

Medical Justification: Complete Analysis of Heart Disease Attributes

Subject: Clinical Validation of Attribute Dependencies for Bayesian Network Construction

Dataset: UCI Heart Disease (14 Attributes)

Reference: Medical Plausibility Analysis & UCI Repository Data

1. Introduction

This document validates the relationships for all 14 attributes in the UCI Heart Disease dataset. It serves as the "Ground Truth" for the Bayesian Network structure, ensuring that the model's probability calculations are based on real medical science rather than random correlations.

Relationship Key:

- **Inverse (↓):** As Factor A increases, Factor B decreases.
- **Direct (↑):** As Factor A increases, Factor B (or Risk) increases.
- **Associative (↔):** Factors are linked by biological category (e.g., Sex).

2. Inverse Relationship

There is one primary inverse relationship dictated by human biology regarding aging.

1. Age → 8. Max Heart Rate (thalach)

- **Relationship: Inverse (Negative Correlation).**
- **The Rule:** Older patients have a lower "speed limit" for their heart rate.
- **Medical Logic:** Aging reduces the intrinsic heart rate maximum due to a decline in beta-adrenergic responsiveness.
- **Formula:** Max HR is roughly 220 minus Age.
- **In the Model:** A high value for age heavily increases the probability of a low value for thalach.

3. Direct & Causal Relationships

These attributes represent the "Domino Effect" where biological aging leads to physical damage.

Hypertension Pathway

- 1. Age → 4. Resting Blood Pressure (trestbps)
 - **Relationship:** Direct.
 - **Logic:** Arteries naturally stiffen with age (arteriosclerosis), leading to higher systolic blood pressure.
- 4. Resting Blood Pressure (trestbps) → 7. Rest ECG (restecg)
 - **Relationship:** Causal.
 - **Logic:** Chronic high blood pressure forces the heart muscle to thicken (Left Ventricular Hypertrophy) to pump against the resistance. This physical thickening creates a specific abnormality (Value 2) on the resting ECG.

The Symptom Attributes (Subjective)

These attributes describe what the patient *feels*, which guides the diagnosis.

- 3. Chest Pain Type (cp) → 14. Disease Target (num)
 - **Relationship:** Qualitative Predictor.
 - **Logic:**
 - **Value 1 (Typical Angina):** The strongest clinical predictor of obstructive disease.
 - **Value 2/3 (Atypical/Non-anginal):** Lower risk, often related to other issues.
 - **Value 4 (Asymptomatic):** Can indicate "Silent Ischemia" (dangerous but quiet).
- 9. Exercise Induced Angina (exang) → 14. Disease Target (num)
 - **Relationship:** Direct / High Specificity.
 - **Logic:** If the patient specifically feels chest pain *during* exercise (exang=1), it is a highly specific sign that the coronary arteries are blocked and cannot supply oxygen under stress.

4. The Test Result Attributes (Objective Evidence)

These are the strongest predictors in the network because they measure the disease directly.

- 10. ST Depression (oldpeak) → 14. Disease Target (num)
 - **Relationship:** Direct / Strong.
 - **Logic:** This measures the magnitude of the electrical drop on the ECG during stress. Deeper depression (higher mm) = More severe oxygen starvation.
- 11. Slope → 10. ST Depression (oldpeak)
 - **Relationship:** Structural Modifier.
 - **Logic:** The slope describes the *shape* of the ST depression. A "Downsloping" (Value 3) shape is functionally worse than an "Upsloping" (Value 1) shape. It modifies how doctors interpret the oldpeak value.

- 12. Fluoroscopy Vessels (ca) → 14. Disease Target (num)
 - **Relationship:** Direct / Anatomical (Strongest).
 - **Logic:** Fluoroscopy uses dye to see the arteries. ca counts the number of blocked vessels (0–3). ca > 0 is visual proof of plaque accumulation.
- 13. Thallium Test (thal) → 14. Disease Target (num)
 - **Relationship:** Direct.
 - **Logic:** A nuclear imaging test.
 - 3 (Normal): Healthy blood flow.
 - 6 (Fixed): Scar tissue from past heart attack.
 - 7 (Reversible): Active ischemia (blood flow blocked during stress).

5. Risk Factors

These attributes increase the baseline probability of disease but do not guarantee it.

- 2. Sex → 5. Cholesterol (chol)
 - **Relationship:** Associative.
 - **Logic:** Hormonal differences (estrogen vs. testosterone) influence baseline lipid profiles (HDL/LDL ratios).
- 6. Fasting Blood Sugar (fbs) → 14. Disease Target (num)
 - **Relationship:** Direct Risk.
 - **Logic:** fbs > 120 indicates Diabetes. Diabetes accelerates atherosclerosis (plaque buildup), increasing the long-term probability of disease.
- 5. Cholesterol (chol) → 14. Disease Target (num)
 - **Relationship:** Weak Direct.
 - **Logic:** While high cholesterol is the raw material for plaque, it is a poor short-term predictor compared to Fluoroscopy or Angina. Many people live with high cholesterol without developing acute blockage.

6. Summary Table: The Full 14-Attribute Model

ID	Attribute	Full Name	Role in Network	Relationship Type	Connects To (Child)
1	age	Age	Root Cause	Inverse / Direct	thalach (Inv), trestbps (Dir)
2	sex	Sex	Root Cause	Associative	chol, thalach
3	cp	Chest Pain Type	Symptom	Predictive	num, exang
4	trestbps	Resting BP	Intermediate	Causal	restecg
5	chol	Cholesterol	Risk Factor	Weak Direct	num
6	fbs	Fasting Blood Sugar	Risk Factor	Direct Risk	num
7	restecg	Resting ECG	Diagnostic	Associative	num
8	thalach	Max Heart Rate	Symptom	Functional	num
9	exang	Exercise Angina	Symptom	Specific	num
10	oldpeak	ST Depression	Test Result	Strong Direct	num
11	slope	Slope of Peak ST	Modifier	Structural	oldpeak
12	ca	Fluoroscopy Vessels	Test Result	Strongest Direct	num
13	thal	Thallium Stress	Test Result	Strong Direct	num
14	num	Target Diagnosis	Outcome	N/A	(None)