

離散Bayes識別則による標的マーカー探索と患者層別化 — 個別化医療の一例として治験への適用 —

山口大学 名誉教授 浜本 義彦

hamamoto@yamaguchi-u.ac.jp

治験の課題等技術の背景

①成功確率の低さ

探索的試験のPhase II でPOC (Proof of Concept) 成功確率は、30%以下である。

②主原因

従来のRCTでは、薬剤効果が患者によって異なることを考慮に入れていない。 薬剤効果がない患者が治療群へランダムに混入されてしまうために、十分な薬剤有効性が認められなかった。

③課題

薬剤効果の期待される患者を選別する患者層別化が成否の鍵である。そのため患者層別化実現のための「標的マーカー探索」と「多種多様な標的マーカーを用いた識別法」の確立が喫緊の課題となる。

④ AI技術への期待

AI技術は画像データには活用されているが、主要な臨床データが非画像データの治験ではAI技術の活用は進んでいない。臨床現場からは、治験を対象とした新しいAIアプローチが切望されている。

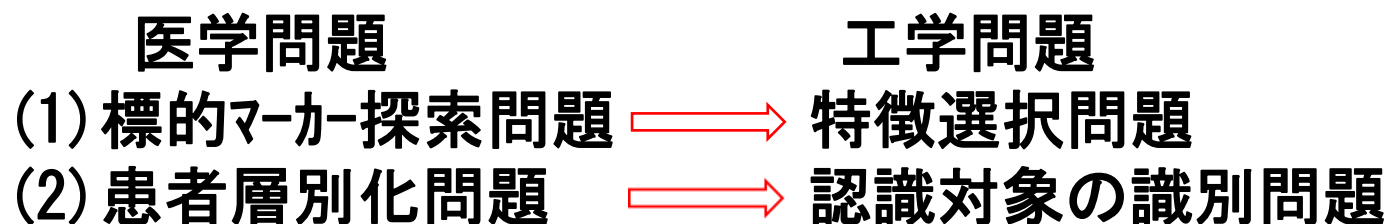
統計的パターン認識と標的マーカー探索、患者層別化との関係

統計的パターン認識の定義

- ①観測：認識対象（患者）から観測（マーカー）によってデータ（測定値）の獲得
- ②特徴選択：観測（マーカー）の中から認識に有用な特徴（標的マーカー）の組合せの選択
- ③識別：特徴（標的マーカー）の組を用いて認識対象（患者）の、それが属すべきクラス（レスポンドー、ノンレスポンドーのいずれか）への識別



解くべき問題
の変換



この二つの問題を解くには、実績のある統計的パターン認識に基づいて

- 1) **特徴選択技術**により標的マーカーの組合せを探索し、
- 2) 標的マーカーの組を用いた**識別技術**により患者をレスポンドーかノンレスポンドーかに識別することになる。しかし事はそう簡単ではなく、越えるべき大きな壁がある。

越えるべき大きな壁と解決策

センシング技術の発展により次世代センサーからの遺伝子変異等、各種ミックス情報のデータを獲得する技術は確立された。しかし、患者層別化（標的マーカー探索を含む）を目的とした、タイプの異なる膨大なデータを同じ枠組みで一括して処理できる解析法はない。

ブレイクスルーその1

①データの「こと」（記号）化

②「こと」を処理できる解析法 ← **独自の離散Bayes識別則と離散系最近傍解析**

数値データだけでなく、医学上重要な記号データ（遺伝子変異等）も混在する臨床データを一括解析できる。

ブレイクスルーその2

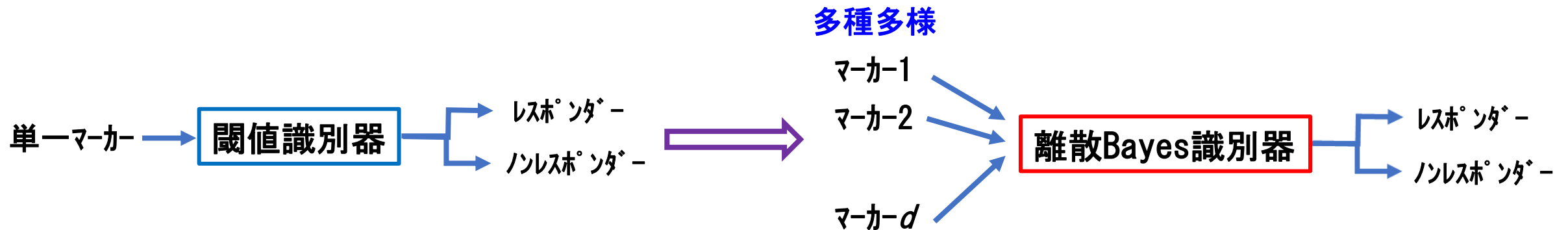
Lancet論文 [[doi.org/10.1016/S0140-6736\(03\)12775-4](https://doi.org/10.1016/S0140-6736(03)12775-4)]から進化した**独自の特徴選択技術**
インパクトファクター18.316(2003年当時) 被引用件数443回(Scopus)

がんは特にサンプル（症例）数が少なく、しかしデータ（次元）数は莫大で、正に「情報の爆発」である。

Small Sample Size and High Dimensionalityの状況下でも、サンプル変動にロバストな（汎化性の高い）特徴選択（標的マーカー探索）が可能。

離散Bayes識別則とそれに基づく標的マーカー探索

- ① マーカーの測定値であるデータを全て記号データとして「こと（医学的概念）」（言葉）化
（数値データはcut-off値により離散化、記号データは必要に応じて適度な粗さに再記号データ化）
- ② 記号データとしての「こと」を扱える独自の離散Bayes識別則に基づき、統計的パターン認識の特徴選択と識別の両方が行える解析法の確立
- ③ 数値データと記号データのようにタイプが異なり、多種多様で莫大な数のマーカーを候補とし、その中から有力な標的マーカーの組合せを、高速に、網羅的に探索する。
標的マーカー（遺伝子変異等）を個々に考えるのではなく、標的マーカーの組合せに着目する



従来の識別（単一マーカーでは不完全）

新技術の識別（標的マーカーの組合せ利用）

個々のマーカーが不完全であっても、それらを組合せて互いに補うことで不完全さを低減させ、識別の高精度化が図れる

新旧の比較

既存の統計的パターン認識
(数値データのみ対象)

① 識別

- ・ Bayes識別則
- ・ 最近傍識別則

② 特徴選択 (標的マーカー探索)

(1) 大規模な高速計算機が必要

(2) データは改正個人情報の対象



独自の統計的パターン認識 (記号データを扱える)

① 離散系識別 (患者層別化)

- ・ 離散Bayes識別則
- ・ 離散系最近傍識別則 (Webテスト, アンケートの記号データ解析も可)

② 離散系特徴選択 (標的マーカー探索)

(1) 大規模な高速計算機は不要

「次元の呪い」がない。データ数が増加しても
計算量の急激な増大はない。

低価格のデスクトップ型PCで解析可能 (院内で解析可能)

(2) データは改正個人情報の対象外

(統計情報処理による非個人情報化)

AI開発で必須である「データの第三者提供」に道が開かれる

「こと」化のメリット

AIの結果「こと (医学的概念)」の組合せ (知識の生成)

医師は、医学知識としてAIの結果を理解できる。

これまで単独で用いられていた遺伝子変異が組合せで評価
され、新しい可能性が期待される。

個人情報保護法におけるデータの第三者提供（AI開発に必須）

本人の同意なしで目的外利用の第三者提供は可能か

データ利用：提供元（病院等）では当初の目的は患者本人の診療で、提供先（企業）のAI開発は目的外利用

個人情報保護法（平成27年改正）

対策1：匿名加工情報

背景：医療情報は要配慮個人情報である。このためアウトは不可。このままでは目的外利用の第三者提供はできない。

特徴：特定の個人が識別できない情報に加工する（非個人情報化）

欠点：提供元がどこまで加工しなければならぬのかは不明。もし不完全であれば提供元に責任が残る。更に法のもとで定まった手続きが複雑である。非現実的

次世代医療基盤法（平成29年） （医療に特化した、個人情報保護法の特別法）

対策2：匿名加工医療情報

背景：病院側に責任がなく、企業からは対策1よりも手続きがもっと簡便な加工法が求められた。

特徴：大臣認定事業者が全ての加工と責任を引き受ける（非個人情報化）

欠点：加工について口を出せない。そのため、AI開発においてデータが使い物にならないリスクがある。

個人情報保護法（令和2年改正） （令和4年4月施行）

対策3：仮名加工情報

背景：企業からは、対策1より使い易く、対策2より自由度のある加工法が求められた。

特徴：特定の個人が、単独の情報では識別できないが、他の情報と照合すれば識別できる加工。敢えて不完全さ（完全性で責任が問われないという解釈）を利点とする。

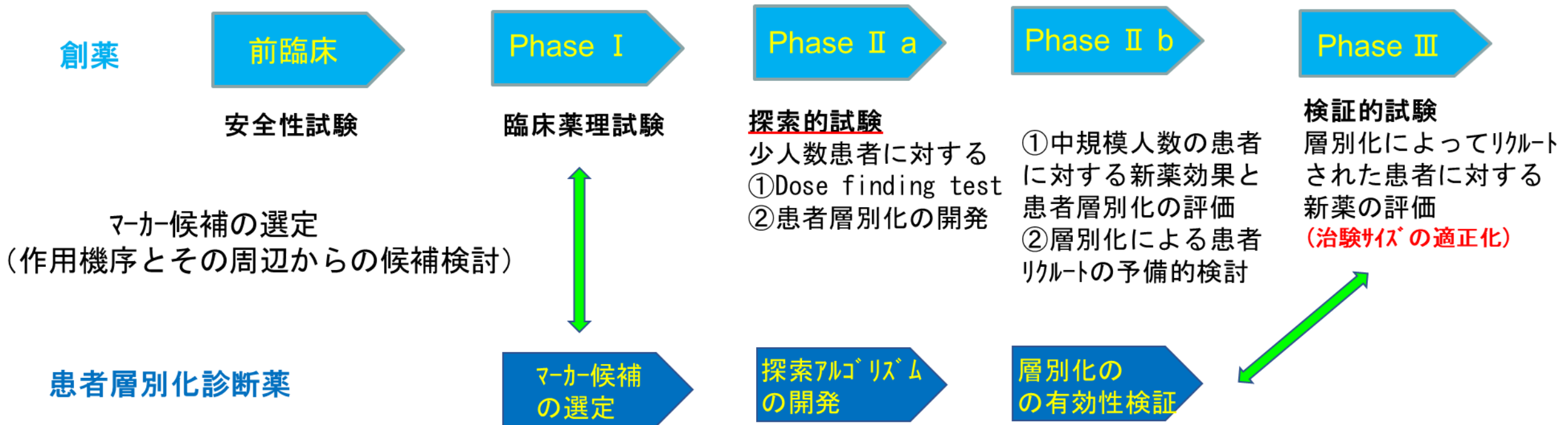
欠点：対策1, 2と異なり、第三者提供はできない。但し、病院からの委託、又は病院との共同利用という病院の管理下であれば第三者提供、AI開発が可能（欠点というより制約）

個人情報の利活用促進（規制緩和）の流れ

離散Bayes識別則のアプローチは、既存の数値を対象とした解析法とはまったく異なり、対策1, 2, 3は不要

構想

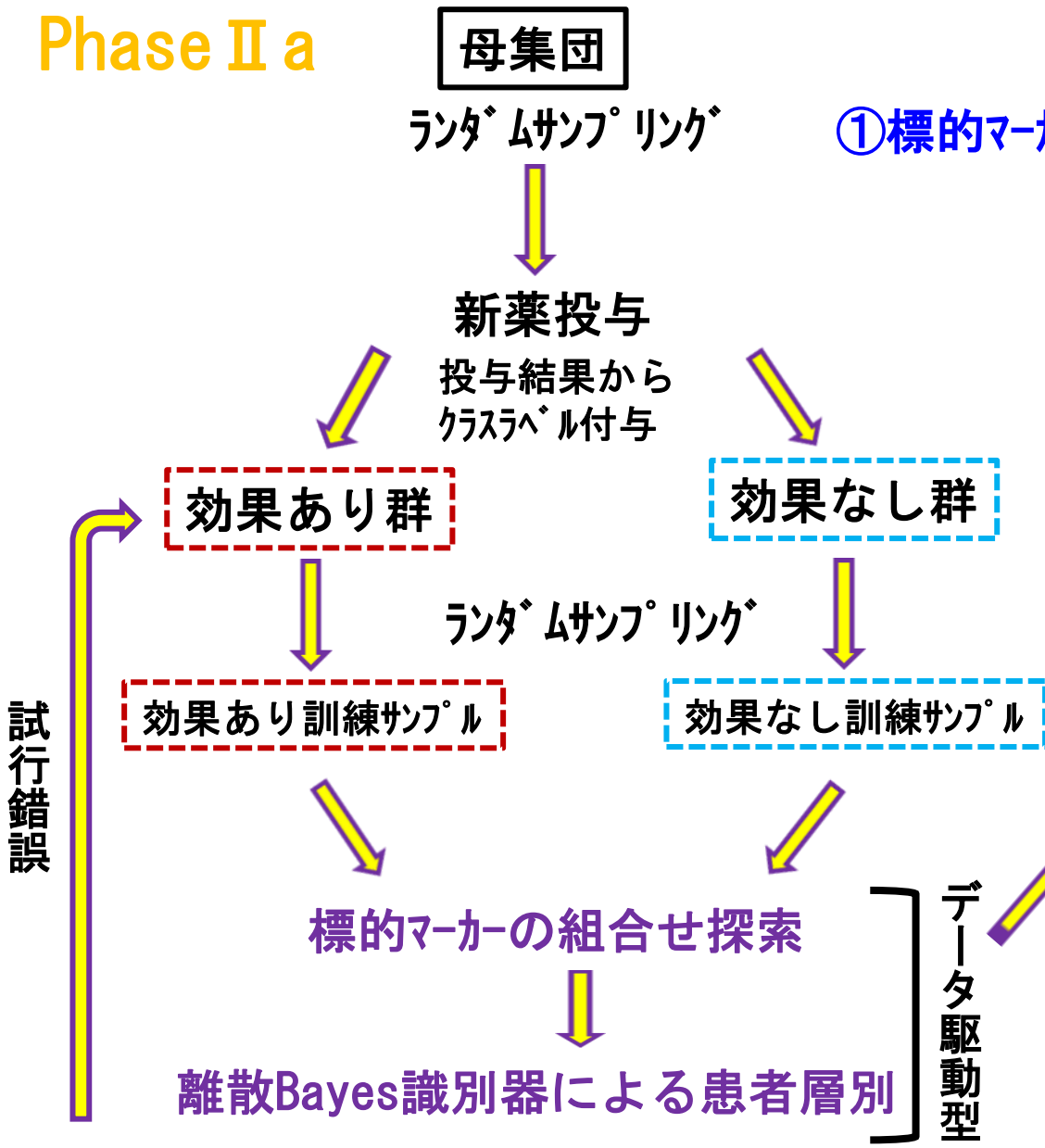
創薬と患者層別化診断薬の同時開発



一番難しく、解析の専門家が手薄の探索的試験 (Phase II) に注力する。

効果：**治験サイズの適正化による①成功確率の向上、②開発期間・開発費の縮小**

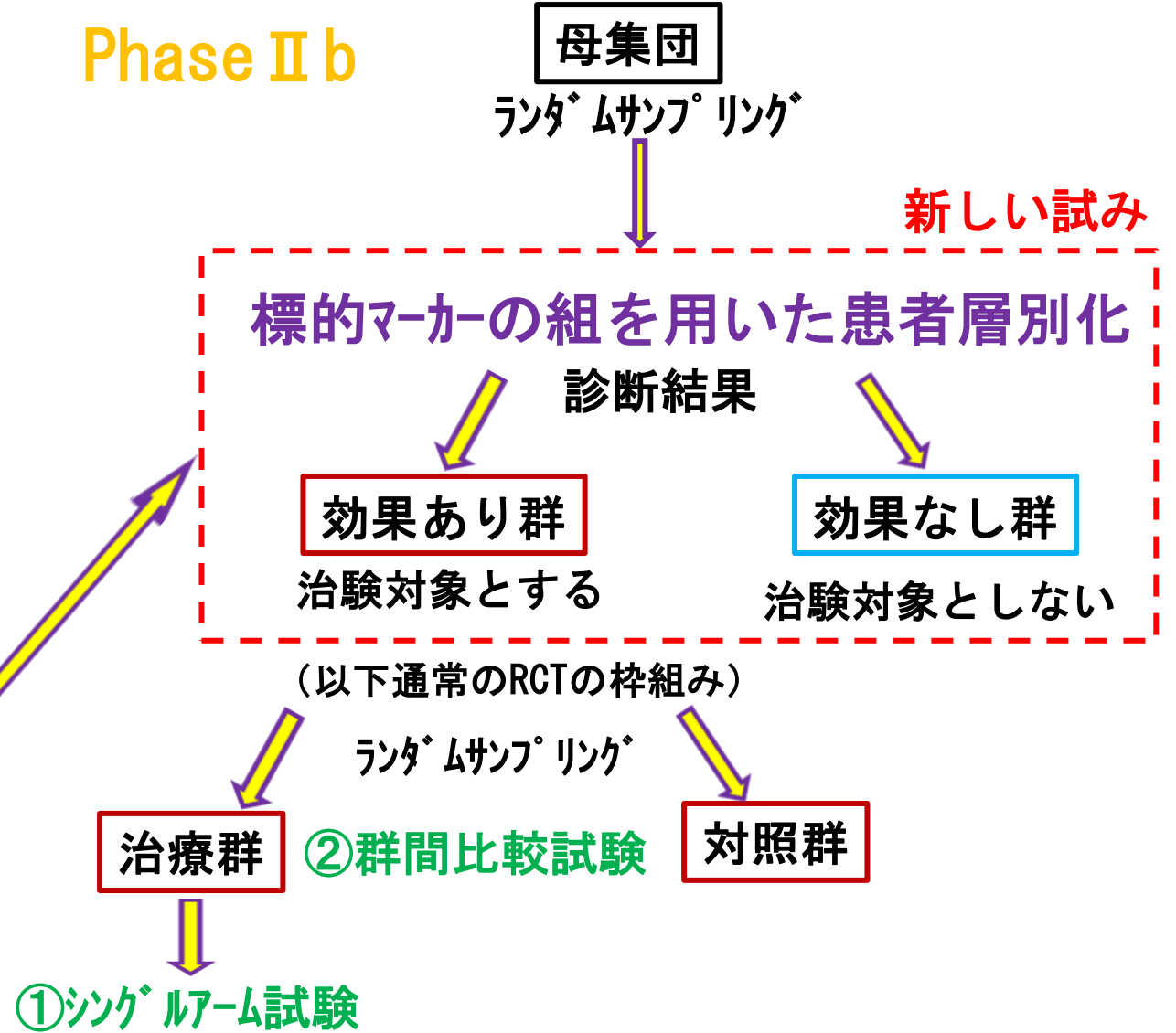
Phase II a



患者層別化開発の流れ

①標的マーカーの探索 ②標的マーカーを用いた離散Bayes識別器の設計

Phase II b



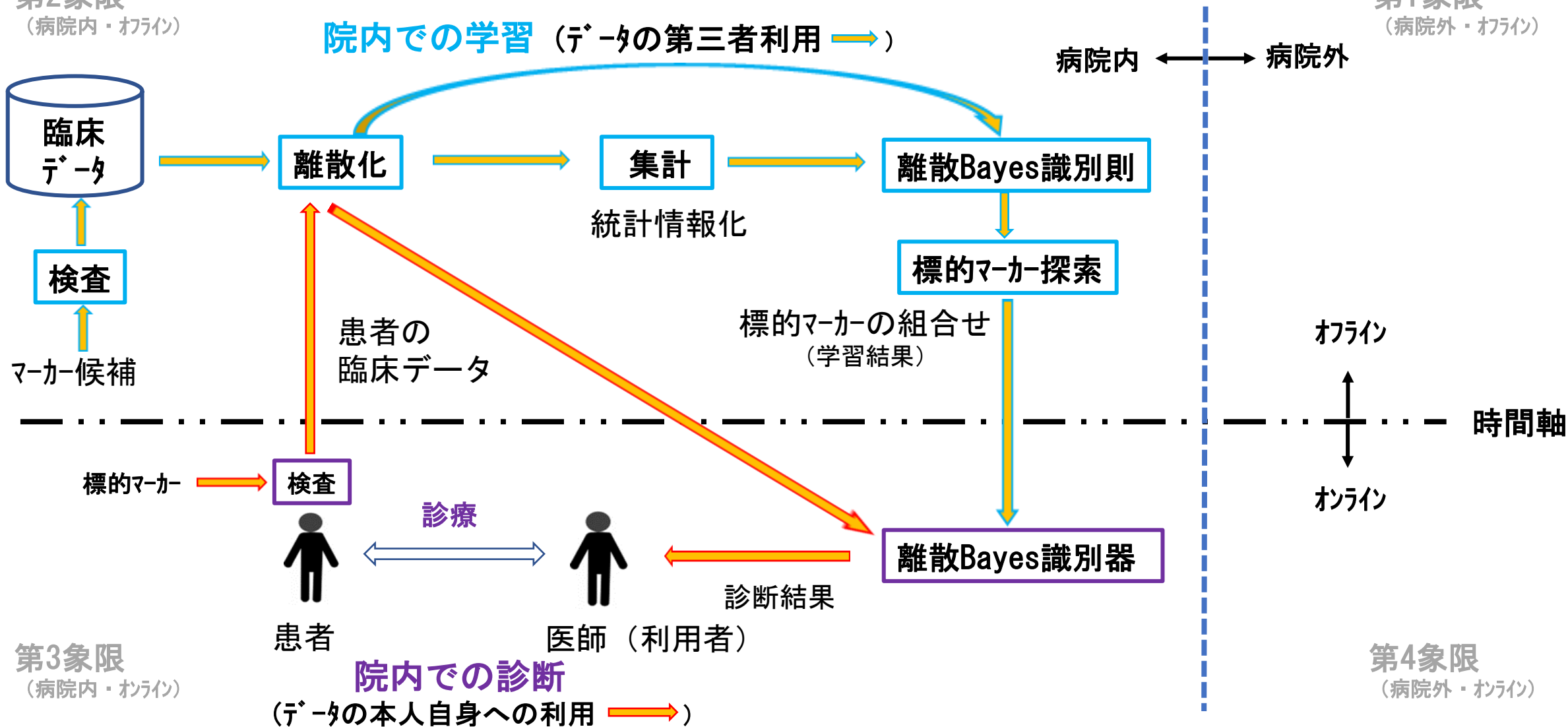
標的マーカー（組合せ）の臨床的意義の検証

患者層別化の有効性の検証

個別化医療実現のイメージ図

第2象限
(病院内・オフライン)

第1象限
(病院外・オフライン)



患者層別化診断薬の応用例：リキッドバイオプシーによる投与前の免疫チェックポイント阻害剤の効果予測

個別化医療における標的マーカー探索と患者層別化に特化した 離散Bayes識別則の研究結果1: 肝がんの早期再発の予測

背景

肝がんの怖さは高い再発率。完全切除しても術後1年以内に約30%が再発する。再発を予測できれば効果的な先制医療を実施でき、もし無再発ならば副作用のある抗がん剤やCT検査が不要となり、医療費の抑制にもなる。

医学問題

手術で肝がんを完全切除した患者を対象に、容易に入手できる保険適用の臨床データを用いて術後1年以内の早期再発を予測せよ。

工学問題

医師が選定した11マーカー候補の中から離散Bayes識別則により標的マーカーを選択し、早期再発を予測せよ。

結果

選択された5標的マーカーの組合せは、腫瘍数、腫瘍サイズ、ICG（数値データ）、vp、Liver damage（記号データ）の混在であり、テストサンプルに対して感度86%（癌再発の検出率）、特異度49%を得た。良く知られているTNM分類やModified JISよりも高感度で、ROC解析でも有効性を示した。

Prediction of early recurrence of liver cancer by a novel discrete Bayes decision rule for personalized medicine, *BioMed Research International*, 2016. doi.org/10.1155/2016/8567479

離散Bayes識別則の研究結果2: 早期胃がんのリンパ節転移の診断

背景

早期胃がんは内視鏡的治療で対応できるが、リンパ節転移が疑われると外科手術となる。しかし実際にはリンパ節転移のない患者にも不要な外科手術が行われている。

医学問題

早期胃がんに対する内視鏡的治療後に、容易に入手できる保険適用の臨床データを用いて、リンパ節転移を診断せよ。

工学問題

医師が選定した8マーカー候補の中から離散Bayes識別則により標的マーカーを選択し、転移を診断せよ。

結果

選択された3標的マーカーの組合せは、深達度、リンパ管侵襲、静脈侵襲（全て記号データ）であり、テストサンプルに対して感度100%（転移の見逃しゼロ）、特異度86%を得た。

Lymph node metastasis can be determined by just tumor depth and lymphovascular invasion in early gastric cancer patients after endoscopic submucosal dissection,
European Journal of Gastroenterology and Hepatology, 2017. doi: 10.1097/MEG.0000000000000987

離散Bayes識別則の研究結果3： 進行大腸がんの再発予測

背景

治癒的切除後の進行がんに対して予後良好の予測は困難で、有効な標的マーカーが求められている。

結果

10マーカー候補の中から離散Bayes識別則により3標的マーカーを選択し、テストサンプルに対して感度71%, 特異度67%の精度で進行大腸がんの再発を予測できた。

CD4 and FOXP3 as predictive markers for the recurrence of T3/T4a stage II colorectal cancer: applying a novel discrete Bayes decision rule, *BMC Cancer*, 2022. doi.org/10.1186/s12885-022-10181-7

離散Bayes識別則の研究結果4： 抗うつ薬の治療反応性の予測

背景

測定系の影響を受けやすい遺伝子情報を用いて抗うつ薬治療効果を予測することは困難であるが、臨床応用上重要であるため、その確立が求められている。

結果

広島大の訓練サンプルを用いて8マーカー候補の中から離散Bayes識別則により3標的マーカーを選択し、他施設である山口大及び徳島大のテストサンプルに対して感度91%, 特異度75%の精度で治療効果を予測できた。

Interferon signaling and hypercytokinemia-related gene expression in the blood of antidepressant non-responders, *Heliyon*, 2023. doi: Heliyon, 9(1), art. no. e13059

離散Bayes識別則の研究成果5: 術後合併症の予測 (論文投稿中)

背景

術後合併症の予測は、術後管理が大変で医療従事者を疲弊させないために、また患者にとっても重要である。

結果

患者を2群に層別化し、山口大の訓練サンプルを用いて群毎に13マーカー候補の中から離散Bayes識別則により標的マーカーを選択し、他施設である大阪大のテストサンプルに対して2群全体として感度86%, 特異度71%の精度で合併症を予測できた。

A novel prediction model of pancreatic fistula after pancreaticoduodenectomy using only preoperative markers, to be submitted

AIによる診療の支援とは

立場： あくまで主役は医師、コンピュータ(AI)は支援のみ

診療におけるAIの活用とは、コンピュータが医師の気づかない点を指摘しヒントを与え、医師の臨床推論力や発想力を高めることである。

盲点

サンプル(症例)の収集やサンプルからのデータ獲得については、機械学習は何も言えない。同様に、多変量解析も何も答えない。単に与えられたデータを処理するだけである。

そもそもデータに必要な情報が含まれているか否かは、
解析前は不明(やってみないと分からない)。しかも何ら理論的道すべもない。

実際はサンプルの収集、データの獲得が一番重要である。にもかかわらず、軽視されている。どの教科書にも書かれていない。結局、経験からでしか学べない。

医学問題では、医師のみが医学知識をもって議論できる。

医師は、コンピュータ(AI)の結果を決して鵜呑みにせず、医学的に解釈すること

参考文献

1. 渡辺 慧、認識とパタン、岩波新書、1978
2. 渡辺 慧、知ること、認知科学選書、東京大学出版会、1986.
(こちらの方が入手し易い ちくま学芸文庫 ISBN:978-4-480-09381-3)
3. C. R. Rao (藤越、柳井、田栗共訳)、統計学とは何か、丸善、1993.
4. 奥野、久米、芳賀、吉澤 共著、多変量解析法、日科技連、1971.
5. 浜本義彦、統計的パターン認識入門、森北出版、2009.

ご清聴、ありがとうございました。

本日の講演原稿のPDFファイルをご希望の方には送付致します。
お問合せは下記までお願い致します。

hamamoto@yamaguchi-u.ac.jp