



# 소화기 암 연구에서 방향성 비순환 그래프 활용

이명지

연세대학교 의과대학 의생명시스템정보학교실 의학통계지원실

## Utilizing Directed Acyclic Graphs in the Study of Digestive Cancers

Myeongjee Lee

Biostatistics Collaboration Unit, Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Korea

**Received** November 3, 2025  
**Revised** November 13, 2025  
**Accepted** November 25, 2025

**Corresponding author:**

Myeongjee Lee

**E-mail:** mlee1004@yuhs.ac

<https://orcid.org/0000-0003-2368-5180>

Directed acyclic graphs (DAGs) have emerged as a foundational framework for causal inference in medical and epidemiological research, offering a transparent and systematic approach to visualizing causal assumptions and informing analytic choices. By mapping hypothesized relationships among exposure, outcome, and other covariates, DAGs help researchers identify confounding, mediation, and collider structures, determine appropriate adjustment sets, and avoid bias from inappropriate variable control. In digestive cancer research, DAGs have been especially useful for clarifying complex causal pathways, improving the validity of observational analyses, and informing advanced study designs such as mediation analysis and Mendelian randomization. This paper outlines the conceptual foundations of DAGs, including d-separation, the backdoor criterion, and M-bias, and demonstrates their practical application through the web-based tool DAGitty. We also describe how published digestive cancer studies have incorporated DAG-based frameworks to strengthen study design, ensure appropriate variable adjustment, and improve the transparency and reproducibility of causal interpretation. Collectively, these approaches show that DAGs are not merely graphical aids but essential methodological tools for causal reasoning in contemporary biomedical research.

**Key Words:** Confounding factors; Bias; Causality; Observational study

## INTRODUCTION

소화기암은 전 세계의 암 발생 및 암으로 인한 사망률에서 큰 비중을 차지한다[1]. 특히, 한국의 국가암등록통계에 따르면 소화기계 암이 주요 다빈도 암종을 구성하며, 사망 기여도 측면에서도 폐암에 이어 소화기암이 전체 암 사망에서 큰 비중을 차지한다[2]. 위암·대장암의 경우 검진 및 치료 고도화로 생존율 향상이 보고되고 있으나, 간암·췌장암 등은 여전히 사망위험이 높아 의료자원 소모, 삶의 질 저하, 생산성 손실을 초래한다. 이러한 역학적 특성은 임상 및 공중보건 측면에서 상당한 사회경제적 부담으로 이어

진다. 따라서 과학적 근거에 기반한 소화기암의 예방, 조기 검진, 치료 개선 전략의 개발은 암 관련 사망 감소와 의료비 절감 측면에서 매우 중요한 공중보건 과제이다.

의학연구에서 치료나 중재(intervention)와 같은 노출의 효과가 결과에 미치는 인과적 관계를 규명하기 위한 원칙적 방법은 무작위 대조 시험(randomized controlled trial, RCT)이다. RCT는 노출 이외의 잠재적 교란을 연구 시작 전에 통제함으로써, 노출-결과간 인과성을 평가할 수 있게 한다. 그러나, RCT는 실제 연구 수행 시 윤리적·현실적 제약이 많아서 시행이 어려운 경우가 많고, 실제로는 관찰연구(observational study)가 더 자



주 사용된다. 다만, 관찰연구는 RCT와 달리 교란요인을 사전에 통제할 수 없어 인과 관계를 규명하기 어렵다. 그럼에도 불구하고, 연구 설계 단계에서 통제해야 할 변수와 통제하지 말아야 할 변수들을 명확히 규정하고, 분석 모형에서 그 변수들을 적절하게 조정할 수 있다면 관찰연구도 RCT에 근접한 수준의 결과를 얻을 수 있다[3-5].

최근 인과추론(causal inference) 분야에서는 변수 간의 관계를 시각적이고 체계적으로 표현할 수 있는 방향성 비순환 그래프(directed acyclic graph, DAG)가 널리 활용되고 있다. DAG는 연구자의 전문 분야 지식과 기존 문헌에 근거하여 노출, 결과, 기타 변수들의 관계를 명시적으로 표현하여, 연구자가 가정한 인과 구조를 명확히 확인하도록 돕는다[6-8]. 또한 어떤 변수를 조정해야 하며 어떤 변수는 조정하지 말아야 하는지를 직관적으로 판별할 수 있게 한다. 이러한 접근은 관찰연구의 가장 큰 한계인 교란변수에 의한 편향을 체계적으로 탐지하고 최소화할 수 있게 하며, 나아가 연구 설계의 투명성과 재현성을 향상시킨다. 특히 병태생리적 요인, 임상 중재, 생활습관, 사회경제적 요인이 복합적으로 작용하는 소화기암 연구에서 DAG를 활용한 변수 선정 및 인과구조 탐색은 단순 연관성 분석을 넘어 보다 근거 기반적 인과 해석을 가능하게 하는 중요한 도구로 주목받고 있다.

본 연구에서는 DAG의 기본 개념을 개괄하고, 실제로 DAG를 사용한 소화기암 연구 사례들을 검토한다. 이를 통해 DAG 기반 접근이 임상연구의 설계 및 분석 과정에서 어떻게 활용되는지를 구체적으로 제시하고, 향후 인과 추론 기반 의학 연구의 확장 가능성을 논의하고자 한다.

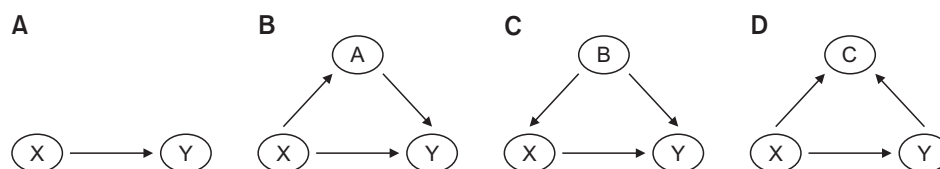
## MAIN SUBJECTS

### DAG의 기본 용어 및 원리

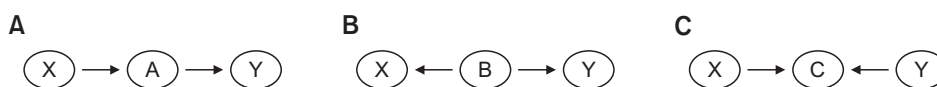
그래프는 정점(node)들과 이들을 연결하는 선(edge)들의 집합으로 이루어진 수학적 구조로 변수들 간의 관계

를 표현하는 데 사용된다. 특히, 모든 선이 방향성을 가지며 순환이 존재하지 않는 그래프를 DAG라고 정의한다. Fig. 1A는 두 정점 X, Y와 한 개의 화살표로 이루어진 단순한 DAG이다. X를 노출변수, Y를 결과변수라고 한다면 'X가 Y의 발생 가능성에 영향을 미친다.'를 그래프로 표현한 것이다. 이때 X는 Y의 선행변수(ancestor)이고 Y는 X의 후행변수(descendant)라고 한다. 만약, X에서 Y로의 화살표가 없으면 'X가 Y에 직접적으로 영향을 미치지 않는다.'는 가정을 표현한다. Fig. 1A에 하나의 변수(정점)를 추가하여 다양한 DAG를 살펴보자. Fig. 1B는 X의 영향을 받고 동시에 Y에 영향을 주는 변수 A를 추가하여 그린 DAG이다. Fig. 1C는 X와 Y에 모두 영향을 미치는 변수 B를 추가하여 그린 DAG이다. Fig. 1D는 X와 Y로부터 모두 영향을 받는 변수 C를 추가하여 그린 DAG이다. 변수를 추가함으로써 인과 변수 X에서 변수 Y로의 경로(path)가 다양해지며, 이들을 인과적경로(causal path)와 비인과적 경로(noncausal path)로 구분할 수 있다. X에서 Y로의 경로에서 모든 화살표가 X에서 Y로 향하고 있는 경로를 인과 경로라고 하고, 그렇지 않은 화살표가 하나라도 있는 경우를 비인과 경로라고 한다. Fig. 1의 각 DAG를 자세히 살펴보자. Fig. 1A와 같이 노출변수 X가 직접적으로 결과변수 Y에 영향을 미치는 경로를 직접 인과 경로(direct causal path)라고 한다. Fig. 1B는 노출변수 X가 변수 A를 거쳐 결과변수 Y에 영향을 주는 X에서 Y로 향하는 경로를 포함하고 있으며, 이 경로  $X \rightarrow A \rightarrow Y$ 를 간접 인과 경로(indirect causal path)라고 한다. 이 때, 변수 A를 매개변수(mediator)라고 한다. 반면에 Fig. 1C, 1D 각각에는  $X \rightarrow Y$ 의 인과 경로 외에  $X \leftarrow B \rightarrow Y$ ,  $X \rightarrow C \leftarrow Y$ 와 같은 비인과 경로를 포함하고 있는데, 변수 X와 변수 Y 모두의 선행변수인 변수 B는 교란변수(confounder)라 하고, 변수 X와 변수 Y 모두의 후행변수인 변수 C는 충돌변수(collider)라고 한다.

Fig. 2는 Fig. 1의 각 DAG에서 직접 인과 경로인  $X \rightarrow Y$ 를 제외한 경로들을 도식화해 놓은 것이다. 이 DAG들



**Fig. 1.** Fundamental types of causal pathways in directed acyclic graphs: (A) a direct causal path only, (B) a combination of direct and mediated causal paths, (C) a combination of direct and confounded causal paths, and (D) a collider structure.



**Fig. 2.** Basic causal structures in directed acyclic graph: (A) chain, (B) fork, and (C) collider.

은 두 개의 화살표와 세 개의 변수로 이루어진 DAG의 기본 인과 구조들이다. Fig. 2A는 모든 화살표가 같은 방향을 향하고 있는 간접 인과 경로이며 이를 ‘연쇄(chain)’ 구조라고 한다. 이 구조에서는 X에서 Y로의 경로가 매개변수 A에 의해 막힐(block) 수도 있음을 의미한다. 예를 들어, X가 흡연, A를 염증여부, Y를 위암발생이라고 한다면, 염증의 수준을 통제한다면 흡연과 위암발생과는 직접적인 관련이 없지만(독립), 염증을 통제하지 않으면 흡연여부가 위암과 연관이 있다. 따라서, 연쇄구조인 DAG에서는 매개변수 A를 통제하면  $X \rightarrow A \rightarrow Y$  경로가 닫히고, A를 통제하지 않으면 해당 경로가 열리게 된다. 즉, 매개변수 A의 통제 여부에 따라 X와 Y의 독립성 여부가 달라지며, 이를 조건부 독립(conditional independence)이 성립한다고 한다. Fig. 2B는 변수 B가 두 변수 X와 Y의 공통 원인이 되는 상황을 표현하고 있는 ‘분기(fork)’ 구조이며, 비인과 경로이다. 예를 들어, X는 위산분비 억제제 사용, B는 헬리코박터 감염, Y를 위암발생이라고 하자. 위산분비 억제제 사용이 위암발생과 연관성이 있어 보이지만, 실은 헬리코박터 감염이 있는 사람은 위염이나 소화불량으로 위산분비 억제제를 처방 받을 가능성이 높고 동시에 위암발생의 위험도 증가한다. 즉, 헬리코박터 감염여부를 고려하지 않는 경우 가짜 연관성을 야기하게 된다. 따라서, 분기 구조의 DAG에서 교란변수인 B는 모형에 포함시켜서 통제하고, 비인과 경로인  $X \leftarrow B \rightarrow Y$ 를 닫히도록 해야 한다. 다시 말해, B를 통제하면 X와 Y는 조건부 독립이 성립한다. Fig. 2C는 변수 C가 두 변수 X와 Y의 공통 결과가 되는 상황으로 ‘충돌(collider)’ 구조라고 하며 비인과 경로이다. 이 DAG에 따르면, X와 Y는 독립이지만 C가 주어질 경우 X와 Y는 더 이상 독립이 아니다. X를 종양의 공격성, C를 수술 성공여부, Y를 외과의의 술기 수준이라고 하면, 종양의 공격성과 외과의의 술기 수준은 독립이다. 하지만, 공격적인 종양일수록 수술 성공률이 낮고, 숙련된 외과의일수록 수술 성공률이 높을 것이며, 만약 공격적인 종양인데 수술이 성공했다면 해당 수술은 숙련된 외과의가 수행했을 가능성이 높고, 덜 공격적인 종양인데 수술이 실패했다면, 해당 수술은 숙련도가 낮은 의사가 집도했을 가능성이 크다고 할 수 있다. 즉, 수술 성

공여부를 고려하게 되면 가짜 연관성이 야기된다. 따라서, 충돌 구조의 DAG에서는 충돌변수 C를 통제하지 않아야 비인과적 경로  $X \rightarrow C \leftarrow Y$ 가 닫히게 되고, X와 Y의 독립성이 유지된다.

## DAG의 활용

### 변수선택

연구를 설계 단계에서 DAG를 활용하면 변수들 간의 관계 구조를 시각적으로 파악할 수 있으며, 이 과정에서 인과효과 추론을 위해 보정이 필요한 변수와 보정하지 말아야 할 변수들을 명확히 구분할 수 있다. 변수 선택에 적용되는 주요 원칙과 유의사항을 살펴보고, DAG를 그리는 도구를 소개하고자 한다.

### D-separation rule

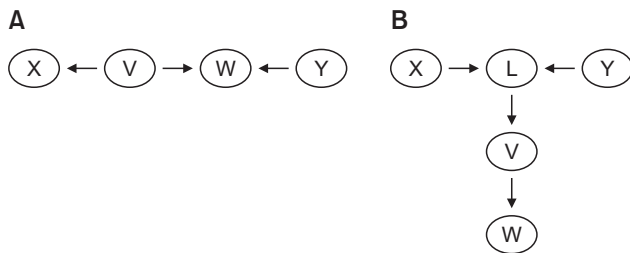
앞에서 간단하게 DAG를 사용하여 기본 구조 및 해석하는 방법에 대해 살펴보았다. 하지만, 실제 연구에서는 더 많은 수의 변수들을 사용한 DAG를 마주하게 된다. 이 때, 어떤 변수들을 통제해야 할까? 노출변수 X와 결과변수 Y의 인과관계를 살펴보고자 할 때, X에서 Y로의 여러 경로들 중 인과 경로들만 열고 비인과 경로들은 닫히도록 하는 제 3의 변수들을 선택하여 통계모형에 포함해야 하는데, 이와 관련한 규칙을 d-separation rule이라고 한다. D-separation rule은 제 3의 변수들을 조건부로 하였을 때, X와 Y가 조건부 독립성을 만족하는 지 판별하는 방법이며 각 규칙들은 아래와 같다.

규칙 1. X에서 Y로의 경로에서 변수(L)가 연쇄 구조나 분기 구조에 있을 경우, 변수 L을 보정하면 해당 경로는 닫히게 된다.

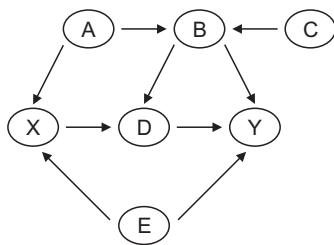
규칙 2. X에서 Y로의 경로에서 변수(L)가 충돌 구조에 있을 경우, 변수 L 또는 그 변수의 후행변수들을 통제하면 해당 경로는 열리게 된다.

규칙 3. 규칙 1 또는 규칙 2에 의해서 경로가 한 번 닫히게 되면, 그 경로는 닫히게 된다.

다음 Fig. 3의 DAG에서 위 규칙을 적용해보도록 하자. Fig. 3A의 변수 X에서 변수 Y로의 경로는 변수 V만 보정



**Fig. 3.** Illustrative examples demonstrating the d-separation rules in directed acyclic graphs: (A) Rule 1 (blocking), Rule 2 (opening), and Rule 3 (re-closing) in the X-Y path; (B) Rule 2 applied to a collider L and its descendants.

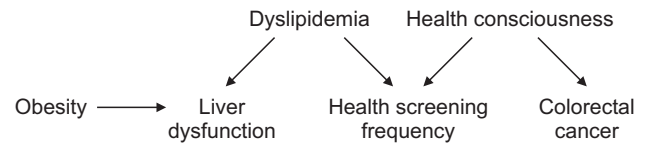


**Fig. 4.** An illustrative example demonstrating the backdoor criterion in a directed acyclic graph.

하면 규칙 1에 의해 해당 경로는 닫히게 되며, 변수 W만 보정하면 규칙 2가 적용되어 해당 경로는 열리게 되지만, 변수 V와 변수 W를 동시에 통제하면 규칙 3을 적용하여 해당 경로는 닫히게 된다. Fig. 3B의 변수 X에서 변수 Y로의 경로는 변수 L을 보정하거나 또는 후행변수에 해당 하는 변수 V 또는 변수 W를 보정하면 해당 경로는 열리게 된다(규칙 2).

#### Backdoor criterion

변수의 수가 증가할수록 DAG도 커지고 복잡해지며, 다양한 경로가 포함된 DAG에서 변수 X에서 변수 Y로의 인과효과를 추정하기 위해 X에서 Y 사이의 비인과 경로 (noncausal path, backdoor path)들을 차단시키는 변수들의 집합을 선택하는 원리를 backdoor criterion이라고 한다[9]. Backdoor criterion을 만족하는 변수들의 집합을 Z라고 하면, Z에 속하는 변수들은 변수 X의 후행 변수가 아니어야 하며, 변수 X에서 변수 Y로의 모든 경로들 중 변수 X로 화살표가 향하는 경로(backdoor path)들을 닫게 하는 변수들이어야 한다. Fig. 4에서, X에서 Y로의 경로들 중 위 조건을 만족하는 비인과 경로들은  $X \leftarrow A \rightarrow B \rightarrow Y$ ,  $X \leftarrow A \rightarrow B \rightarrow D \rightarrow Y$ ,  $X \leftarrow E \rightarrow Y$ 이



**Fig. 5.** An illustrative example demonstrating the M-bias in a directed acyclic graph depicting the hypothesized causal pathways between obesity and colorectal cancer.

고  $Z = \{A, E\}$ 이다. 즉,  $Z = \{A, E\}$ 를 통계 모형에서 보정함으로써, 변수 X에서 변수 Y로의 경로들 중 비인과 경로들은 닫히게 된다. 여기서, 만약 변수 A가 DAG 상으로는 데이터에 존재해야 할 변수이지만, 실제 분석에서 구하기 어려운 변수라고 하자. 이 때, 변수 A를 보정하지 않으면  $X \leftarrow A \rightarrow B \rightarrow Y$ ,  $X \leftarrow A \rightarrow B \rightarrow D \rightarrow Y$ 가 열리게 되고, X에서 Y로의 인과효과를 확인할 수가 없다. 그러나, 해당 경로들은 측정되지 않은 변수 A 대신 변수 B를 보정하더라도 해당 경로들이 닫히게 되므로  $Z = \{B, E\}$ 로 정의될 수도 있다. 더 나아가 집합 Z는  $\{A, E\}$ ,  $\{B, E\}$ ,  $\{A, B, E\}$ ,  $\{A, C, E\}$ ,  $\{B, C, E\}$ ,  $\{A, B, C, E\}$ 로도 구성될 수 있다는 것을 알 수 있으며, 이들 중에서 원소의 수가 가장 적은 부분 집합인  $\{A, E\}$  또는  $\{B, E\}$ 를 최소 충분 보정 집합 (minimally sufficient adjustment sets)이라고 한다.

#### M-bias

일반적으로, 연구자들은 변수 X와 변수 Y의 연관성에 대하여 살펴보고자 할 때, 제 3의 변수들 중 단변량 모형 분석 결과의 p value가 유의수준 0.05 미만인 변수들을 교란변수로 정의하고, 이렇게 정의된 변수들을 포함하여 분석하고 있다. 다음의 예를 살펴보자. 비만이 대장암 발생에 미치는 영향에 대해서 살펴보고자 하는 연구를 수행하기 위하여 그려본 DAG이다(Fig. 5). 이 때, ‘고지혈증’ 여부와 ‘건강의식’ 변수가 가용 데이터에 없다고 가정하자. 각 ‘간기능이상’과 ‘건강검진빈도’ 변수의 단변량 분석 결과 유의수준 0.05에서 p value가 0.05 미만이라면, 두 변수를 다변수(multivariable) 분석 모형에 포함시켜야 할까? DAG에 의하면, ‘간기능이상’과 ‘건강검진빈도’를 보정하지 않는 것이 ‘비만’과 ‘대장암’간의 인과효과 추정이 가능하다. 그럼에도 불구하고 ‘간기능이상’과 ‘건강검진빈도’ 모두를 다변수 분석에 추가하여 해당 변수들의 영향을 통제하면, 해당 경로는 열리게 되고 비인과 경로가 생겨 인과효과 추정 시 편향이 생기게 된다. 이를 M-bias라고 한다.

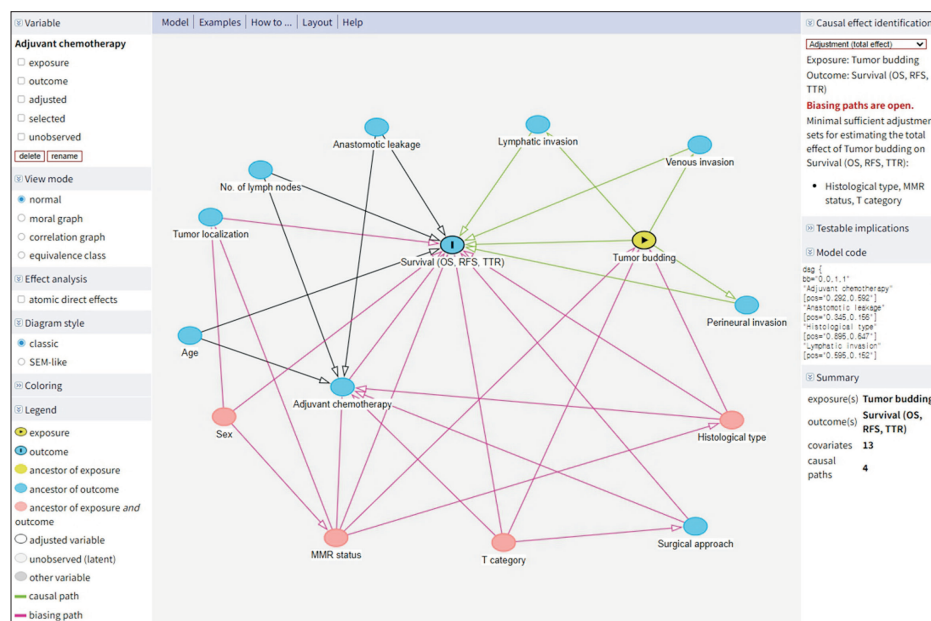


### DAGitty

빅데이터를 기반으로 하여 연구를 수행하며 DAG를 그려보면 많은 수의 변수와 복잡한 구조로 인하여 연구자가 직접 d-separation rule과 backdoor criterion을 적용하여 어떤 변수들을 통제하는 것이 적절한 지 판단하는 것은 쉽지 않다. DAGitty 프로그램을 사용하면 연구자들이 직접 DAG를 그릴 수 있고 통계 분석 모형에 포함되어야 할 변수들을 동시에 확인할 수 있다[10]. DAGitty 프로그램 사용방법은 다음과 같다. 해당 프로그램은 웹 기반 사용자 인터페이스 형태로 제공되며 <https://www.dagitty.net>에서 이용할 수 있다. 첫 화면에서 'Launch DAGitty online in your browser'를 선택하고 가운데 회색 창 상단 메뉴에서 'Model'-'New model'을 선택한다. 회색 창 내부 아무 곳에서 클릭을 하면 변수를 추가할 수 있으며, 각 변수의 역할(exposure, outcome 등)은 변수를 선택한 후, 왼쪽 'Variable' 메뉴에서 선택하여 지정할 수 있다. Exposure로 선언된 변수는 내부에 검정색 삼각형이 포함된 초록색 타원으로 표시되며, outcome로 선언된 변수는 내부에 I가 포함된 파란색 타원으로 표시된다. 선행변수에서 후행변수로의 연결은 각 변수를 순서대로 클릭하여 화살표로 설정하며, 동일한 작업을 반복하면 해당 화살표가 제거된다.

Pihlmann Kristensen et al. (2024) [11]이 제시한 DAG를 DAGitty 프로그램을 이용하여 Fig. 6에 제시하였다. 종양발아(tumor budding)는 Union for Interna-

tional Cancer Control Tumor, Node, Metastasis 분류 제 8판에서 조직학적 예후 인자로 포함된 잘 알려진 바이오마커이며, 특히 고등급 종양발아는 2기 결장암(stage II colon cancer)의 불량한 예후 인자로 보고되고 있다. 이에 Pihlmann Kristensen et al. [11] 연구진은 2기 결장암 환자들을 대상으로 종양발아와 예후 간의 관련성을 분석하였으며, 노출변수와 결과변수를 포함한 15개의 변수들을 이용하여 DAG를 제시하였다. DAGitty 프로그램을 사용하여 변수 'Tumor budding'을 exposure로, 변수 'Survival'을 outcome으로 설정하고 나머지 13개 변수 간의 관계를 화살표로 도식하였으며, 그 결과 인과 경로는 초록색 선으로 표시되고, 그 외 경로들은 분홍색 또는 검정색으로 구분되어 나타났다. 분홍색으로 표시된 경로들은 biasing path들을 나타내며, 이러한 경로들은 적절한 변수 조정을 통해 차단해야 인과효과의 올바른 추정 가능성이 가능하다. 회색 창 우측의 'Causal effect identification' 메뉴에서는 Tumor budding이 예후(overall survival, recurrence-free survival, time to recurrence)에 미치는 causal (total) effect를 추정하기 위해 모형에 포함해야 할 변수들의 minimal sufficient sets로 histological type, mismatch repair status, T category 변수들을 제안하였다. Pihlmann Kristensen et al. (2024) [11]의 원문에서도 동일한 세 변수를 보정하여 분석을 수행하고 결과를 보고하였다.

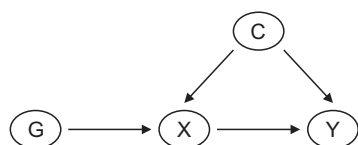


**Fig. 6.** Directed acyclic graph illustrating the hypothesized causal structure between the exposure (tumor budding) and outcome (time to overall survival, recurrence-free survival, and recurrence). The conceptual structure was based on the framework reported by Pihlmann Kristensen et al. (Hum Pathol 2024;146:15-22) [11].

### DAG 기반 연구방법의 선택

DAG는 인과효과를 추론하기 위해 조정해야 할 변수를 선택하는 도구의 역할을 넘어 연구의 목적 설정 및 분석 방법을 결정하는 데에도 활용될 수 있다. Fig. 1B는 노출 변수 X에서 결과 변수 Y로 향하는 경로들이 모두 인과 경로임을 보여주며, 그 중 경로  $X \rightarrow Y$ 는 직접 인과 경로, 경로  $X \rightarrow A \rightarrow Y$ 는 간접 인과 경로로 구분된다. 이 때, 간접 인과 경로상에 위치한 변수 A를 매개변수라고 한다. 매개분석(mediation analysis)은 노출변수 X가 결과변수 Y에 미치는 총 인과효과를 직접 인과 경로를 통한 직접효과(direct effect)와 매개변수가 존재하는 간접 인과 경로를 통한 간접효과(indirect effect)로 구분하여 추정하는 분석방법이다. 즉, 전체 효과 중 매개변수에 의해 설명되는 간접효과의 기여도를 평가함으로써 인과관계의 기전을 구체적으로 이해할 수 있다. 예를 들어, 변수 X를 비만, 매개변수 A를 염증표지자, 변수 Y를 대장암 발생이라고 설정한 경우 매개분석을 통해 비만이 대장암 발생에 미치는 총 효과를 추정할 때, 비만이 염증반응을 유발하고 이를 통해 대장암 발생에 영향을 미치는 간접효과의 크기를 정량적으로 평가할 수 있다.

한편, Fig. 7은 연구 설계 단계에서 도출된 또 다른 형태의 DAG를 보여준다. 해당 DAG에 의하면 변수 G는 다음 세 조건을 만족한다. 변수 G는 (1) 노출 변수 X에만 영향을 미치고, (2) 결과 변수 Y에는 직접적인 영향을 주지 않으며, (3) 변수 G는 변수 X를 통해서만 변수 Y에 영향을 미치고, 변수 X와 변수 Y의 교란변수(C)와 독립이다. 이러한 특성을 갖는 변수 G를 도구변수(instrumental variable, IV)라고 한다. 예를 들어, low-density lipoprotein (LDL) 콜레스테롤을 낮추는 유전자 변이가 대장암 발생을 감소시키는지에 대한 연구를 한다고 하면 유전자 변이(G), LDL 콜레스테롤 수치(X), 대장암 발생(Y) 변수들의 관계를 Fig. 7의 DAG와 같이 표현할 수 있다. 유전자는 멘델의 독립의 법칙(Mendel's law of independent assortment)에 의해 무작위적으로 분리되므로 유

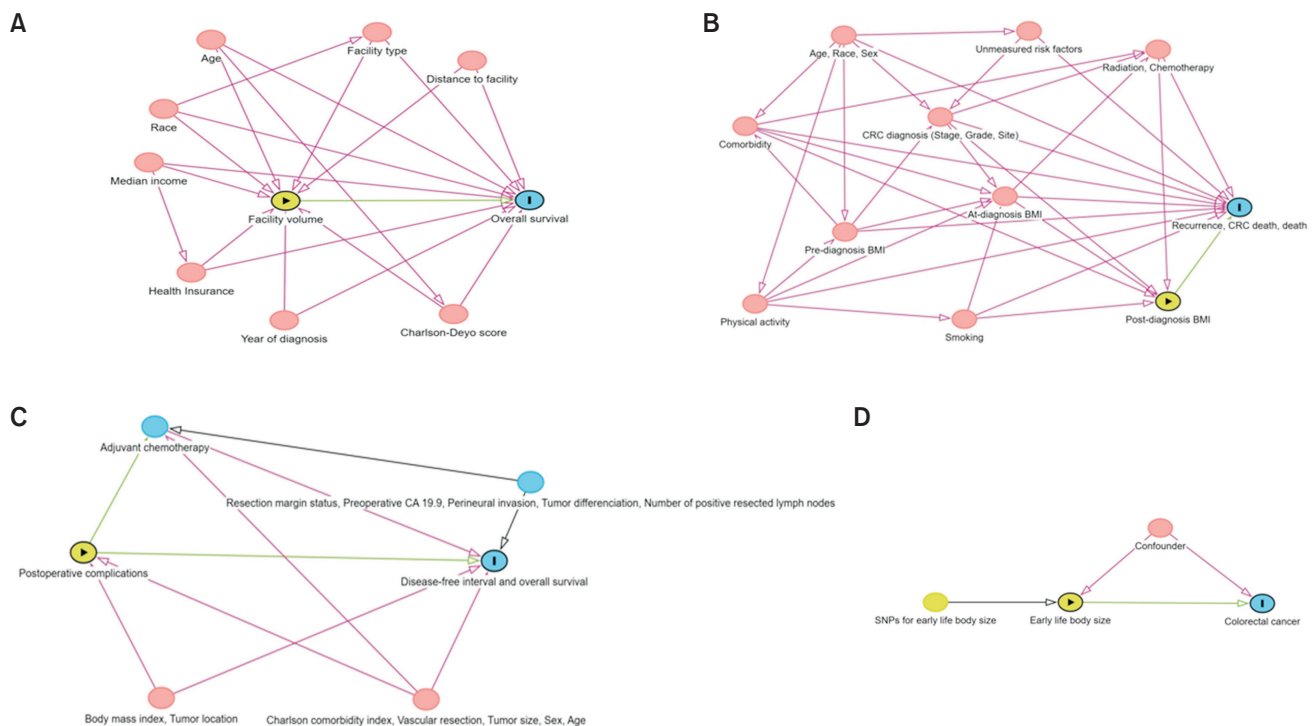


**Fig. 7.** Directed acyclic graph illustrating the causal structure of Mendelian randomization analysis.

전적 변이를 이용한 분석은 RCT와 유사한 인과추론이 가능하다[12]. 이처럼 유전적 변이를 IV를 활용하여 인과효과를 추정하는 접근방법을 멘델의 무작위화(Mendelian randomization, MR) 분석이라고 하며, 이는 관찰연구에서 인과관계를 보다 엄밀하게 검증할 수 있는 분석방법으로 활용되고 있다.

### 소화기암 연구에서 DAG가 활용된 사례

실제로 출판되고 있는 여러 소화기암 연구들에서 DAG는 연구 설계와 타당성을 확보하고 인과효과 해석을 명확히 하는 데 활용되고 있다. 그 중 몇 가지 사례들을 선정하여 각 연구에서 사용된 DAG를 DAGitty 프로그램을 이용해 재구성하였다(Fig. 8). 이를 통해 각 연구 분석에서 최소 충분 보정집합으로 선택된 변수들이 적절하게 적용되었음을 확인하였다. Elshami et al. (2022) [13]은 병원의 간담췌암 환자 진료경험의 정도(facility volume)가 수술을 시행하지 않은 췌장 선암 환자의 생존에 미치는 영향을 평가한 연구를 수행하였다. 이 연구에서는 인과효과를 구하기 위한 변수 선택 과정에서 DAG를 활용하였으며, DAG에 의해 선택된 변수들을 이용하여 inverse probability (IP)-weight를 산출하고 IP-weighted Cox regression model을 적용하였다(Fig. 8A). Kroenke et al. (2016) [14]은 대장암 환자들을 대상으로 체질량 지수(body mass index, BMI)가 사망률에 미치는 영향을 분석하였다. 이 연구에서도 DAG를 이용하여 잠재적 교란변수들을 사전에 정의하였으며 암 진단 시점의 BMI뿐 아니라 암 진단 이후의 BMI가 사망률에 미치는 영향을 구분하여 평가하였다(Fig. 8B). DAG에 의하면, 다른 교란변수들이 외에 암 진단 이전에 측정된 BMI가 보정해야 할 변수로 선택되어 다변량 모형에 추가로 보정하여 인과효과를 추정하였다. Henry et al. (2023) [15]은 네덜란드 췌장암 연구 그룹과 함께 췌장암 절제술을 시행한 환자들을 대상으로, 수술 후 발생한 합병증이 재발 및 생존에 미치는 영향을 분석하였다. 이 연구에서는 DAG를 사용하여 교란변수들을 정의하였고, 더 나아가 보조항암치료가 매개변수로 작용함을 DAG를 통해 확인하였다. 이에 따라 연구진은 합병증이 생존에 미치는 총 효과뿐 아니라 보조항암치료의 매개효과도 함께 평가하였다(Fig. 8C). 마지막으로, Papadimitriou et al. (2023) [16]은 성장 초기 및 성인기의 비만도가 대장암 발생에 미치는 인과적 영향



**Fig. 8.** Directed acyclic graphs (DAGs) based on published studies, illustrating various applications of causal inference methods: (A) DAG representing the hypothesized causal structure between facility volume and overall survival, conceptually informed by the framework of Elshami et al. (HPB [Oxford] 2022;24:1878-1887) [13]. (B) DAG representing the hypothesized causal structure between post-diagnosis body mass index (BMI) and outcomes, conceptually informed by Kroenke et al. (JAMA Oncol 2016;2:1137-1145) [14]. (C) DAG representing the hypothesized causal structure between postoperative complications and outcomes, conceptually informed by Henry et al. (BJS Open 2023;7:zrac174) [15]. (D) DAG representing the hypothesized causal structure between early-life body size and colorectal cancer, conceptually informed by Papadimitriou et al. (BMC Med 2023;21:5) [16].

을 구분하기 위해 유전체연관연구데이터를 사용하여 성장기 체형 및 성인기 비만도와 관련된 유전자 변이를 IV로 정의하여 MR 분석을 수행하였다(Fig. 8D).

## CONCLUSION

최근 의료계 다양한 분야에서 빅데이터가 축적됨에 따라, 관찰연구를 통해 편향을 최소화하고 최적화된 인과추론을 하기 위한 다양한 방법이 제안되고 있고, 그 출발점으로 DAG가 많이 사용되고 있다. 노출변수와 결과변수 간의 단순 연관성을 평가하는 연구에서는 DAG 제시가 필수적이지는 않다. 그러나 관찰자료를 기반으로 인과관계를 규명하고자 하는 경우, 인과 관계를 왜곡하거나 혼란을 초래할 수 있는 변수들을 식별하고 그 선택의 근거를 명확히 제시하기 위해 DAG를 함께 제시하도록 권장하고 있다. 본 연구에서는 DAG를 소개하고, DAG를 이용하여 노출변수와 결과변수 사이의 인과관계를 규명하기

위해 모형에 고려해야 할 변수들을 선택하는 방법에 대해 살펴보았다. DAG를 사용하면 노출변수 X와 결과변수 Y 사이의 인과관계를 확인할 때 관련되는 다양한 변수들 사이의 관계를 직관적으로 확인할 수 있고, 이를 통해 연구자 간 투명하게 가설을 공유할 수 있다. D-separation rule과 backdoor criterion을 통해 무엇을 보정하고 무엇을 보정하지 말아야 하는지를 사전에 명시함으로써 과대·과소보정 및 충돌변수 보정에 의한 편향을 줄이는 변수 선택이 가능해진다. 뿐만 아니라 DAG를 이용하여 매개분석, IV 지정 후 MR 분석 등 다양한 인과추론 분석 방법을 연구 가설 규명을 위해 고려해 볼 수 있다.

하지만, 동시에 연구 수행에서 DAG 사용 시 한계점들은 다음과 같다. 첫째, DAG를 사용함으로써 통제해야 할 변수를 파악하고, 변수 통제 여부에 따라 어떤 경로가 열리고 차단되는지 확인할 수 있고, 추정량의 편향 존재 여부를 판단할 수는 있지만 편향의 크기와 방향을 직접 산정하지는 못한다. 둘째, DAG는 연구자의 사전 지식과 가



정에 기반하여 구성되므로 주관적일 수 있으며, 데이터에서 모든 관계가 도출되는 것은 아니다. 따라서, 해당 분야의 배경 지식이 충분하지 않거나 인과관계에 대한 이해가 부족한 경우, 제시된 DAG의 현실 적합성을 검증하기가 어렵다. 또한, DAG를 통해 통제해야 할 변수를 명확히 정의하였다 하더라도, 실제 자료에서 해당 변수가 관측되지 않거나 정확히 측정되지 않은 경우에는 잔여 교란(residual confounding)을 피하기 어렵다. 따라서 연구자는 임상적 및 문헌적으로 타당한 근거를 제시하여 DAG를 설명할 수 있어야 하며, 다학제 전문가들이 함께 참여하여 다양하고 가능한 시나리오의 DAG를 병렬적으로 구성하고 각 시나리오에서 도출되는 최소 보정 집합을 비교·검토하는 과정이 필요하다. 더불어, 각 DAG를 기반으로 분석을 수행한 후, 결과의 변동 정도를 평가하는 민감도 분석과 외부자료를 통한 검증을 병행함으로써 연구 결과의 견고성(robustness)을 확보하는 것이 바람직하다 [7].

결론적으로, 소화기암 연구에서 DAG는 단순한 도식이 아니라 인과 구조를 시작화하고 연구 설계 및 분석 전략을 체계화하는 핵심 도구로 활용될 수 있으며 인과 추론의 투명성과 정확성을 높이는 실행 가능한 설계 인프라이다. 연구 설계 단계에서부터 임상 의사, 역학자, 통계학자, 데이터 과학자의 다학제적 협업을 통해 DAG를 적극적으로 도입하여 다양한 시나리오의 DAG를 구축해보고, 각 DAG에 근거한 최소 충분 보정집합의 제시 및 적절한 최신 인과 분석방법의 시행 및 시나리오별 민감도 분석의 통합적 운용을 결합한다면 관찰자료에서도 임상적으로 의미 있는 인과 추론이 가능해진다. 이는 빅데이터 시대의 풍부한 정보를 편향 통제와 재현성이라는 원칙 아래 임상적 가치로 전환하여, 소화기암의 예방·조기검진·치료의 최적화 전략을 보다 견고하게 뒷받침하는 토대가 될 것이다.

## FUNDING

None.

## CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

## REFERENCES

1. Zhou Y, Song K, Chen Y, et al. Burden of six major types of digestive system cancers globally and in China. *Chin Med J (Engl)* 2024;137:1957-1964. <https://doi.org/10.1097/CM9.0000000000003225>
2. Park EH, Jung KW, Park NJ, et al. Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2022. *Cancer Res Treat* 2025;57:312-330. <https://doi.org/10.4143/crt.2025.264>
3. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;70:41-55. <https://doi.org/10.1093/biomet/70.1.41>
4. Vanderweele TJ. *Explanation in causal inference: methods for mediation and interaction*. New York: Oxford University Press, 2015. p.xvi, 706.
5. Hernán MA, Robins JM. *Causal inference: what if*. Boca Raton: Chapman & Hall/CRC, 2020.
6. Byeon S, Lee W. Directed acyclic graphs for clinical research: a tutorial. *J Minim Invasive Surg* 2023;26:97-107. <https://doi.org/10.7602/jmis.2023.26.3.97>
7. Swarbrick CJ, Blake HA, Martin P, Partridge JSL, Moppett IK. Directed acyclic graphs to minimise bias and optimise causal inference in SNAP-3: an observational cohort study of frailty, multimorbidity, and delirium in older surgical patients. *Br J Anaesth* 2025;135:177-187. <https://doi.org/10.1016/j.bja.2025.04.027>
8. Tennant PWG, Murray EJ, Arnold KF, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *Int J Epidemiol* 2021;50:620-632. <https://doi.org/10.1093/ije/dyaa213>
9. Pearl J. Causal diagrams for empirical research. *Biometrika* 1995;82:669-688. <https://doi.org/10.1093/biomet/82.4.669>
10. Textor J, van der Zander B, Gilthorpe MS, Liskiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *Int J Epidemiol* 2016;45:1887-1894. <https://doi.org/10.1093/ije/dyw341>
11. Pihlmann Kristensen M, Korsgaard U, Timm S, et al.



- The prognostic value of tumor budding in a thoroughly characterized stage II colon cancer population in the context of a national screening program. *Hum Pathol* 2024;146:15-22. <https://doi.org/10.1016/j.humpath.2024.02.010>
12. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* 2008;27:1133-1163. <https://doi.org/10.1002/sim.3034>
  13. Elshami M, Ahmed FA, Kakish H, et al. Average treatment effect of facility hepatopancreatobiliary cancer volume on survival of non-resected pancreatic adenocarcinoma. *HPB (Oxford)* 2022;24:1878-1887. <https://doi.org/10.1016/j.hpb.2022.07.007>
  14. Kroenke CH, Neugebauer R, Meyerhardt J, et al. Analysis of body mass index and mortality in patients with colorectal cancer using causal diagrams. *JAMA Oncol* 2016;2:1137-1145. <https://doi.org/10.1001/jamaoncol.2016.0732>
  15. Henry AC, van Dongen JC, van Goor IWJM, et al. Impact of complications after resection of pancreatic cancer on disease recurrence and survival, and mediation effect of adjuvant chemotherapy: nationwide, observational cohort study. *BJS Open* 2023;7:zrac174. <https://doi.org/10.1093/bjsopen/zrac174e>
  16. Papadimitriou N, Bull CJ, Jenab M, et al. Separating the effects of early and later life adiposity on colorectal cancer risk: a Mendelian randomization study. *BMC Med* 2023;21:5. <https://doi.org/10.1186/s12916-022-02702-9>