暴露和结局的共同原因，未充分调整混杂变量会导致混杂偏倚。

### 3.5

#### 因果推断 causal inference

基于真实世界数据，刻画暴露或干预与临床结局或健康结局的因果关系路径，充分考虑各种协变量和已测或未测混杂因素的影响，并控制可能的偏倚，采用恰当的统计模型和分析方法，做出暴露或干预与临床结局或健康结局的因果关系的推断结论。

### 3.6

#### 中介变量 intermediate variable

处于从暴露到结局的因果路径上、既受暴露或干预影响、同时又影响结局的变量。

### 3.7

#### 碰撞变量 collider

同时受暴露和结局影响的变量。

注：暴露和结局的共同结果。错误调整碰撞变量会引入碰撞偏倚，导致虚假关联。

### 3.8

#### 最小充分调整集 minimally sufficient adjustment set

为尽可能阻断暴露与结局间的非因果路径而必须在统计模型中调整的一组协变量。

注：该集合通常包括混杂变量，但不含中介变量或碰撞变量，以期在减少偏倚的同时避免过度调整引起的估计偏差或统计功效降低。

## 4 缩略语

下列缩略语适用于本文件。

ICD：国际疾病分类（International Classification of Diseases）

MeSH：医学主题词（Medical Subject Headings）

PECO：人群-暴露-对照-结局（Population-Exposure-Comparator-Outcome）

UMLS：一体化医学语言系统（Unified Medical Language System）

## 5 构建原则

### 5.1 系统性

需进行充分且可复制的文献检索，识别所有与因果效应有关的潜在变量。在医学观察性研究的设计阶段即开始构建有向无环图，而不是仅在数据分析阶段。除混杂变量外，中介变量、碰撞变量也体现在有向无环图中，避免对中介变量进行不恰当的调整，以及避免因调整碰撞变量或选择特定

研究对象而引入碰撞偏倚。

## 5.2 可循证

有向无环图中的每个节点和每条边均需有循证医学证据支持。

## 5.3 一致性

根据本文件推荐的流程，确保针对同一个研究问题，不同研究者构建的有向无环图是一致的。

# 6 构建流程

## 6.1 概述

DAG 的构建本质是对多个变量间关系的文献证据进行综合，可参考系统综述的相关指南。本文件针对医学观察性研究中 DAG 的构建，参考《Cochrane 干预措施系统评价手册》，提出以下六个步骤：明确研究问题、全局证据检索、证据纳入与排除、信息提取、有向无环图构建、有向无环图优化。各步骤环环相扣，以实现有向无环图构建和报告的标准化、透明化和可循证。

## 6.2 明确研究问题

采用 PECO 框架明确研究目的和关键变量。例如“在特定人群中，暴露于 X（相对于不暴露）是否与结局 Y 的风险增加/降低有关”。该表述有助于将研究问题拆解为基本要素，确保后续检索与分析不偏离研究目标。

## 6.3 全局证据检索

根据已明确的研究问题，制定针对暴露与结局的检索策略，进行全局检索（文献并集而非交集），包括：

- a) 围绕暴露 X 的研究，用于寻找 X 的父节点和子节点；
- b) 围绕结局 Y 的研究，用于寻找 Y 的父节点和子节点；
- c) 同时涉及 X 与 Y 的研究，直接探讨 X 与 Y 关系。

以上检索既包括了焦点关系（即 X 和 Y）的证据，也包含了 X 或 Y 各自的影响因素和结局事件的证据，保证“前因后果”的全局检索。

## 6.4 证据纳入与排除

### 6.4.1 纳入标准

纳入证据宜包括以下内容。

- a) 根据循证医学证据金字塔，纳入来自不同证据等级的研究结果。
- b) 若存在相关系统综合和荟萃分析，宜优先纳入，以提高证据可靠性。存在多个荟萃分析时，可进行伞式综述，整合现有荟萃分析。
- c) 需阅读摘要或全文，提取与变量间效应相关的结论性陈述。例如结论部分“暴露于 X 会显著增加/降低结局 Y 的风险...”，以及背景部分“X 是广泛认可的 Y 的危险因素”等类似表述。

## 6.4.2 排除标准

排除以下文献类型。

- a) 评论、观点、社论、研究方案等无实证结果或仅提出理论假设的文献。
- b) 未明确报告与暴露或结局存在效应关系的文献证据。
- c) 与研究问题定义的人群特征差异过大的文献证据。

## 6.5 信息提取

### 6.5.1 指标提取与结构化

提取暴露与结局关系的效应指标、效应量、置信区间、效应方向等信息，并借助 PECO 框架将不同研究结果统一表示为 PECO 格式，例如“在人群 P 中，暴露 E（相较于对照 C）会显著增加结局 O 的风险”。可采用大语言模型等前沿人工智能工具和人工审核相结合的方式。

### 6.5.2 术语统一

将不同文献中出现的同义或不同层级的变量统一命名，例如“脑卒中”和“中风”可映射到同一编码下，并对节点进行合并。可借助常用医学术语编码体系，例如 MeSH、ICD、UMLS 等，按照研究问题将同义或相关上下层级概念进行术语映射与统一。通过这种映射，能够将表述不同但含义相同的变量合并，保证 DAG 中各节点均具有唯一、明确的含义。

## 6.6 DAG 构建

### 6.6.1 根据循证医学证据构建初始有向无环图

经过以上明确研究问题、全局证据检索、证据纳入与排除、信息提取四个过程，形成了拟研究的暴露-结局各自的成因和后果变量，以及支持证据的结构化表格。根据图 1 和以下步骤，构建初始 DAG。

- a) 识别暴露的父节点和子节点集合 V1。
- b) 识别结局的父节点和子节点集合 V2。
- c) 识别同时属于 V1 和 V2 的变量集合，即两者的交集 ( $V1 \cap V2$ )，根据箭头方向可找出共同原因、中间变量、共同结果（见图 1）。
- d) 用有向边连接有文献支持的变量，构建初始 DAG，并为每一条边建立“边索引”，标注支持文献及结论。
- e) 构建的初始 DAG 涉及变量原则上宜包括四大类：
  - 1) 人口统计学变量（性别、年龄、社会经济水平等）；
  - 2) 合并症或疾病变量；
  - 3) 用药变量；
  - 4) 生活方式和行为变量。

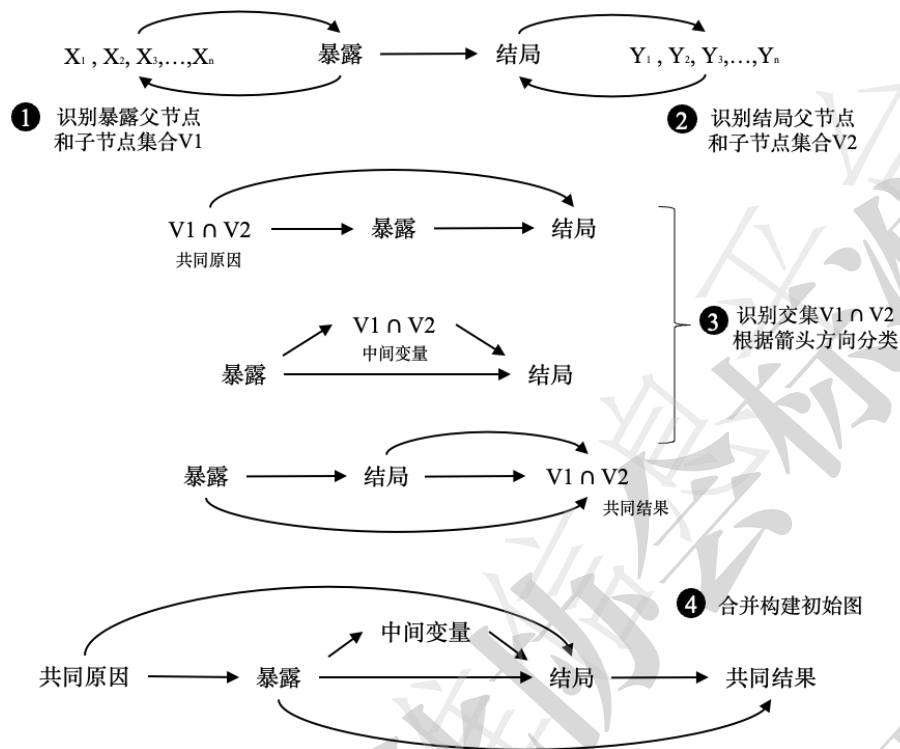


图1 构建初始 DAG 流程示意图

### 6.6.2 根据专家经验对节点和关系进行补充

上述初始 DAG 已确定暴露变量与结局变量的共同父节点、共同子节点，以及从暴露到结局的中间节点。可根据以下步骤对初始 DAG 中涉及的变量进行两两关系的补充。

- 对于初始 DAG 中任意两个变量之间是否应添加带箭头的连线，可依据专家经验进行判断。
- 若专家经验无法确定某些变量间的关系，宜借助国内外权威的知识查询平台，获取两个变量间是否存在影响关系的文献。

## 6.7 DAG 优化

### 6.7.1 基于因果关系准则处理冲突关系

由于医学研究的异质性，以上文献检索和处理过程会产生结果冲突甚至反向因果等问题，可基于以下因果关系准则，结合现有的可用数据进行检查和处理。

- 时间顺序：暴露发生先于结局。
- 生物学合理性：因果关系有生物学机制支持。
- 一致性与效应强度：有至少 2 项研究一致支持且效应量较大。
- 剂量-反应关系：存在暴露水平高低与结局风险程度的正相关关系。

不满足上述准则的关系可暂删除或标记为“不确定”，并经医学/临床专家组讨论最终决定。

### 6.7.2 根据变量时间顺序处理有环路径

有环路径常由反向因果引起(如  $X \rightarrow Y$  和  $Y \rightarrow X$  同时存在)。需根据研究具体情境厘清时间顺序，并借助实际数据或医学/临床专家组进行处理。

## 6.8 识别第三方变量

在构建完成的 DAG 中，需明确识别以下三类关键的第三方变量（见图 2）。

- 混杂变量**：是同时影响暴露（如出现了某个情况后开始用药）和结局（如肝损伤）的共同原因，其存在会导致药物与不良事件之间产生虚假关联。在 DAG 中表现为从该变量指向暴露和结局的“叉形”结构，如图 a）。
- 中介变量**：是暴露到结局的中间传导机制，代表暴露因素作用的生物学或临床路径。在 DAG 中表现为暴露指向中介，中介再指向结局的“链式”结构，如图 b）。
- 碰撞变量**：是同时受暴露和结局影响的共同结果，DAG 中表现为两个箭头指向同一节点后形成“碰撞”结构，如图 c）。

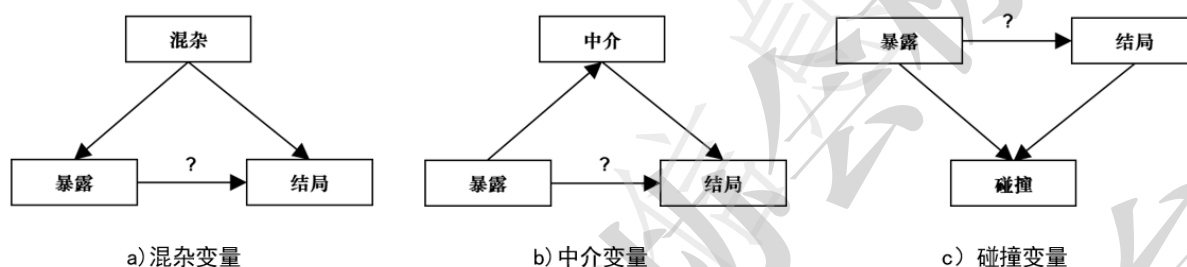


图 2 第三方变量图结构示意图

## 7 应用策略

### 7.1 生成最小充分调整集

7.1.1 在完成对混杂变量、中介变量和碰撞变量的角色划分后，可生成统计分析所需的最小充分调整集，即基于 DAG 对暴露与结局之间的非因果路径进行阻断时，必须且仅需纳入统计模型的一组协变量。该集合旨在既减少因遗漏调整带来的偏倚，又避免过度调整引起的估计偏差或模型效率下降。

7.1.2 在实践中，研究者可借助专门工具确认最小充分调整集。通过绘制或导入已经构建好的 DAG，并标记研究所关注的暴露与结局后，基于图结构的 d-分离原理，即判断给定变量集是否能使暴露与结局在图中条件独立，可以自动检索出能够阻断所有“后门路径”的一组或多组变量。研究者可根据其推荐列表，结合具体研究问题和数据可用性，选择其中适用性最高、规模最小的调整集。

### 7.2 制定数据统计分析策略

#### 7.2.1 混杂变量

将最小充分调整集纳入模型。对于未能测量的潜在混杂因素，可在 DAG 中以“U”标记，以提示残余混杂风险。

#### 7.2.2 中介变量

若目标是估计暴露对结局的总效应，一般不调整；若研究直接效应或作用机制，则可选择性调整。

#### 7.2.3 碰撞变量

不纳入模型，以避免打开新的非因果路径并引入偏倚。

## 参考文献

- [1] 世界医学会.赫尔辛基宣言(2024).2024
- [2] 国际人用药品注册技术协调会. E9 (R1): 临床试验中的估计目标与敏感性分析[Z].2021
- [3] 国家药品监督管理局药品审评中心. 用于产生真实世界证据的真实世界数据指导原则(试行)[Z].2021年第27号, 2021年04月15日.
- [4] 国家药品监督管理局药品审评中心. 药物真实世界研究设计与方案框架指导原则(试行)[Z]. 2023年第5号, 2023年02月16日.
- [5] 国家卫生健康委员会.医疗卫生机构开展研究者发起的临床研究管理办法[Z].国卫科教发(2024)32号, 2024年09月18日.
- [6] 刘慧鑫, 汪海波, 汪宁. 有向无环图在混杂因素识别与控制中的应用及实例分析[J].中华流行病学杂志, 2020,41(4):585-588
- [7] 曹雨滋,袁驰,刘刚,孙凤,李舍予.有向无环图在观察性研究多因素分析协变量选择中的应用[J]. 华西医学, 2025,40(6):1-8.
- [8] 郝允逸,夏雪,胥芹,赵性泉,王安心.有向无环图的构建规范:ESC-DAGs 方法介绍及脑血管病研究案例解读[J]. 中国卒中杂志, 2024,10:1221-1229.
- [9] 白永梅,孙华鸽,杜建.知识图谱:一种系统性构建因果图的方法[J].首都医科大学学报,2022,43(4):584-591.
- [10] MURAD M H, ASI N, ALSAWAS M, et al. New evidence pyramid[J]. BMJ Evidence-Based Medicine. 2016,21(4):125-127.
- [11] KYRIACOU DN, GREENLAND P, MANSOURNIA MA. Using Causal Diagrams for Biomedical Research[J]. Ann Emerg Med. 2023,81(5):606-613.
- [12] LIPSKY AM, GREENLAND S. Causal Directed Acyclic Graphs[J]. JAMA. 2022 Mar 15;327(11):1083-1084.
- [13] FEENEY T, HARTWIG FP, DAVIES NM. How to use directed acyclic graphs: guide for clinical researchers[J]. BMJ. 2025, 388: e078226.
- [14] MANSOURNIA MA, NAZEMIPOUR M. Recommendations for accurate reporting in medical research statistics[J]. Lancet. 2024,403(10427):611-612.
- [15] COCHRANE. Cochrane Handbook for Systematic Reviews of Interventions (中文译名:Cochrane 干预措施系统评价手册)[M]. Wiley-Blackwell, <https://training.cochrane.org/sites/training.cochrane.org/files/public/uploads/resources/CochraneHandbookChineseDec2014.pdf>
- [16] BAI Y, SHI X, DU J. A computable biomedical knowledge system: Toward rapidly building candidate-directed acyclic graphs[J]. J Evid Based Med. 2024,17(2):307-316.
- [17] SHI X, ZHAO W, CHEN T, YANG C, DU J. Evidence triangulator: using large language models to extract and synthesize causal evidence across study designs[J]. Nat Commun. 2025,16(1):7355.
- [18] MURAD MH, ASI N, ALSAWAS M, ALAHDAB F. New evidence pyramid[J]. Evid Based Med. 2016,21(4):125-7.

中国标准化协会  
管理及服务平台

---

ICS 11.020

CCS C 00

关键词：医学观察性研究、有向无环图、因果推断

---