

# The Benefit of Diagnostic Hypotheses in Clinical Reasoning: Experimental Study of an Instructional Intervention for Forward and Backward Reasoning

Norman, G. R., Brooks, L. R., Colle, C. L., and Hatala, R. M  
Cognition and Instruction, 17(4), 433-448

## 1. はじめに

### 1.1. 検討する問題

- 医学教育における仮説とデータ (特徴) の関係 (Elstein, Shulman, & Sprafka, 1978)
    - ◇ 立場 1: 体系的にデータを集めた後に診断を下すように教育するのが良い
    - ◇ 立場 2: 一時的な仮説を頭においたのちに, データを探すように教育するのが良い
- 診断に及ぼす病歴の影響や熟達化による方略の差異に関する研究が行われてきた

### 1.2. 先行研究 1: 病歴が診断に及ぼす影響

- 放射線診断における病歴の効果 (Barbaum et al., 1986; Schreiber, 1963; Swensson, Hessel, & Herman, 1985)
  - ◇ フィルムを見る前に病歴から一時的な仮説を立てるのが良いのかどうか
  - ◇ 様々な立場が存在 (仮説を立てるほうが良いときもあれば, 良くないときもある)
- ECG (心電図) 診断における病歴の効果 (Hatala, Norman, & Brooks, 1999)
  - ◇ 特定の仮説を暗示する病歴が最終的な診断に及ぼす効果
    - 専門医 11%, 研修医 29%, 学生 31%
    - 学生だけでなく, 専門医も病歴からの影響を受ける
    - 病歴が診断を正確にする場合もあれば, 悪くする場合もある
- 病歴の影響に関する 3 つの考え方
  1. 最終的な診断のみに影響 . 病歴によって仮説の確信が高まる . バイアスではない
  2. データ取得に影響 . 仮説駆動的な見方を促す . 確証バイアスとなる
  3. データ解釈に影響 . 得られたデータの解釈がゆがめられる
- 上記 2・3 を支持する証拠 (Hatal et al., 1999; Norman, Brooks, Coblenz, Babcook, 1992)
  - ◇ 最終的な診断結果だけでなく, 抽出される特徴にも病歴は影響を与える

### 1.3. 先行研究 2: 熟達化による方略の差異

- 力学の問題解決における熟達化 (Larkin, McDermott, Simon, & Simon, 1980 p. 319)
  - ◇ ノービスは解を求めるために, 得られている量を解釈
  - ◇ エキスパートは着実に与えられたものから求める量を導く

前向き推論は熟達化に必要な能力?
- 医学 (カルテからの診断) における熟達化 (Patel & Groen, 1986; Patel, Groen, & Arocha, 1990)

- ◇ ノービス: 後ろ向き推論・後ろ向き推論と前向き推論の混合方略「感染症なら熱がある」
- ◇ エキスパート: 前向き推論「熱があるなら感染症」  
熟達化にしたがって、後ろ向き推論を身に着ける？病歴は特徴抽出のバイアス？

#### 1.4. 特徴の曖昧さが推論方略の良さに関係している可能性

- 前向き推論の結果、曖昧だが重要な特徴は見落としがある
  - ◇ 仮説なしには、曖昧な特徴を見ることができない
- 前向き推論の結果、特徴の拾いすぎがある
  - ◇ 曖昧な特徴を見逃すまいとし、有意ではない特徴を拾ってしまう
- 後ろ向き推論は曖昧な特徴を歪める
  - ◇ 仮説検証を目的に特徴を見る結果、確認バイアスが生じる  
前向き・後ろ向き共に、ネガティブな側面があると考えられる

#### 1.5. 特徴の曖昧さとカテゴリ化に関する研究 (Wisniewski and Medin, 1994)

- 特徴とカテゴリを結ぶルールを発見する課題
  - ◇ 先行知識(常識知識)を喚起するようなカテゴリ(都会の子供 vs. 田舎の子供)
  - ◇ 曖昧な特徴セット
  - ◇ 先行知識と真のルールは無関連
  - ◇ 被験者にはフィードバックのみを与える
- 現実の医療診断には適合しない手続き
  - ◇ 特徴と疾患の統計的な組み合わせを医師は既に知っている
  - ◇ 医師が行うことは特定の患者における診断を行うこと

#### 1.6. 本研究の目的

- 前向き推論と後ろ向き推論を教示により操作
- 前向き推論と後ろ向き推論のどちらが優れているかを検証
- 領域として ECG 診断を設定
  - ◇ 著者たち自身が研究を行った経験をもち、この領域に関する知識があるから
  - ◇ 視覚的な素材を用いた診断(放射線医学・病理診断)に比べ、特徴の曖昧さが低い(ECG 診断における自動診断は画像診断における自動診断に比べ、実用的)
  - ◇ 実験結果の一般化をしやすい

## 2. 実験 1

- 教示による方略の操作によって診断の正確さが変化するかを検討

## 2.1. 方法

### 2.1.1. 被験者

- McMaster University の心理学部の大学生: 16 人
    - ◇ D/B ( diagnosis/backward-reasoning ) 群: 8 人
    - ◇ F/F ( feature/forward-reasoning ) 群: 8 人
- 医学部の学生は診断方略に関する教育を受けた可能性があるため心理学部の学生を採用

### 2.1.2. 材料

- 第 1 冊子
  - ◇ ECG の読み方
  - ◇ ECG の導入と基礎用語の説明 ( 正常 ECG , ST 降下 , ST 上昇 )
- 第 2 冊子
  - ◇ MacMaster University 医学部で使用されているマニュアル
  - ◇ 代表的な心臓病関連の ECG  
RVH , LVH, LVHS , LBBB , RBBB , AAMI , AIMI , APMI , I , P , H
  - ◇ 全 5 章から校正  
“ 心室肥大 ” , “ 血管拘束 ” , “ 心筋梗塞 ” , “ 心筋虚血 ” , “ その他 ”  
各章に練習課題 ( ECG の提示 )
  - ◇ 疾患ごとの特徴リスト ( LVH の例 )
    - V1・V2 における S 波が大きい
    - V5・V6 における R 波が大きい
    - R 波と S 波の合計が 35mm 以上
- 回答冊子
  - ◇ 特徴チェックリスト ( Appendix )
- 問題冊子 1 ( 学習を促すためのクイズ )
  - ◇ トレーニングで学習した 10 症例の確認
- 問題冊子 2 ( テスト )
  - ◇ まだ学習していない ECG 10 症例

### 2.1.3. 実験デザインと手続き

- 総実験時間は 2 時間 ~ 3 時間 20 分
- 1. トレーニングフェーズ 1 ( F/F 群・D/B 群で共通 )
  - ◇ 実験者と 1 対 1 で学習
    - A) 第 1 冊子を使って ECG の導入
    - B) 疾患別の特徴リストの提示
    - C) 11 疾患の見分け方を第 2 冊子を使って学習

- D) 章ごとに小テスト・フィードバック
2. トレーニングフェーズ 2 (F/F 群・D/B 群で共通)
    - ◇ 問題冊子 1 を使用して診断経験を積む
  3. テストフェーズ (F/F 群・D/B 群で異なる教示)
    - ◇ 問題冊子 2 を使用して診断を行う
    - D/B 群への教示.** 疾患リストと ECG を見ながら診断してください。診断を終えたのちに、特徴をチェックリストに記述してください (疾患名がリストされた用紙 [疾患リスト] を始めに渡す。その間、特徴のチェックリストを示さない)
    - F/F 群への教示.** 各 ECG について、特徴のチェックリストを埋めなさい。リストを完成させたら、ECG を取り除き、診断をしてもらいます。疾患リストを使って診断をしてください。(被験者が特徴をチェックしている間、疾患リストを見せない。最終的な診断を下すときに ECG を参照できない。疾患リストを見るまえに、特徴のチェックを完了する必要)

## 2.2. 分析

- 診断の正確さ (正解/不正解)
- 各症例で同定された特徴を分類
  1. Consistent: ECG に存在する特徴で、かつ正しい診断と整合的
  2. Errors: ECG 上にない特徴
  3. Present but irreverent: ECG には存在するけど、正しい診断と不整合な特徴
  4. Others: ECG に存在するけど、一般的過ぎて診断に使えない特徴

## 2.3. 結果

### 2.3.1. 診断の正確さを指標とした $2 \times 10$ AVOVA (群 $\times$ 症例)

- 群間の差が有意 ( $F(1,14) = 7.04, p < .02$ )
  - ◇ D/B 群の正解率 = 61.3%
  - ◇ F/F 群の正解率 = 41.9%
- 症例の主効果も有意だったが、考察の対象とはしない (症例間での難易度が異なる)
- 交互作用は有意ではない

### 2.3.2. 総特徴数を指標とした $2 \times 10$ AVOVA (群 $\times$ 症例)

- 群間の差が有意傾向 ( $F(1,14) = 4.48, p < .06$ )
  - ◇ D/B 群の総特徴数 = 3.45
  - ◇ F/F 群の総特徴数 = 4.64
- 症例の主効果も有意だったが、考察の対象とはしない (症例間での難易度が異なる)
- 交互作用は有意ではない

### 2.3.3. 個別カテゴリの特徴数を指標とした $2 \times 10$ AVOVA (群 $\times$ 症例)

- Present irreverent での群間の差のみが有意 ( $F(1,14) = 17.57, p < .001$ )  
F/F 群は D/B 群に比べて不適切な特徴を多く挙げた

TABLE 1  
Experiment 1: Mean Number and Standard Deviation of Features per Case by  
Type of Feature and Condition

	<i>Consistent</i>		<i>Error</i>		<i>Present Irrelevant</i>		<i>Other</i>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Diagnose/Backward	1.18	0.27	0.88	0.92	0.98	0.35	0.41	0.17
Feature/Forward	1.26	0.30	1.27	0.58	1.70	0.34	0.41	0.19

## 2.4. 考察

- 実験 1 の結果, 望ましい診断方略が後ろ向き推論であることが示された
  - ◇ D/B 群の診断成績は 50% 以上
  - ◇ D/B 群は無関連な特徴をとってこない
- F/F 群において最終診断時に ECG を取り除いたことの問題点
  - ◇ D/B 群は最終診断時に ECG を参照可能であったため, 前向き推論と後ろ向き推論の両方が行われた?
  - ◇ F/F 群は最終診断時に ECG を取り除かれたため, 特徴の顕著性や明瞭さまでも覚えておかなければならなかった?

## 3. 実験 2

- 特徴の顕著性や明瞭さを覚えることができなかつたから診断成績が低かつたのか
- 一度同定した特徴を無視することができなかつたために診断成績が落ちたのか  
F/F 群への実験手続きを変更した群を設定

### 3.1. 方法

#### 3.1.1. 被験者

- McMaster University の心理学の学生: 24 人
  - ◇ D/B (diagnosis/backward-reasoning) 群: 8 人
  - ◇ F/Fm (modified feature/forward-reasoning) 群: 16 人  
検定力を挙げるために F/Fm 群の被験者を増やした

#### 3.1.2. 材料

- 実験 1 と疾患リスト・解答用紙以外は同様
  - ◇ 2 つの特徴をリストから削除 (Q 波が存在しない, R 波が低い)

実験 1 の被験者にとって理解が困難であり，診断ともあまり関係しない

### 3.1.3. デザインと手続き

- テストフェーズにおける F/Fm 群のみ違い

F/Fm 群に対する教示：各 ECG について，解答用紙を埋めなさい。どの特徴が存在しますか。リストを完成させたら，診断をしてもらいます。疾患リストを使ってください（実験 1 とは異なり，最終診断時に ECG を見ることができた）

## 3.2. 結果

### 3.2.1. 診断の正確さを指標とした 2×10AVOVA（群×症例）

- 群間の差が有意 ( $F(1,14) = 5.40, p < .05$ )
  - ◇ D/B 群の正解率 = 61.9%
  - ◇ F/F 群の正解率 = 49.4%
- 症例間の主効果も有意 ( $F(1,14) = 27.88, p < .0001$ )
- 実験 1 の F/F 群との差は有意ではない ( $t(22) = .504, p = .61$ )

### 3.2.2. 総特徴数を指標とした 2×10AVOVA（群×症例）

- 群間の差が有意 ( $F(1,14) = 11.36, p < .01$ )
  - ◇ D/B 群の総特徴数 = 3.87
  - ◇ F/F 群の総特徴数 = 5.22
- 症例間の主効果も有意だったが，考察の対象とはしない（症例間での難易度が違った）
- 交互作用は有意ではない

### 3.2.3. 個別カテゴリの特徴数を指標とした 2×10AVOVA（群×症例）

- Error での群間の差 ( $F(1,14) = 5.89, p < .02$ )
- Present irrelevant での群間の差 ( $F(1,14) = 15.37, p < .001$ )

F/F 群は D/B 群に比べて不適切な特徴を多く挙げた

TABLE 2  
Experiment 2: Mean Number and Standard Deviation of Features per Case by  
Type of Feature and Condition

	Consistent		Error		Present Irrelevant		Other	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Diagnose/Backward	1.31	0.33	1.05	0.80	1.14	0.42	0.37	0.13
Feature/Forward	1.46	0.27	1.66	0.58	1.71	0.41	0.39	0.16

## 3.3. 考察

- 前向き推論の問題点が明らかになった

- ◇ 一旦特徴を集めてしまうと、その特徴を捨てることができない
- ◇ 見つけた特徴を何とか説明するための診断を考えてしまう
- 前向き推論における診断時の ECG の提示が不適切な特徴の棄却を促進させなかった
  - ◇ 実験 2 の F/Fm 群での特徴数 = 5.52
  - ◇ 実験 1 の F/F 群での特徴数 = 4.64
  - 特に不適切な特徴で実験 1 と実験 2 の差が大きい
  - ◇ 内省報告「たくさん特徴を同定したのになぜ診断時に ECG を見なければならぬのか、訳が分からない」

## 4. 総合考察

### 4.1. 実験結果のまとめ

- 実験結果は前向き推論に利点がないことを示す
  - もしかしたら、ECG 中の特徴が曖昧だから前向き推論に利点がないのかもしれない
    - ◇ F/F・F/Fm 群: consistent      irreverent
    - ◇ D/B 群: consistent < irreverent
- 曖昧な状況下では後ろ向き推論はバイアスよりも利点のほうが多いのかもしれない

### 4.2. 本研究の問題点

- 医学部の学生ではない被験者を採用したことによる生態学的妥当性の低下
  - ◇ ECG に関する先行研究 (Hatala et al., 1999) との対比
    - 医学部における学生 36%      F/F・F/Fm 群
    - 研修医 65%      D/B 群
  - 医学部での実際の学習とパフォーマンスに違いはない
- ECG だけでなく、他の医療診断でも同じような結果が得られるのか
  - ◇ 後ろ向き推論の優位性を示す研究
    - 内科検診 (Norman, Brooks, Cunnington, Shai, & Marriott, 1977)
    - ECG (Hatala et al., 1999)
    - 放射線診断 (Norman, Brooks, Coblentx, & Babcook, 1992)
    - 医学における特徴の曖昧さが後ろ向き推論の良さを決定付けている可能性
  - ◇ 医学以外の領域での理論・仮説とデータの関係 Winiewski & Medin (1994)
    - 新たな刺激とカテゴリを与えられ、カテゴリ化のルールを発見する課題
    - 素朴理論に基づいて特徴を解釈しようとする
    - 医学診断との違い: 理論とデータの結びつきにおける曖昧さ
    - 後ろ向きは理論とデータがきちんと定義されているときでも有効なのかも

### 4.3. 熟達化研究への貢献

- 熟達化による前向き推論の増大を示した研究は観察的・横断的研究

- 熟達化に従ったストラテジーの変化の詳細はわからない
  - ◇ 熟達化によって推論方略そのものが変化したのか・記憶構造の変化に伴うものなのか？
  - ◇ エキスパートは記憶構造が良く組織化されているために、暗示的にスキーマが活性化され、スムーズに問題解決されるのか？
  - ◇ Patel らの結果はカルテ診断でオリジナルの曖昧さがないからなのか？

#### 4.4. まとめ

- 本研究は後ろ向き推論と前向き推論を実験的に学習させた初の研究
- たとえエキスパートが前向き推論を日常的に行っているとしても、我々の結果は後ろ向き推論がとてもよいことを示した

#### APPENDIX: Response Sheet

##### LIST OF FEATURES

- Q waves  -present in V1 to V4  
 -present in II, III, AVF
- Predominant R wave in V1
- R waves present in V1 and V2
- Large R waves in V5 and V6
- R wave in V1 > 7 mm
- R wave getting progressively smaller from V2 to V4
- Large S waves in V1 and V2
- Sum of largest S wave in V1 and V2 and R wave in V5 and V6 > 35 mm
- Widened QRS complex
- RSR' complexes (rabbit ears):  -V1 and V2  
 -V5 and V6
- Peaked T waves in more than 1 lead, particularly from V1 to V6
- T wave inversion  -V4 to V6   
 -II, II, AVF   
 -I, AVL   
 -V1 to V6
- ST depression  -V4 to V6   
 -II, II AVF   
 -I, AVL   
 -V1, V2, V3   
 -V1 to V6
- ST elevation  -across multiple leads   
 -V1 to V4   
 -II, III, AVF