

Lecture 06 - Contents

An overview of the parts in the clinical Chain-of-Thought lecture.

Part 1

Reasoning Fundamentals

Part 2

Advanced CoT Techniques

Part 3

Clinical Applications

Hands-on

CoT Implementation Hands-on

This outline is for guidance. Navigate the slides with the left/right arrow keys.

Lecture 6:

Chain-of-Thought: Teaching LLMs Clinical Reasoning

Ho-min Park

homin.park@ghent.ac.kr

powersimmani@gmail.com

Chain-of-Thought in Medical Context

CoT Prompting

- Step-by-step reasoning process
- Makes LLM thinking transparent
- Breaks complex problems into steps
- Intermediate reasoning shown explicitly

Medical Applications

- Diagnosis generation
- Treatment planning
- Clinical decision support
- Patient assessment

Standard vs CoT Comparison

Standard Prompting

Direct answer only

VS

CoT Prompting

Reasoning + Answer

Performance Improvement: **20-80% accuracy increase on complex medical reasoning tasks**

Lecture 6: Chain-of-Thought for Clinical Reasoning

Part 1 Reasoning Fundamentals

- Clinical Reasoning Process
- Diagnostic Reasoning Chains
- Zero-Shot CoT Medical
- Few-Shot CoT Examples
- Self-Consistency Voting
- Reasoning Path Selection
- Medical Knowledge Integration

Part 2 Advanced CoT Methods

- Tree of Thoughts Medical
- Graph of Thoughts
- Differential Diagnosis Trees
- Bayesian Reasoning Integration
- Uncertainty Propagation
- Counterfactual Reasoning

Part 3 Clinical Applications

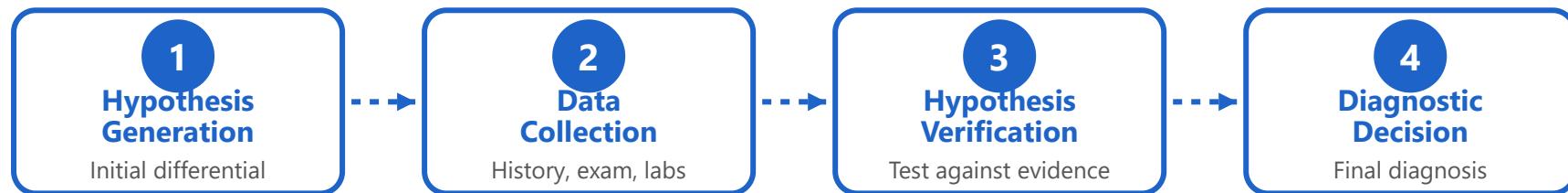
- Emergency Triage CoT
- Treatment Planning Chains
- Drug Interaction Reasoning
- Lab Result Interpretation
- Clinical Note Generation
- Error Analysis Patterns
- Reasoning Verification
- Performance Comparison
- Case Studies Analysis
- Hands-on Implementation
- Best Practices Guidelines

Part 1/3:

Clinical Reasoning Fundamentals

- 1.** Clinical Reasoning Process
- 2.** Diagnostic Reasoning Chains
- 3.** Zero-Shot CoT in Medical Context
- 4.** Few-Shot CoT Examples
- 5.** Self-Consistency & Voting
- 6.** Reasoning Path Selection
- 7.** Medical Knowledge Integration

Clinical Reasoning Process

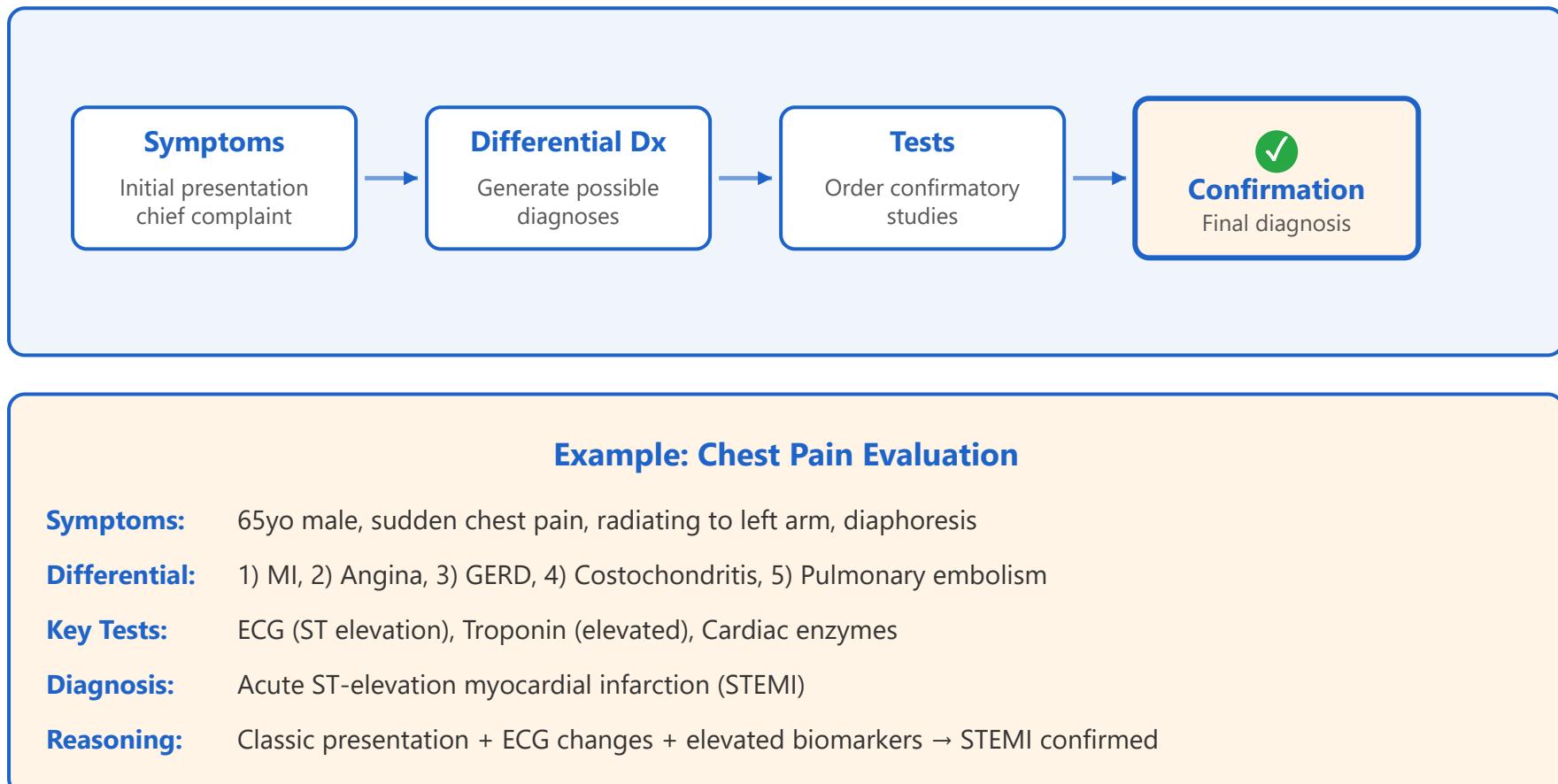


Iterative Clinical Reasoning Cycle



Key Insight: Clinical reasoning is iterative and hypothesis-driven, continuously updating based on new evidence

Diagnostic Reasoning Chains



Zero-Shot Chain-of-Thought in Medical Context

Standard Medical Query:

Patient presents with fever (39°C), productive cough, and pleuritic chest pain.
What is the most likely diagnosis?

"Let's think step by step"

Zero-Shot CoT Medical Prompt:

Patient presents with fever (39°C), productive cough, and pleuritic chest pain.
What is the most likely diagnosis? **Let's think step by step.**

✓ Success Case

Query: Fever + cough + pleuritic pain
With CoT: Systematic evaluation of symptoms → ruled out alternatives → pneumonia diagnosis
Accuracy: 85% → 92%

✗ Limitation Case

Query: Rare genetic syndrome
With CoT: Detailed reasoning but incorrect conclusion
Issue: Limited knowledge, not reasoning problem

Few-Shot Chain-of-Thought Examples

Few-Shot CoT Structure

2-5 Examples: Provide diverse medical cases

Show Reasoning: Include step-by-step clinical thinking

Format Consistency: Use same structure across examples

Example Template

Case: [Patient presentation]

Step 1: Identify key symptoms and risk factors

Step 2: Generate differential diagnosis list

Step 3: Determine discriminating tests

Step 4: Integrate results and reach conclusion

Diagnosis: [Final diagnosis]

Key: Examples teach the LLM both the format and the medical reasoning process

Self-Consistency & Voting Mechanism

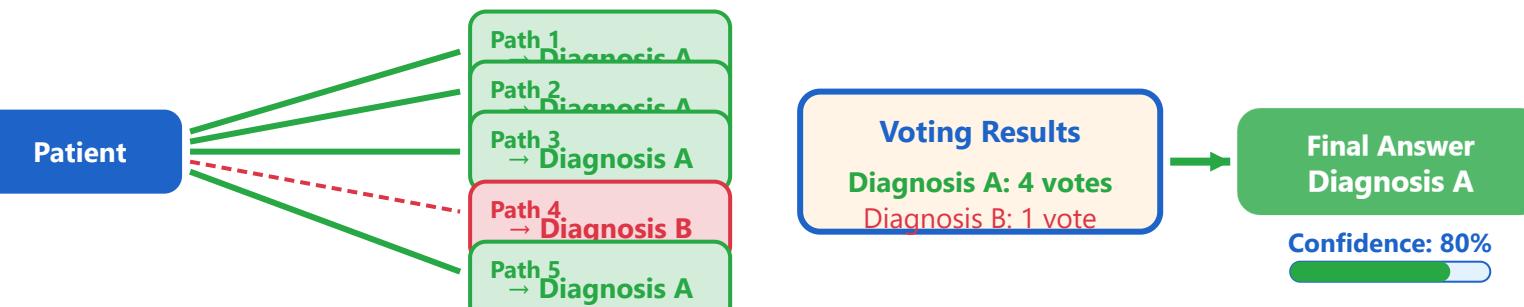
Multiple Reasoning Paths

Generate Multiple Solutions: Sample 5-10 different reasoning paths

Voting Mechanism: Select most consistent answer across paths

Confidence Boost: Agreement increases reliability

Example: 5 Reasoning Paths



Performance improvement: **5-15% accuracy gain** over single-path CoT

Reasoning Path Selection & Optimization

🎯 Path Quality Metrics

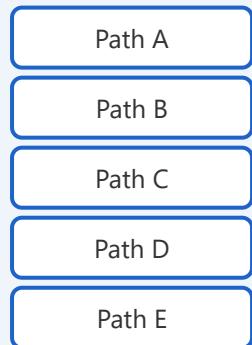
- Logical consistency
- Evidence alignment
- Medical accuracy
- Completeness of reasoning

🔍 Selection Strategies

- Confidence scoring
- Expert validation
- Guideline adherence
- Outcome verification

Path Selection & Pruning Process

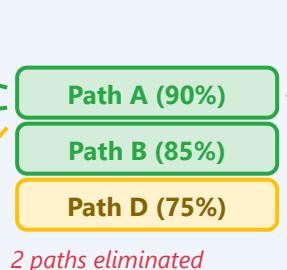
1. Generate



2. Assess Quality



3. Prune & Select



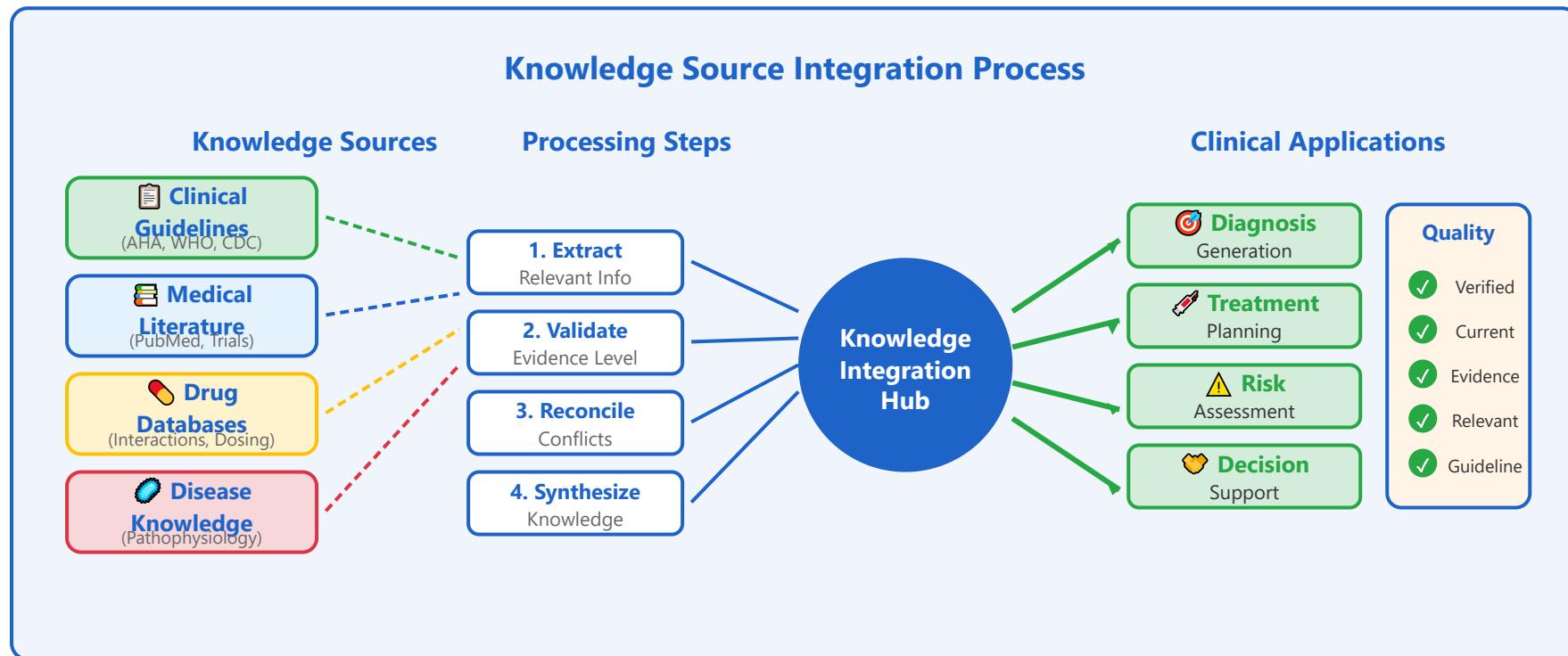
4. Final Selection

✓ SELECTED
Path A (Best Score)



Pruning Strategy: Eliminate paths with logical contradictions or unsupported claims early

Medical Knowledge Integration



Integration Example: Hypertensive Emergency

Patient Case: BP 220/120, chest pain, confusion

Guideline Integration: JNC-8 → Target BP reduction 25% in first hour

Evidence Integration: Recent meta-analysis → IV labetalol preferred

Drug Database: Check contraindications, adjust for CrCl

Output: Evidence-based treatment protocol with monitoring plan

Critical: Always reference authoritative medical sources

Part 2/3:

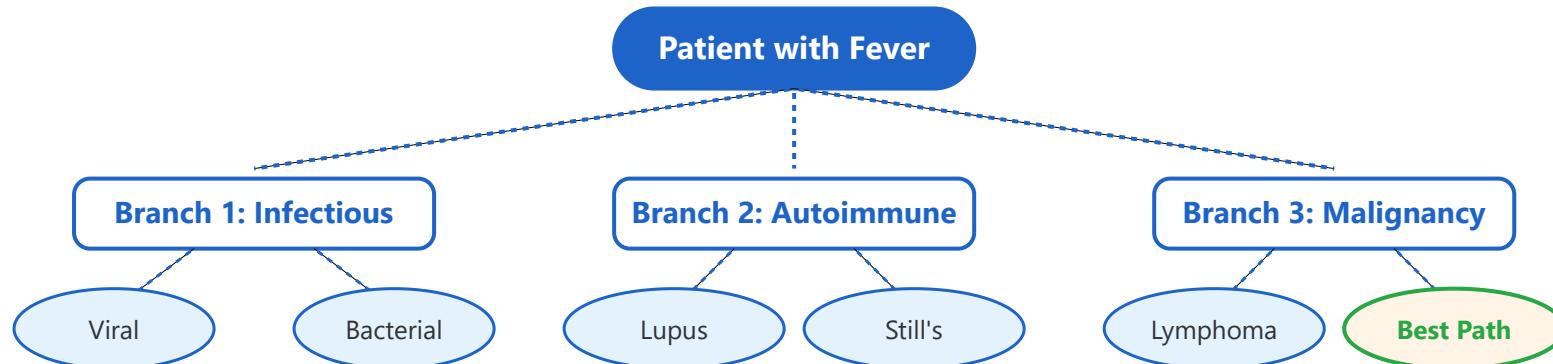
Advanced CoT Methods

- 1.** Tree of Thoughts for Medical Reasoning
- 2.** Graph of Thoughts
- 3.** Differential Diagnosis Trees
- 4.** Bayesian Reasoning Integration
- 5.** Uncertainty Propagation
- 6.** Counterfactual Reasoning

Tree of Thoughts (ToT) for Medical Reasoning

ToT Structure

- **Branching Exploration:** Multiple diagnostic pathways simultaneously
- **Depth-First/Breadth-First:** Systematic exploration strategies
- **Backtracking:** Return to earlier decision points if path fails
- **Evaluation:** Score each branch based on evidence



Advantage: Explores multiple hypotheses systematically before committing to one

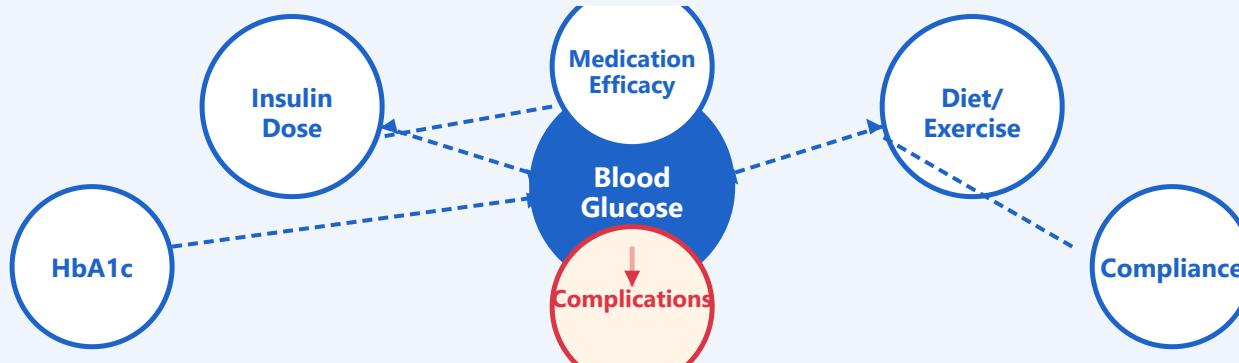
Graph of Thoughts (GoT)

Graph Structure

- Non-linear reasoning paths
- Cyclic relationships
- Bidirectional connections
- Complex interdependencies

Medical Applications

- Comorbidity interactions
- Drug-drug interactions
- Multi-system diseases
- Feedback loops in treatment



Key: Captures complex medical relationships beyond simple trees

Differential Diagnosis Trees

Probabilistic Branching

- **Prior Probabilities:** Based on prevalence and risk factors
- **Likelihood Ratios:** Each test updates probabilities
- **Bayesian Updates:** Continuous refinement with new data
- **Threshold Decision:** Diagnostic certainty level to act

Example: Chest Pain Differential

Initial:

MI (30%)

Angina (25%)

PE (15%)

GERD (20%)

MI (92%)

After ECG:

MI (65%)

Angina (20%)

PE (10%)

GERD (3%)

Angina (5%)

Others (3%)

After Troponin:

Treat as MI (>90% threshold)

Quantitative reasoning: Numbers guide clinical decisions

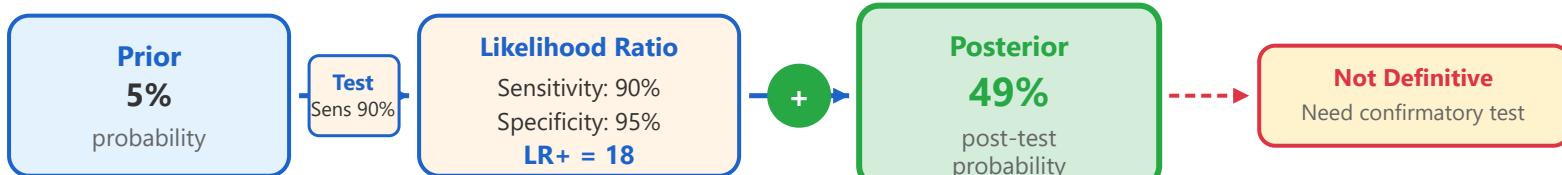
Bayesian Reasoning Integration

Bayesian Framework

- **Prior Probability:** Pre-test likelihood based on prevalence
- **Likelihood Ratio:** How much test changes probability
- **Posterior Probability:** Post-test diagnostic certainty
- **Sequential Testing:** Chain multiple tests together

Bayes' Theorem in Diagnosis

$$P(\text{Disease}|\text{Test}) = [P(\text{Test}|\text{Disease}) \times P(\text{Disease})] / P(\text{Test})$$



Critical: Integrating prevalence and test characteristics improves diagnostic accuracy

Uncertainty Propagation in Clinical Reasoning

Sources of Uncertainty

- Incomplete patient history
- Test measurement error
- Diagnostic ambiguity
- Treatment response variability

Managing Uncertainty

- Confidence intervals
- Probability distributions
- Decision thresholds
- Risk-benefit analysis

Propagation Example

Initial
Symptom Interpretation
 $\pm 20\%$

After Test
Diagnosis
 $\pm 15\%$

Treatment
Outcome Prediction
 $\pm 25\%$

Total Uncertainty: Compounds through reasoning chain

Action: Make decisions robust to uncertainty levels

Counterfactual Reasoning in Clinical Context

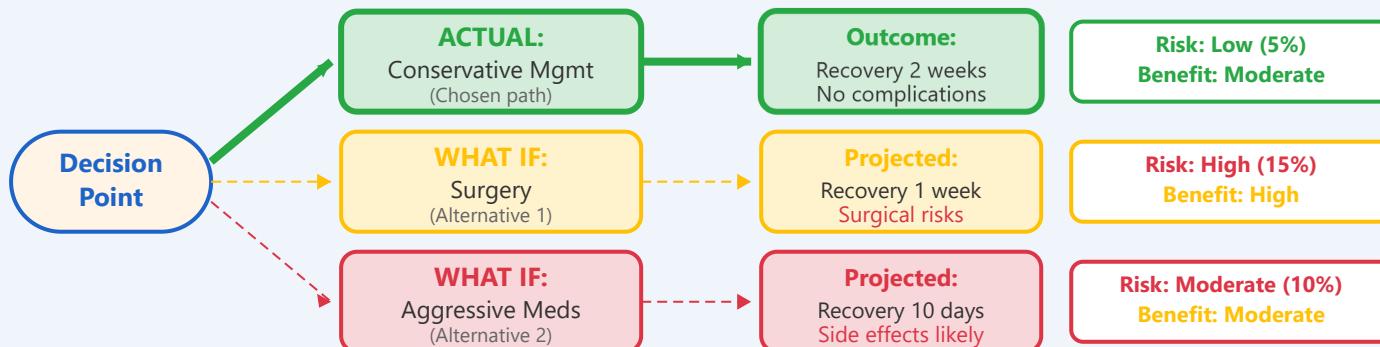
"What If" Scenarios

- Alternative treatments
- Timing variations
- Different diagnostic paths
- Modified risk profiles

Analysis Benefits

- Validate decisions
- Identify improvements
- Risk-benefit clarity
- Learning from outcomes

Counterfactual Decision Analysis



Analysis Result: Conservative management was optimal given risk-benefit profile ✓

Benefit: Exploring alternatives validates chosen approach and identifies improvements

Part 3/3:

Real Clinical Applications

- 1.** Emergency Triage with CoT
- 2.** Treatment Planning Chains
- 3.** Drug Interaction Reasoning
- 4.** Lab Result Interpretation
- 5.** Clinical Note Generation
- 6.** Error Analysis & Verification
- 7.** Performance & Case Studies
- 8.** Hands-on Implementation

Emergency Triage with Chain-of-Thought

Triage CoT Process

- **Rapid Assessment:** ABC (Airway, Breathing, Circulation) first
- **Priority Scoring:** ESI (Emergency Severity Index) 1-5
- **Resource Prediction:** Anticipate needed tests/interventions
- **Time-Critical Flags:** STEMI, stroke, sepsis alerts

Example: Chest Pain Triage

Step 1: Vital Signs
BP 160/95, HR 110
→ **concerning**

Step 2: Symptoms
substernal pressure
→ **cardiac?**

Step 3: Risk Factors
age 65, DM, HTN
→ **high risk**

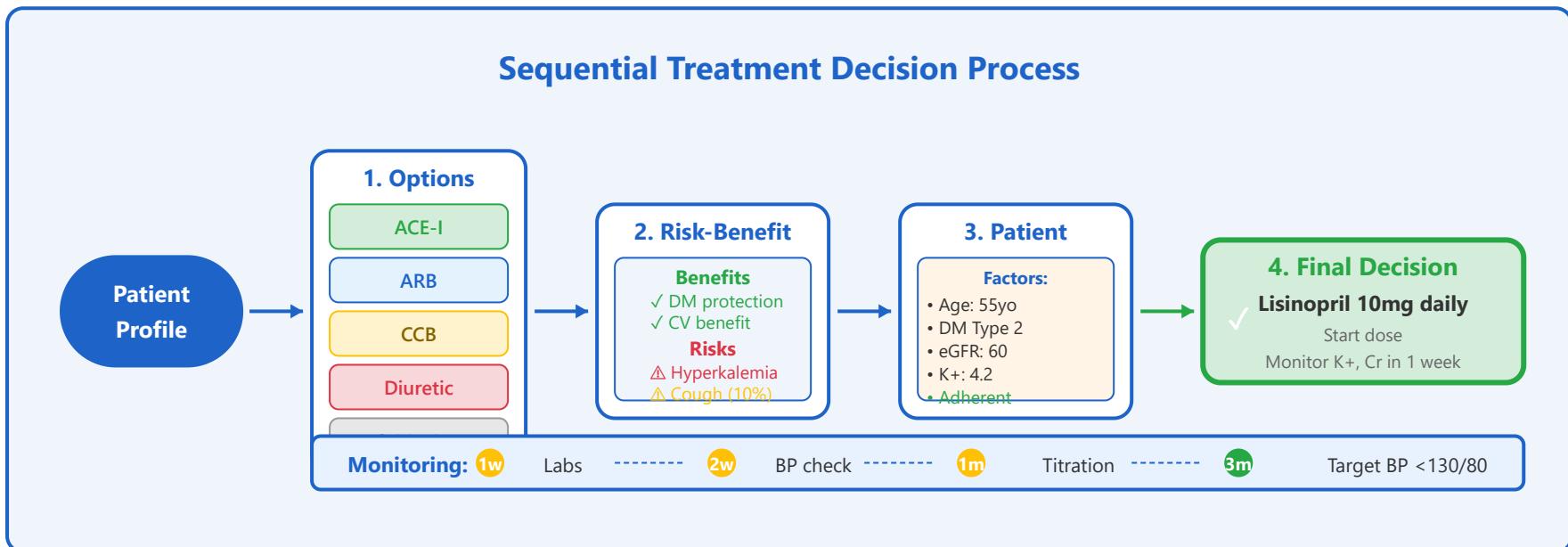
Step 4: Alert!
Red flags present
→ potential ACS

Triage: ESI Level 2 - IMMEDIATE
Bedside ECG, labs, cardiac monitoring NOW



Critical: CoT enables rapid, systematic prioritization under time pressure

Treatment Planning with CoT

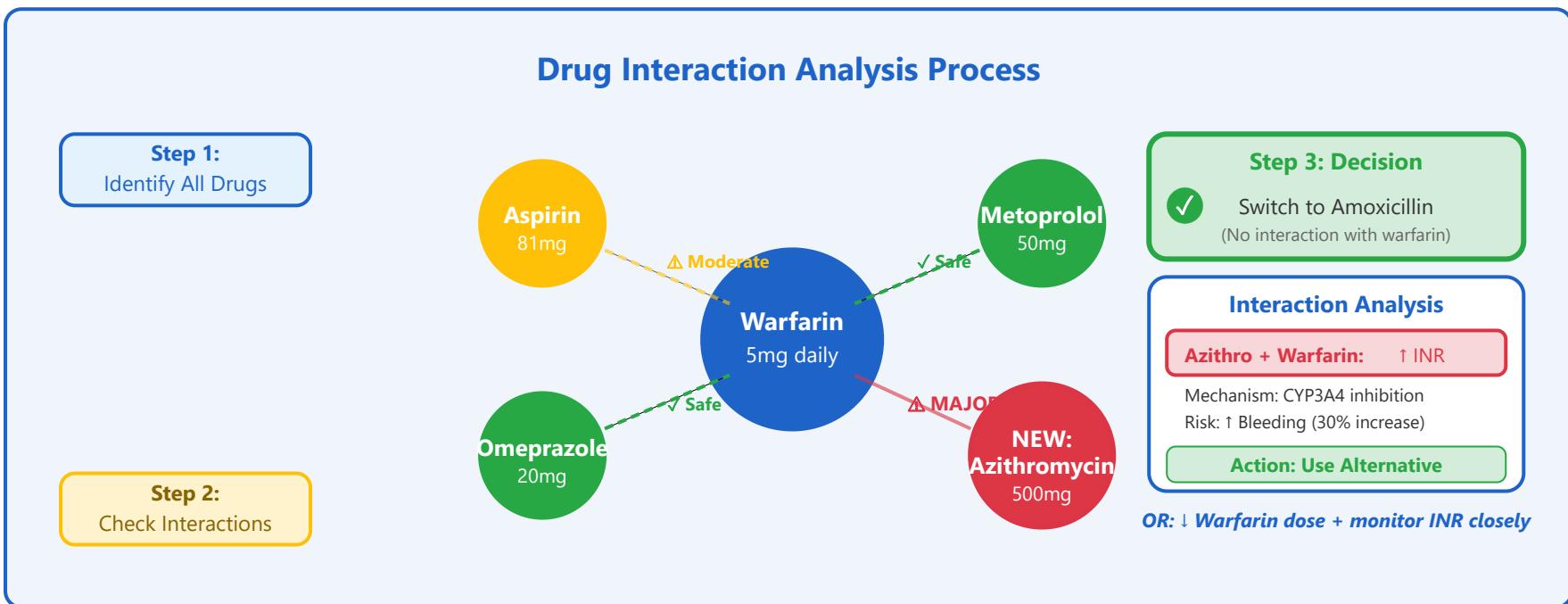


CoT Reasoning Chain

- Step 1:** List all antihypertensive options → ACE-I, ARB, CCB, Diuretic, β-blocker
- Step 2:** Consider comorbidities → DM present → ACE-I/ARB preferred (renal protection)
- Step 3:** Check contraindications → K+ normal, eGFR acceptable → No contraindications
- Step 4:** Select based on guidelines + patient factors → ACE-I (Lisinopril) first-line
- Step 5:** Define monitoring plan → Labs at 1 week, BP at 2 weeks, titrate monthly

Guideline-based + personalized = optimal treatment

Drug Interaction Reasoning with CoT

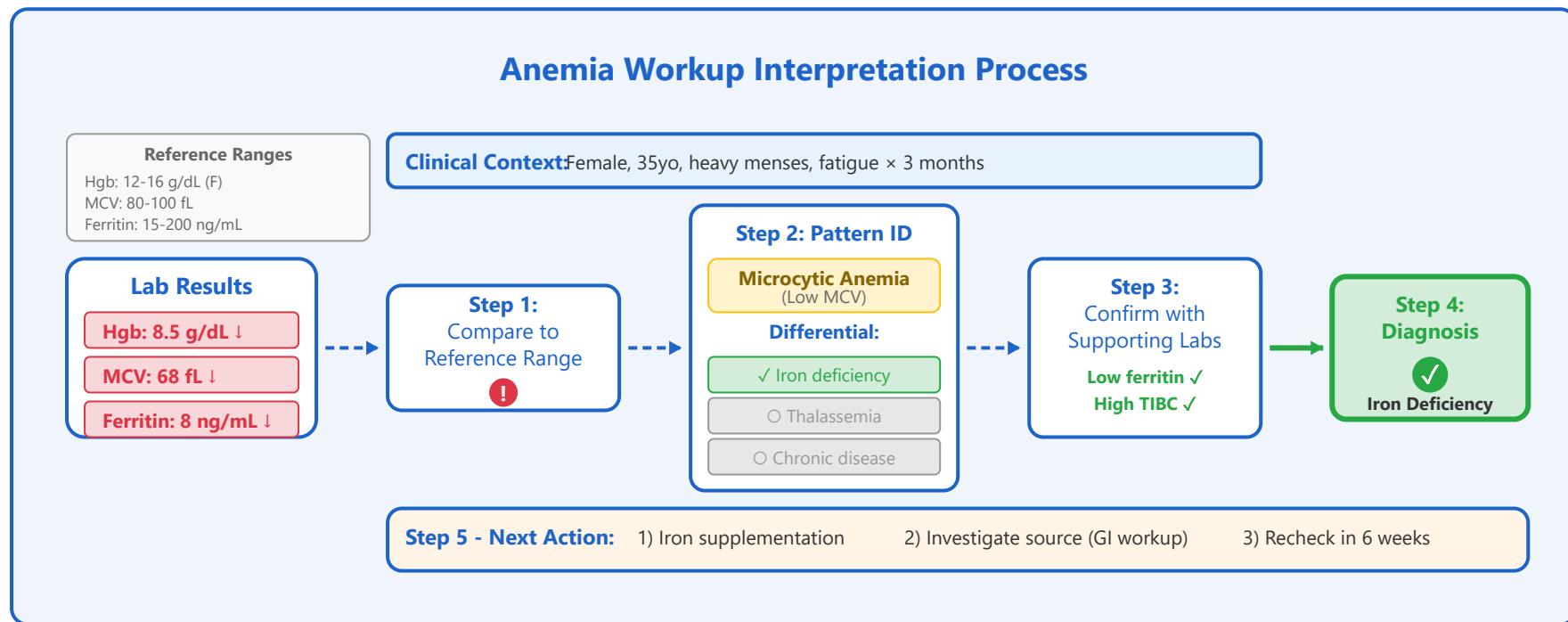


CoT Reasoning Steps

- 1. Identify:** List all current medications → Warfarin, Aspirin, Metoprolol, Omeprazole
- 2. New drug:** Azithromycin for pneumonia → Check database for interactions
- 3. Analysis:** Major interaction found → CYP3A4 inhibition → ↑ warfarin levels → ↑ bleeding risk
- 4. Decision:** Use alternative (Amoxicillin) OR adjust warfarin + intensive monitoring

Safety First: CoT catches potential harms before they occur

Lab Result Interpretation with CoT



CoT Reasoning Steps

Step 1: Anemia confirmed (Hgb 8.5 < 12 g/dL) → Severity: moderate

Step 2: Microcytic (MCV 68 < 80 fL) → DDx: Fe deficiency, thalassemia, chronic disease

Step 3: Low ferritin (8 ng/mL) + high TIBC → Confirms iron deficiency

Step 4: Etiology? → Female, menorrhagia likely → Consider GI loss if persists

Diagnosis: Iron deficiency anemia → Treat + investigate source

Integration: Labs + clinical picture = complete assessment

Clinical Note Generation with CoT

SOAP Note with CoT-Enhanced Reasoning

S - Subjective

CC: "Chest pain × 2 hours"
HPI: Substernal pressure, radiates to L arm, diaphoresis
PMH: HTN, DM, dyslipidemia
Meds: Metformin, Lisinopril

O - Objective

VS: BP 160/95, HR 110, O₂ 96%
PE: Diaphoretic, chest clear
ECG: ST elevation V2-V4
Labs: Troponin 2.4 (↑)
CXR: No acute findings

A - Assessment (CoT)

Reasoning Chain:

1. Chest pain + diaphoresis
→ cardiac vs non-cardiac
2. ST elevation + 1 troponin
→ confirms STEMI
3. Risk factors: DM, HTN
→ high TIMI score (5)

P - Plan

IMMEDIATE:

- Cath lab activation STAT
- ASA 325mg, Ticagrelor 180mg
- Heparin bolus + gtt
- Morphine PRN pain
- Cardiology consult

CoT Integration Benefits:

- ✓ Transparent reasoning
- ✓ Auditable decisions
- ✓ Teaching tool
- ✓ Reduces errors
- ✓ Legal documentation

CoT-Enhanced Assessment Example:

"Given presentation of substernal chest pressure + radiation to left arm + diaphoresis (classic ACS triad) + ECG showing ST elevation V2-V4 (anterior wall) + troponin elevation (2.4, significant) + TIMI score 5 (high risk), this is consistent with acute STEMI. DDx considered: unstable angina (ruled out by troponin), PE (low Wells score), GERD (atypical presentation)."

Documentation: CoT makes clinical reasoning explicit and auditable

Error Analysis in CoT Reasoning

⚠ Common Error Types

- Premature closure (stopping too early)
- Anchoring bias (fixation on initial diagnosis)
- Missing rare conditions
- Ignoring contradictory evidence

🔧 Improvement Strategies

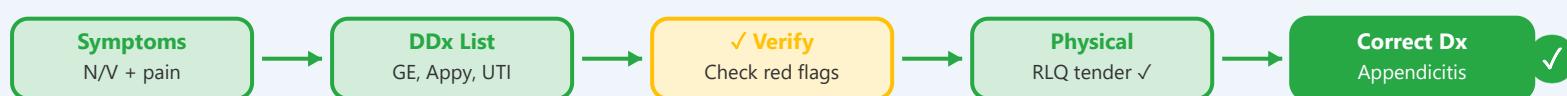
- Force consideration of alternatives
- Explicit verification steps
- Checklist integration
- Human-AI collaboration

Error Detection & Correction Flow

Error Example: Misdiagnosed Appendicitis



Corrected with CoT Enhancement



Root Cause: Premature closure after matching 3 symptoms

Fix: Add mandatory "red flag check" step before finalizing diag.

Results 50% error reduction

Learning: Analyze failures to improve CoT prompting strategies

Reasoning Verification & Quality Control

Verification Steps

Logical Consistency: Check for contradictions in reasoning chain

Evidence Support: Verify each claim has supporting data

Completeness: Ensure all relevant factors considered

Guideline Adherence: Confirm alignment with clinical standards

Verification Checklist

- ✓ Are all steps logically connected?
- ✓ Is each conclusion supported by evidence?
- ✓ Were alternative diagnoses considered?
- ✓ Are contradictory findings explained?
- ✓ Does final answer match clinical guidelines?
- ✓ Are safety concerns addressed?

Quality Assurance: Systematic verification reduces errors by 30-40%

CoT vs Standard Prompting Performance

Standard Prompting

- Direct answer generation
- No intermediate steps
- Faster but less accurate
- Harder to debug errors

CoT Prompting

- Step-by-step reasoning
- Transparent logic
- Higher accuracy
- Auditable process

Benchmark Results

MedQA: 56% → 72% (+16% with CoT)

PubMedQA: 68% → 81% (+13% with CoT)

Clinical Diagnosis: 62% → 78% (+16% with CoT)

Treatment Planning: 70% → 84% (+14% with CoT)

Average Improvement: +15% accuracy across medical tasks

Trade-off: 2-3x slower but significantly more accurate and trustworthy

Real-World Case Studies

Case 1: Sepsis Detection

Presentation: 72yo, fever 39.2°C, confusion, HR 125, BP 85/50

CoT Process: SIRS criteria → infection source → SOFA score → sepsis diagnosis

Outcome: Early recognition → prompt antibiotics → improved survival

Case 2: Atypical Presentation

Presentation: 45yo female, fatigue, no chest pain, mild dyspnea

CoT Process: Atypical symptoms → considered MI (women present differently) → ECG → troponin → diagnosed NSTEMI

Outcome: CoT helped recognize atypical presentation, standard prompting missed it

Case 3: Complex Comorbidity

Presentation: DM + CKD + CHF patient with worsening shortness of breath

CoT Process: Multiple potential causes → systematically evaluated each → identified CHF exacerbation + UTI

Outcome: Treated both conditions appropriately, avoided missing secondary diagnosis

Lesson: CoT excels at complex, atypical cases requiring careful reasoning

Hands-on: Implementing CoT for Medical Tasks

Implementation Steps

- 1. Define Task:** Diagnosis, treatment planning, triage, etc.
- 2. Design Prompt:** Include "think step by step" or provide examples
- 3. Test & Iterate:** Evaluate on sample cases, refine prompt
- 4. Validate:** Compare to expert clinicians, measure accuracy

Code Example (Python)

```
prompt = """  
Patient: {patient_info}  
Task: Provide differential diagnosis.  
Let's think step by step:  
1. Key symptoms and signs  
2. Possible diagnoses  
3. Tests to discriminate  
4. Most likely diagnosis  
"""  
  
response = model.generate(prompt)
```

Practice: Try implementing CoT for your own clinical scenarios

Best Practices for Clinical CoT

✓ DO

- Use clear, explicit step labels
- Reference clinical guidelines
- Verify reasoning logic
- Include safety checks
- Test on diverse cases

X AVOID

- Vague instructions
- Too many steps (>7-8)
- Ignoring contradictions
- Blind trust in output
- Skipping validation

Key Principles

- 1. Clarity:** Each step should be clear and actionable
- 2. Completeness:** Cover all critical decision points
- 3. Safety:** Always include risk assessment
- 4. Evidence:** Ground reasoning in medical literature
- 5. Validation:** Human expert review essential

Remember: CoT is a tool to augment, not replace, clinical judgment

Thank you

Chain-of-Thought for Clinical Reasoning

Ho-min Park

homin.park@ghent.ac.kr

powersimmani@gmail.com