

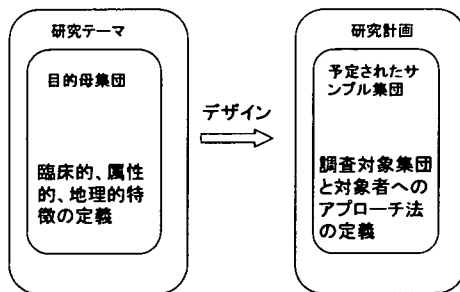
研究結果の一般化

- Framingham Study
 - 3分の1の予定対象者が研究への参加を拒否
 - ボランティアで補う
 - 参加拒否者には健康に問題のある人が多く
 - ボランティアには健康な人が多い傾向
- 予定されたサンプル集団を正確に代表していない可能性
 - 「高血圧が冠動脈疾患のリスクファクターである」という結論に影響を与えるほど大きなものではない
 - そのような誤差のないほかの研究でも同様の知見
 - 「質的判断 (qualitative judgment)」

研究結果の一般化

- 久山町研究 1961年～
 - 脳梗塞・虚血性心疾患などを研究
 - 前向きホコート研究
 - 調査開始後1000人の自然増のみ
 - 年齢構成・職業構成は全国平均
 - ・ 偏りの少ない平均的な日本人
 - 40歳以上の住民(約7000人)
 - 受診率80%以上
 - 追跡率99%以上
 - 剖検率80%以上

研究対象者を決定するまでのプロセス



選択基準

- 取り込み基準
- 除外基準
- サンプル集団
 - 臨床サンプル
 - 一般集団サンプル

取り込み基準

- 目的母集団の主な特性の定義
 - 「一般健康状態が良好な者」をどう定義するか?
 - 定義はできるだけ明確でなければならない
- 一般化可能性と研究の効率
 - 各種属性(年齢・性別・人種)
 - 臨床的特性
 - 地理的条件
 - 時間的条件

除外基準

- 取り込み基準を満たす人の中で・・・
 - データの質を低下させる、結果の解釈を複雑にする、倫理的に問題のある人を除外
- むやみに除外基準を設けるのは避ける
- 例として・・・
 - 骨粗鬆症の研究にアルコール常飲者を含めるか?
 - 含める場合⇒アルコールの過剰摂取と骨の脱灰化との関係を検討できる。反面フォローアップが困難。
 - 含めない場合⇒データの信頼性は増すが、アルコール常飲者とはどういふ状態かを明確に定義する必要がある。
- 研究結果を一般化する場合の妥当性 (Validity) を検討する必要がある

サンプルサイズ

- くわしくは第5・6章で・・
- サンプルサイズは多ければ多いほどよい？
 - 小児科医の当直回数と忙しさについて調べる
 - 調べる病院は？
 - 専門病院・大学病院・基幹病院・市中総合病院
 - 当直の形態
 - 医師数

サンプル集団

- 臨床サンプル
 - 実際の患者さんを対象とする
 - 利点・集めるのが容易でコストが少なくてすむ
 - 欠点・医療施設により選択バイアスがかかる
- 一般集団サンプル
 - コミュニティーを代表するサンプル
 - 利点・地域や社会集団の公衆衛生にかかわる重要な情報が得られる。
 - 欠点・集めるのが困難でコストがかかる

サンプリング

- 簡易サンプリング
 - 簡易サンプリング
 - 取り込み基準を満たしアクセスが容易な人を選ぶ
 - 連続サンプリング
 - すべての人を連続的にサンプリングする
- 確率的サンプリング
 - すべての人がランダム(無作為)に、等しい確立で対象者に選ばれる方法
 - 単純ランダムサンプリング
 - 層化ランダムサンプリング
 - クラスタサンプリング
 - 系統的サンプリング

確率的サンプリング

- 単純ランダムサンプリング
 - 目的母集団の全員からランダムにサンプリングする
 - 乱数表(やコンピューター)を用いて行う。
- 層化ランダムサンプリング
 - 目的母集団を性別、人種などの特性にもとづく層に分け、それぞれの層からランダムにサンプリングする。

確率的サンプリング

- クラスタサンプリング
 - 母集団の中の部分集団(クラスタ)をひとつのサンプル単位としてクラスタをランダムにサンプリングする方法
- 2段階クラスタサンプリング
 - クラスタサンプリングを行った集団の中でさらにランダムにサンプリングする方法
- データ解析が複雑になる

確率的サンプリング

- 系統的サンプリング
 - 母集団全員に番号をつけ、あらかじめ決められた周期性をもってサンプリングする方法
 - 臨床研究にはふさわしくないサンプリング方法

サンプリングデザインのまとめ

- 臨床研究で確率的サンプルを用いることができることはきわめてまれ
- ほとんどの臨床研究では、簡易サンプルで行われることが多く、連続サンプルが現実的にもっとも適したサンプル方法

研究対象者のリクルート

- リクルートの目標
 - 代表性のあるサンプルの獲得
 - 必要な数の対象者を獲得する
- 現実的に可能か？

代表性のあるサンプルの獲得

- 目的母集団やサンプリング法を適切に選ぶ
- 取り入れ基準(除外基準)を明確に設定
 - 研究者ごとで誤差が出ないように
- その基準が研究の全プロセスで遵守されているかモニタリングが必要
- 記述研究の場合は参加率が代表性を左右
 - 参加率をあげる努力

代表性のあるサンプルの獲得

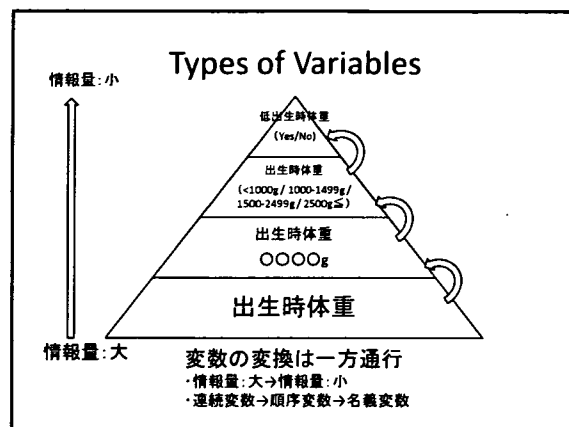
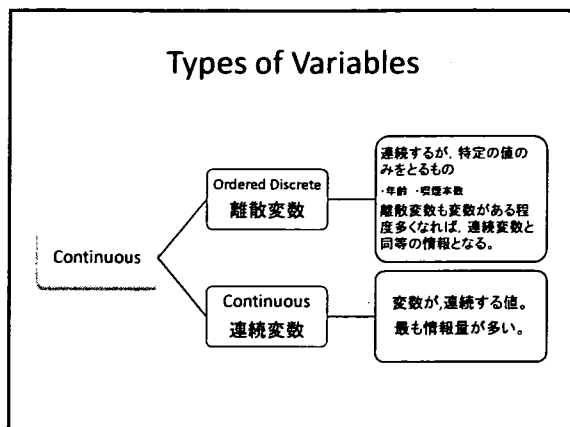
- 参加率をあげるためには……
 - 何度も接触をこころみる
 - 複数の接触方法(電話・手紙・メール・訪問)
 - 侵襲性や不快を伴うような研究デザインを避ける
 - 不安を和らげるためにパンフレットを用意したり直接説明する
 - 交通費を負担する
 - 検査結果を伝えるなどの特典

必要な数の対象者を獲得する

- 予定された対象者が確保できないことはよくある
 - 予備調査を行う
 - 不参加率を見込んで対象者を多めに設定
 - 不足が明らかになった時点で適切な対応
 - 研究期間の延長など
 - リクルートの進み具合を綿密にモニターする
 - 不参加となった対象者の割合と理由を把握しておく
 - 対象者を増やす工夫
 - 関係者・機関の協力、健康診断・郵便・テレビ・新聞・ホームページなど

まとめ

- 母集団を代表するサンプル
- サンプリングの利点は研究効率を高めるが一般化可能性に留意
- 目的母集団の概念を明確にし、取り込み基準を設定、除外基準は最小限に
- 連続サンプリングが最も適切な方法であることが多いが、サンプル数が多すぎるときは確率的サンプリングを組み合わせる
- 研究に必要なサンプル数が確保できるようなリクルート法をデザインする



どの尺度を使って測定するか

【原則】 連続変数が望ましい

- 情報量が多い
- 検出力が高く、サンプルサイズを少なくできる。

- 研究者の関心により、順序尺度を選択する
 - 例) 個々の体重より、発達に影響を与える低出生体重児(BW<2500g)の割合に関心がある場合
- その場合も、連続データ(生データ)を残しておく。
- 順序尺度を選択する場合、区切りを細かくしておく。

測定法の要件

本当に知りたいことが、正確に測定されること

- Precision 精度
- Accuracy 正確度
- Validity 妥当性

Precision 精度

- Degree of refinement in measurement

Minimum Random Error
偶然誤差が小さい!

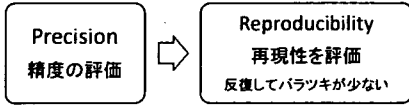
Good Reproducibility
再現性が高い
測定を反復してもばらつかない

Variability
ばらつき

Precision 精度に影響を与えるもの

- Random Error 偶然誤差
 - まったく偶然によって生じる誤差
 - 両方向に均等にばらつく
- Variability
 - Observer variability 測定者によるバラツキ
 - Subject variability 対象者によるバラツキ
 - Instrument variability 測定方法によるバラツキ

Precision 精度を評価する

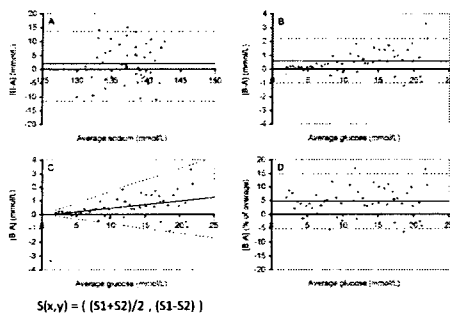


- **Observer Variability** 評価者によるばらつき
 - 同一評価者が反復測定したときの再現性
 - 異なる評価者が測定したときの再現性
- **Instrument Variability** 測定方法によるばらつき
 - 同じ測定方法で反復測定したときの再現性
 - 異なる測定方法で測定したときの再現性

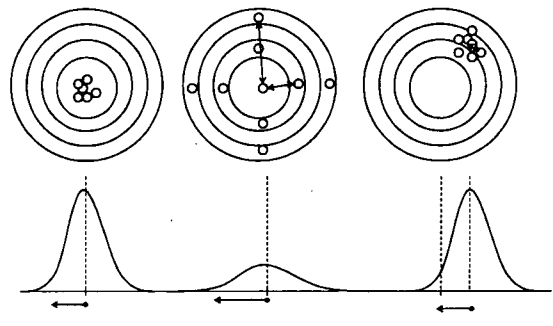
Precision 精度を評価する

- 再現性の指標
- 連続変数
 - 標準偏差 (Standard Deviation)
 - > 平均値とSDに相関がある場合は、相関係数ではなく、変動係数 (SD/mean) を用いて評価する。
- カテゴリー変数
 - 一致率, κ係数が用いられる

Bland-Altman Plot



Precision 精度を評価する



Precision 精度を高める工夫

1. 測定方法の標準化
2. 測定者の訓練
3. 測定方法の改善
4. 測定方法の自動化
5. 測定の反復

Strategy to Improve Precision	Source of Variation	Example of Variation	Example of Strategy to Reduce Variation
1. Standardize the measurement method	Observer	Observer A uses a pocket square to measure the width of a table.	Specify the method of measurement to all observers.
2. Train the observer	Observer	Observer B uses a pocket square to measure the width of a table.	Specify the method of measurement to all observers.
3. Improve the measurement method	Observer	Observer C uses a pocket square to measure the width of a table.	Specify the method of measurement to all observers.
4. Automate the measurement	Observer	Observer D uses a pocket square to measure the width of a table.	Specify the method of measurement to all observers.
5. Repeat the measurement	Observer	Observer E uses a pocket square to measure the width of a table.	Specify the method of measurement to all observers.

Accuracy 正確性

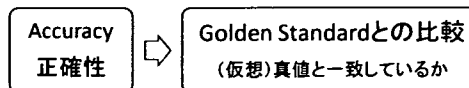
- Confirming exactly with the truth or a standard (Pocket Oxford)
- 測定したいと考えている真の値をどれだけ忠実に表現できているか。

- What is the Truth?
- Golden standard

Accuracy 正確性に影響を与えるもの

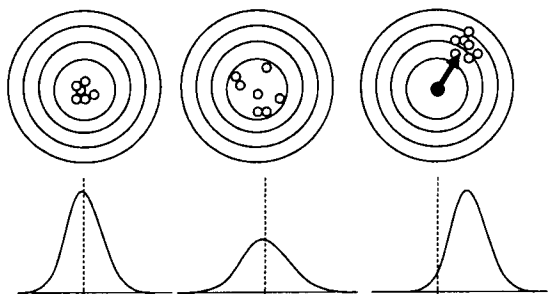
- System Error 系統誤差
 - 系統的におこる誤り
- Bias
 - Observer bias
 - Subject bias
 - Instrument bias

Accuracy 正確性を評価する



- 連続変数
 - (測定値 - 真値) を評価する
- カテゴリー変数
 - 真値に対する、感度と特異度を測定

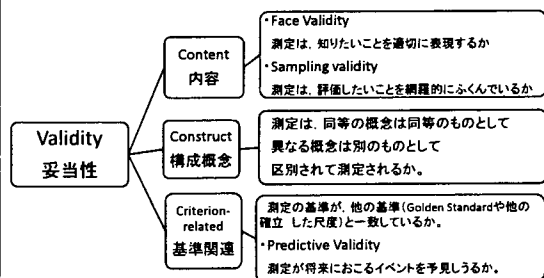
Accuracy 正確性



Accuracy 正確性を高める工夫

1. 測定方法の標準化
2. 測定者の訓練
3. 測定方法の改善
4. 測定の自動化
5. 対象者に気づかれぬ方法での測定
6. 盲検化
7. 測定方法のキャリブレーション

Validity: 測定の妥当性



測定方法の要件: その他

- 感度 (sensitivity)
 - 特異度 (specificity)
 - データの適切な分布 (Distribution)
 - 客観性 (objectivity)
 - 客観的な指標を絞りこむことは大切。
 - しかし、視野が狭くなりすぎて、得られる情報が限定されてしまう。あるいは予想外の知見を見逃してしまう可能性に注意する。
- ひとつの妥協策
- 主観的で質的な評価も付け加えておく。

まとめ

- 測定
知りたいこと(=変数)を数値化すること。
- 測定の要件
 - 精度 Precision
 - 正確性 Accuracy
 - 妥当性 Validity

まとめ Precision & Accuracy

	Precision 精度	Accuracy 正確性
定義	与 Reproducibility 反復して測定した測定値の バラツキの少なさ	測定値が、目的とする真の値 と一致する度合
最良の評価法	反復測定値と比較	Golden standardと比較
研究にとっての 意義	効果判定の検出力↑ サンプルサイズ↓	結果の妥当性↑
影響する因子	Random Error 偶然誤差	Systematic Error 系統誤差

2008年1月22日

診断アルゴリズムを核とした 小児急性疾患の初期対応における コンピュータ重症度判定の試み

国立成育医療センター
臨床研究フェロー 矢作尚久

目次

1. 診断アルゴリズムの概要
2. データベースの構築
3. システムの精度の検証
4. 今後の展開

医師の一般的な思考過程

1. 症状から鑑別疾患を予測し、問診
2. 鑑別疾患を絞り込むために診察
3. 絞り込まれた鑑別疾患を確定するために検査

診断アルゴリズムの概要

診断システムの開発の初期手順

1. 全ての症状一覧を作成
2. 全ての診断名一覧を作成
3. 1, 2各々で想定される問診内容を作成
4. 3の全てにScoreを付与

※ Scoring 重み付けとしての値を付与。
(症状がない=0、致命的な内容=4と設定。
初めは起点を1とし、重み付けとしての1-3を付与。)

データベースとは？

・ゴミ箱にデータを投げ入れるように蓄積するもの？

システムの精度の検証

背景

小児救急医療体制の充実の必要性はすでに数年来の社会問題となっているが、抜本的な解決策は見出されていない。

主たる原因は、養育者や医療提供者の「症状や変化への対応がわからない」という経験と知識の未熟性にある。



重症度を判定できない

重症度判定

小児における症状発生後の重症度判定は、養育者をはじめ、診療に携わる研修医、看護師、医師各々の知識と経験により大きく左右される。

経験豊富な小児科医は患者の主訴と的確な問診から診察前のある程度の重症度を予測する。



1. 重症度判定システムの開発
2. システムの精度を検証

システム精度の検証

仮説:システム判定の精度が高くなれば、システムは経験豊富な医師の判断に近づく。



検証:

- ①外来診療において患者の実際の重症度を入力。
- ②実際の判定とシステム判定の差を分析。

方法

場所	対象	作業内容
待合室	養育者	1.生年月日、体重等を入力 2.全身状態の質問に回答 3.主要3症状の質問に回答
診察室	小児科医	診察開始 (予診内容を確認) 診察終了 4.実際の重症度を入力



step 1

Step 1
Q 必要毎項目、生年月日とを教えてください。

診療者番号

生年月日 年 月 日
(年-月-日の順に入力してください)

性別

兄弟の数(本人含む)

体温

その発熱は から

体温

過去の受診

再診(1か月以上かかるとは)

前から見たお子様の状態

来院のきっかけ

Step 2

お子様は生後1年6ヶ月です。

Step 2
Q 全体的状態について教えてください

■ 外傷についての質問
今回気になる症状に関連して、ぶつけたり、落ちたり、打つようなことがありましたか?
① はい
② わからない
③ ありません

■ 嘔吐状態についての質問
いつもと比べて嘔吐状態はどうですか?
① ぐっぐりしている、反りが強い
② いつもと様子が変わり、嘔一つしている
③ いつもと変わらない

■ 呼吸状態についての質問
いつもと比べて呼吸・顔色はどうですか?
① 顔色が悪い、唇が青い、鼻下黒い、手足が冷たいか、おぼろげに青い

step 3

Step 3
Q 気になる症状を最大3つまで選んでください。

<input type="checkbox"/> 顔色が悪い	<input type="checkbox"/> 目が眩に大きい
<input type="checkbox"/> 顔のどこかが痛い	<input type="checkbox"/> 耳が痛い
<input type="checkbox"/> 鼻が痛い	<input type="checkbox"/> 喉が痛い
<input type="checkbox"/> 喉が痛い	<input type="checkbox"/> 喉が痛く、目の腫れ
<input type="checkbox"/> 背中が痛い	<input type="checkbox"/> 喉が痛く
<input type="checkbox"/> 口の中の変化	<input type="checkbox"/> 喉が痛く
<input type="checkbox"/> 唇が腫れている	<input type="checkbox"/> 飲み込みにくい
<input type="checkbox"/> 吐いた	<input type="checkbox"/> 解熱剤が出る・排便時に痛がる
<input type="checkbox"/> 顔を吐いた	<input type="checkbox"/> 発熱
<input type="checkbox"/> おなかが痛い	<input type="checkbox"/> 白っぽい便が出た
<input type="checkbox"/> おなかがはっている	<input type="checkbox"/> 下痢
<input type="checkbox"/> タマが腫れている	<input type="checkbox"/> 吐くみ・じんましん
<input type="checkbox"/> 血便が出た	<input type="checkbox"/> 鼻出血
	<input type="checkbox"/> あざ(出血斑)

step 4

Step 4
Q 気になる症状について教えてください。

嘔吐
吐れますか? ① 吐れない
② 嘔ても時々目を覚ます程度
③ 嘔れている様子

呼吸
呼吸の状態はどうですか? ① 平/平な状態で息している
② わからない
③ いつもと変わらない


症状の経過は? ① 今日から/徐々に悪化している
② 2-3日前から
③ 上乗れり乾から

鼻汁
鼻汁や鼻血はいつからですか? ① 3週間以上前から
② 数日前から
③ 今日から

鼻汁
鼻汁はつまっていますか? ① はい、鼻がつまって吐れない
② はい、ややつまっている
③ いいえ

鼻汁
鼻汁はどのような感じですか? ① 色がついて粘り
② 黄白色透明で水様/その他

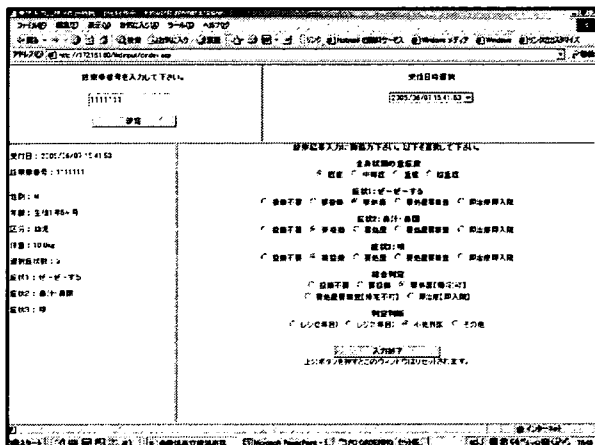
ご協力ありがとうございました。
待合室にてお待ちください。



入力終了して印刷へ戻る

初めて御利用の方はアンケートに御協力下さい

診察室で実際の重症度を入力



対象と基準

期間: 2004年1月1日~6月30日
 場所: 横浜市立市民病院小児科外来、救急外来
 対象: 受診した全患者8145人

判定基準:
 重症 入院(精査・加療)が必要
 中等症 処置後帰宅可能
 軽症 無処置で帰宅可能

結果

重症度	重症	中等症	軽症
実際の判定	22例	238例	1107例
システム判定	247例	316例	804例

重症度	要処置群	無処置群
実際の判定	260例	1107例
システム判定	563例	804例

※ 有効回答数1367データを分析

精度

感度 100%
(実判定で要処置をシステム判定が要処置に判定した割合)

特異度 72.6%
(実判定で無処置をシステム判定が無処置に判定した割合)

陽性的中率 46.1%
(システム判定が要処置のうち実判定が要処置だった割合)

陰性的中率 100%
(システム判定が無処置のうち実判定が無処置だった割合)

考察

見落としのない安全域を確保した精度向上
+

より使いやすさを追求したシステム作り

↓

システムの重症度判定基準を改定し、より優位な結果を確認

対象と基準

期間: 2005年1月1日~12月31日
 場所: 横浜市立市民病院小児科外来、救急外来
 対象: 外来に受診した全患者(n=12600)

判定基準:
 重症 入院(精査・加療)が必要
 中等症 処置後帰宅可能
 軽症 無処置で帰宅可能

結果

重症度	重症	中等症	軽症
実際の判定	63例	161例	776例
システム判定	110例	192例	698例

重症度	要処置群	無処置群
実際の判定	224例	776例
システム判定	302例	698例

※有効回答n=9880から1000データをランダムに抽出

精度の比較

	2004年	2005年
感度	100%	100%
<small>(実判定で要処置をシステム判定が要処置に判定した割合)</small>		
特異度	72.6%	89.9%
<small>(実判定で無処置をシステム判定が無処置に判定した割合)</small>		
陽性的中率	46.1%	74.2%
<small>(システム判定が要処置のうち実判定が要処置だった割合)</small>		
陰性的中率	100%	100%
<small>(システム判定が無処置のうち実判定が無処置だった割合)</small>		

考察

- ・安全性 : 重症、中等症の見落としはなかった。
- ・精度 : 陽性的中率、特異度共に向上した。



本システムは、小児医療に経験の少ない医療従事者(小児科以外の医師、研修医、看護師、救急隊等)のトリアージ支援や安全支援に役立つ。

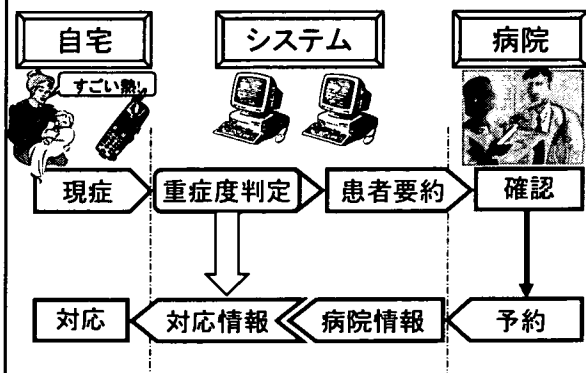
将来的には、養育者の救急外来受診の判断支援に活用できる。

Take Home Message

重症度判定システムは小児の救急医療環境を改善する可能性がある。

つづく

今後の展開



今後の展開

- ・ 医療資源の適正化
病院機能評価、救急車の最適配置
- ・ 感染拡大予測
- ・ 研修効果判定(シミュレーション)
- ・ 医療従事者の診療支援
などへの応用

2008年1月29日 臨床研究センター勉強会
 医学研究のデザイン 第5章
 サンプルサイズを見積もるための準備

国立成育医療センター
 内分泌代謝科 臨床研究フェロー
 磯島 豪

はじめに
 医学研究における統計手法の利用
 (2006年7月13日 臨床試験方法論 第7回統計解析の基礎 松山裕資料より)

- 統計手法の種類: 探索的か検証的か
- 探索的 (exploratory)
 - データは何を語るか? 分布の形, 外れ値
 - グラフ手法の応用
 - 仮説の創設
- 検証的 (confirmatory)
 - 仮説の検証を目的
 - 統計的推測: 検定・推定
 - 解析計画

⇒今回と次回(前川・斉藤先生担当)は, 検証的な手法についての勉強会です。

臨床試験・比較試験を例に取れば
 (2006年7月13日 臨床試験方法論 第7回統計解析の基礎 松山裕資料より)

- データのクリーニング
- 集計、分布状態の記述
- 単純な群間比較
- 交絡要因の補正・調整
- 回帰モデルの当てはめ

・種々の段階で, 種々の統計手法が活躍する
 ・今回は, 仮説検定の基本事項について概説します

本日の内容

- 仮説
 - 帰無仮説・対立仮説
- 仮説検定の統計学的基礎とその他の基本事項
 - α エラー(偽陽性)と β エラー(偽陰性)
 - 有意水準と検出力(パワー)
 - P値
 - 効果量
 - データの変動度
 - 多重仮説検定と事後仮説

サンプルサイズの見積もりは研究に欠かせない

- サンプルサイズが小さいと, 結論を出しにくい
- サンプルサイズが大きいと, 実施が難しくなる
- 研究が動き出したら, 途中の変更は難しい

⇒ 研究対象・研究デザインが決まったら, 早い段階で行う

具体的な方法論は, 次回以降。
 今週は, 概念を理解してサンプルサイズ計算の理解を深められるようにすることが目標。

仮説をたてる

- 研究テーマは, 最終的にはある研究仮説という形にまとめられる
- 研究テーマへの答えを「有意差検定」という統計学的手続きに置き換える基礎を作る
- 記述的研究には必要ないが, 比較を伴う研究には必要である
- 仮説は, 「簡潔に」「的確に」「実施前に立てる」のが良い

研究をまとめたものが論文であり, 今週の内容については, 先週のランセットの論文を具体例に, 仮説と統計学的解析について説明してみる

Effect of maternal multiple micronutrient supplementation on fetal loss and infant death in Indonesia: a double-blind cluster-randomised trial

論文本文のスクリーンショット。右側には「仮説は、題名でわかることも多い」という注釈があり、左側には「仮説は、背景の最後の方に書いてあることが多い」という注釈がある。

SUMMIT Study Group. *Lancet*. 2008; 371:215-25

Thus, additional findings from studies of maternal MMN supplementation on infant outcomes such as fetal loss and mortality are needed. Moreover, findings obtained in the context of prenatal care delivery systems would provide greater relevance compared with those obtained with unsustainable methods. We therefore undertook the Supplementation with Multiple Micronutrients Intervention Trial (SUMMIT)—a double-blind cluster-randomised trial in Indonesia—to examine the effect of maternal supplementation with MMN, compared with IFA, on fetal loss and infant death in the setting of routine prenatal care.

論文から逆に仮説を1つの文章にしてみると

「日常診療において、母体への多種微量栄養素 (MMN) 投与は、鉄剤と葉酸 (IFA) 投与と比較して、胎児死亡や乳児死亡を減らす」

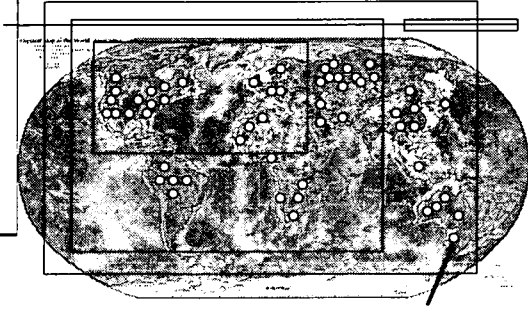
SUMMIT研究の仮説の証明はどのようにするか?

「ランダム化の手続きにより選ばれたサンプルを調べること、対象者全体に対して仮説が正しいかどうかを調べる」時、
対象者全体(母集団)での仮説が真ならば、サンプル内でも真であるから
サンプル内での結果
MMN群: 35.5死亡/1000出生(3.55%)
IFA群: 43死亡/1000出生(4.3%)
を比較して、3.55% < 4.3%とMMN群の方が低いので、
仮説は証明できた!!
としていいのだろうか?

ある医師の反論

「私たちの経験では、IFA投与では死亡者は0人ですが、MMN群では2人死亡者がいたのでMMN投与の方が死亡率を下げるとはとうい思えません」
⇒これもまた事実であり、必ずしも間違っているとは言えない
1つの反例を呈示することで、一般化した仮説は簡単に否定できる!!!
⇒一般化した仮説を、直接証明するのは非常に難しい
自分の都合のいいデータだけを出して喜んでいてもまわりの人は納得してくれない
仮説を反証し、それがどの程度受け入れられるかを考える

「白鳥は白い」は真実か?



Black swanの発見

仮説を示すための推論の第一歩 -研究仮説から検定仮説へ-

- 仮説A: 「対象者全体に対するMMN投与は、IFA投与よりも胎児死亡・乳児死亡は少ない」
が正しいとすると
- 予測B: 「サンプル内において、MMN投与群の方が、IFA投与群よりも、胎児死亡・乳児死亡は少ない」
は正しい

命題として簡単に示すと、A: 真 ⇒ B: 真

昔, 数学で習った内容の復習
<p>A: 真⇒B: 真が正しい時, 次のうち正しいのはどれ?</p> <p>1) B: 真⇒A: 真 (逆)</p> <p>2) A: 偽⇒B: 偽 (裏)</p> <p>3) B: 偽⇒A: 偽 (対偶)</p> <p>答えは, 3)の対偶</p> <p>「サンプル内での2つの死亡率が予測通り(B: 真)だったから, 対象者全体に対する仮説が正しい(A: 真)」と導く推論は, 「逆」の推論過程であり, 証明したい命題の真偽と必ずしも一致しない。</p> <p>死亡率だけを比較して, 結論を導けない理由</p>

そこで, 「命題と対偶の真偽が一致すること」を利用して, 帰無仮説を立てる
<p>予測B: 「サンプル内において, MMN投与群が, IFA投与群よりも, 胎児死亡・乳児死亡は少ない」が, 否定できれば</p> <p>仮説A: 「対象者全体に対するMMN投与は, IFA投与よりも胎児死亡・乳児死亡は少ない」は誤っていることを証明できる。</p> <p>でも, 証明したいのは, 仮説Aであり, 誤っていることを証明したいのは, 仮説Aを否定する仮説である。そこで, 帰無仮説Aを立てる。</p>

基本用語の説明 帰無仮説と対立仮説
<ul style="list-style-type: none"> ■ 帰無仮説: 検定で棄却するための仮説 「予測因子とアウトカムの間には関連がない」 ■ 対立仮説: 研究者が示したい仮説 「予測因子とアウトカムの間には関連がある」

仮説検定の手順
<p>帰無予測B: 「サンプル内において, 胎児死亡・乳児死亡は変わらない」が誤り</p> <p style="text-align: center;">↓</p> <p>帰無仮説A: 「対象者全体において, MMN投与とIFA投与では胎児死亡・乳児死亡は変わらない」が誤り</p> <p style="text-align: center;">↓</p> <p>仮説Aを受け入れる(対立仮説) ← 受け入れる確率を統計学的に評価する (P値を考慮する)</p> <p>受け入れた結果が偶然生じる確率がどの程度であるかを評価する。俗な言い方をすると, 「仮説をどの程度妥協して受け入れるか?」</p>

片側仮説と両側仮説
<ul style="list-style-type: none"> ■ 対立仮説にも2種類ある <ul style="list-style-type: none"> - 両側仮説 関連・効果の向きを限定しない 例: 「井戸水を飲む人と飲まない人では胃潰瘍の頻度が異なる」「MMN投与群とIFA投与群では, 乳児死亡率は違う」 - 片側仮説 関連・効果の向きを一方に限定 例: 「井戸水を飲む人は飲まない人よりも胃潰瘍が多い」「MMN投与群の方が, IFN投与群よりも乳児死亡率が低い」 ■ 統計学的厳密性という観点から片側か両側かを選ぶべきである。一般的には, よほどの根拠がない限り両側の事が多い(SUMMIT研究でも両側検定)

仮説検定のまとめ
<ul style="list-style-type: none"> ■ 仮説は, 母集団を対象とし, 実際の研究は, 母集団の中の一部のサンプルである。 ⇒命題と対偶の真偽が一致することを利用して, サンプルから母集団のことを示す ■ 一般に, 差の存在を, 直接示すことは難しい ⇒帰無仮説を立てて, 差が無いことを示す。その際に統計を利用して, 対立仮説を受け入れる確率を考える

仮説検定の統計学的基礎とその他の基本事項

- 帰無仮説と対立仮説
- 効果量・変動度
- α エラー(偽陽性)と β エラー(偽陰性)
- 有意水準と検出力(パワー)
- P値
- 多重仮説検定と事後仮説

効果量 (effect size)

- 研究者が示したい効果・関連の大きさ
- 効果量は、研究者が比較を行うために、変換する
- 研究者が決定する
 - 臨床的に意味のある値
 - 先行研究からの情報
 - 小規模なパイロット研究
- 効果量が小さいほど必要なサンプルサイズは大きくなる

効果量の指標

- 差,比,オッズ比,率比,・・・
 - 論文では大きく見える値が載っていることもある
 - 先ほどのSUMMIT研究を例に取ると,
MMN群: 35.5死亡/1000出生(3.55%), IFA群: 43死亡/1000出生(4.3%)
 - 絶対差 $3.55-4.3=-0.75\%$
 - 比 $3.55/4.3=0.82$
 - 相対差 $(3.55-4.3)/4.3=-0.18$ 18%の減少
- サンプルサイズの計算方法
 - 指標によって異なる
 - 結果はほぼ同じ
 - 仮説が複数ある時は一番大事なものに注目

Statistical analyses

We calculated the sample size on the basis of proportions,¹⁴ since estimates of death rates per person-time were not available, in an intention-to-treat analysis assuming an estimated neonatal mortality ratio of 46 per 1000 livebirths because this figure would accommodate higher ratios of early infant mortality to 12 weeks post partum. The design effect from clustered randomisation was estimated to be 1.20 for 300 clusters, and anticipated loss to follow-up was 10%, with a 15% allowance for abortions and stillbirths. With a power of 80% and a two-sided significance level of 5%, enrolment and follow-up of 36 100 pregnant women and infants would be needed to detect at least a 15% change in the neonatal or postneonatal mortality in the MMN group relative to the IFA group. Stopping rules drafted by the data and safety monitoring board established a minimum meaningful effect as 10%, thereby dictating a provisional sample size of 83100 euro/lees to assess effect of MMN on infant deaths.

推定の乳児死亡率:
46死亡/1000出生

少なくともMMN群が
IFA群よりも死亡率に
おいて、15%の変化を
検出する必要性

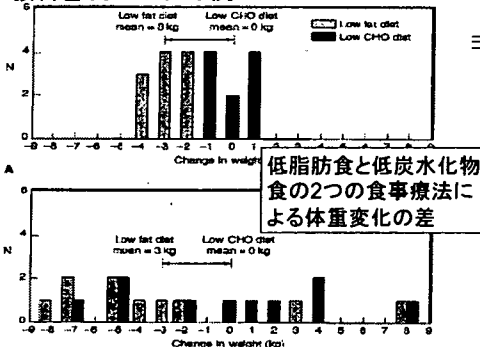
安全評価委員会は、試験中止基準は10%とした。

研究者が、設定した効果量⇒2つの治療の相対差が15%

データの変動度

- 効果のばらつき具合
 - ばらつきやすい指標=サンプルがたくさん必要
- 教科書(60ページ)の例
 - 同じ効果量(平均体重差5kg)でも、データのばらつきにより、結論できる場合とできない場合がある
- できるだけ安定した指標,安定した測定を選ぶ
 - SUMMIT研究では、「出産後90日以内の死亡」

教科書60ページの例



Statistical analyses
 We calculated the sample size on the basis of proportions,¹⁷ since estimates of death rates per person-time were not available, in an intention-to-treat analysis assuming an estimated neonatal mortality ratio of 46 per 1000 livebirths because this figure would accommodate higher ratios of early infant mortality to 12 weeks post partum. The design effect from clustered randomisation was estimated to be 1.20 for 300 clusters, and anticipated loss to follow-up was 10%, with a 15% allowance for abortions and stillbirths. With a power of 80% and a two-sided significance level of 5%, enrolment and follow-up of 36 500 pregnant women and infants would be needed to detect at least a 15% change in the neonatal or postneonatal mortality in the MMN group relative to the IFA group. Stopping rules drafted by the data and safety monitoring board established a minimum meaningful effect of 10%, thereby dictating a provisional sample size of 83 100 enrollees to assess effect of MMN on infant deaths.

死亡という測定は安定した指標と考えられるが、ランダム化後に全員把握できるわけではない

「10%は経過観察できない」、「15%は中絶または死産する」と予想

観察される効果量の差(1):
 例えばSUMMIT研究の2つの治療の死亡率の差

帰無仮説A:「MMN投与とIFA投与では胎児死亡・乳児死亡は変わらない」が正しければ、
 仮想的に同じ規模の研究を繰り返した時に、
 「2つの治療による死亡率の差」は0を中心に左右対称にばらつくはず。下記のような0を中心とする正規分布になるはず。

でも、SUMMIT研究における結果である-0.75%は、ゼロから結構離れているのでは？
 斜線の面積が十分小さければ、死亡率の差は0から十分離れていると考える⇒帰無予測が誤りと考える
 斜線の面積がP値である
 -2つの治療の死亡率の差

観察される効果量の差(2):
 2つの治療の死亡率の差

研究によっては、仮想的に研究を繰り返した時に取る分布は様々である。同じ群間差でもP値は違う。上の分布の方が、P値が小さく精度の高い研究といえる。

P値: 帰無仮説が正しいと仮定した時に、研究結果がサンプリング誤差によって偶然発生する確率
 (群間差が観察された値と同じか、それよりももっと仮説から離れた値をとる確率)

群間差 0% 2つの治療の死亡率の差

結論を誤る確率

- 対象集団が母集団をうまく代表していないかもしれない
- 測定にはランダム誤差がある
- エラーをいくつまで許容するかを事前に設定する
 P値が事前に定めた値(有意水準)よりも小さければ、仮説を棄却する

研究の結論	母集団の状態	効果がない	効果がある
		正しい	エラー
帰無仮説を採用		正しい	エラー
対立仮説を採用		エラー	正しい

αエラー(偽陽性)とβエラー(偽陰性)

- αエラー(偽陽性)
 - 母集団では(本当は)、関連や効果がないのに、偶然データで効果が示してしまう確率
 - あわて者のα(効果があると早とちりした)
 - αエラーをする確率が、αで、統計学的有意水準と言う
 - 有意水準は、5%や1%とすることが多い
- βエラー(偽陰性)
 - 母集団では(本当は)、関連や効果があるのに、偶然データで効果が示せない確率
 - ぼんやり者のβ(効果を見逃した)
 - βエラーをする確率が、βで、(1-α)を統計学的パワー(検出力)という
 - αは、20%や10%とし検出力を80%や90%に設定することが多い

Statistical analyses
 We calculated the sample size on the basis of proportions,¹⁸ since estimates of death rates per person-time were not available, in an intention-to-treat analysis assuming an estimated neonatal mortality ratio of 46 per 1000 livebirths because this figure would accommodate higher ratios of early infant mortality to 12 weeks post partum. The design effect from clustered randomisation was estimated to be 1.20 for 300 clusters, and anticipated loss to follow-up was 10%, with a 15% allowance for abortions and stillbirths. With a power of 80% and a two-sided significance level of 5%, enrolment and follow-up of 36 500 pregnant women and infants would be needed to detect at least a 15% change in the neonatal or postneonatal mortality in the MMN group relative to the IFA group. Stopping rules drafted by the data and safety monitoring board established a minimum meaningful effect of 10%, thereby dictating a provisional sample size of 83 100 enrollees to assess effect of MMN on infant deaths.

検出力:80%
 有意水準:5%
 両側検定を採用

MMNIによる乳児死亡に対する効果を評価するためのサンプルサイズは83,100人と見積もった

多重性の問題

- 同時に複数の仮説を検定することを、多重仮説検定という
 - 多重仮説検定を行うと偶然有意になる確率(αエラー)が高くなる
- 複数の仮説が、独立な仮説とし、有意水準を5%とすると、偶然に仮説が1つでも有意になる確率は、
仮説の数が2つならば、 $1-0.95^2=0.0975$ (9.75%)
3つならば、 $1-0.95^3=0.142625$ (14.26%)
4つならば、 $1-0.95^4=0.18549375$ (18.55%)
仮説が20個あると、1つでも偶然有意になる確率は64%まで到達する!!

多重性の対策

- ボンフェローニ法(Bonferoni method)
 - それぞれの有意水準を「α/検定回数」にする
 - 検定が厳しくなる分サンプルはたくさん必要
- ゲノム研究で多数の遺伝子型と病気の関連を調べる時などに用いられている
- でも、同じ仮説を別々の研究で検定する時には、有意水準は調整しないのに、ここまで厳しくする必要はあるの?
⇒ 筆者達は、臨床研究には、少し厳しすぎると考えている
- 有意水準を厳しくする必要性は、
- 偽陽性の確率が高い場合
 - 事前確率が低い場合
- 一般に、どの有意水準を用いるかは、仮説の数よりも、検定する仮説の事前確率の方により依存する
- もちろん、P値の小さい結果だけ報告するのは本末転倒。

複数の仮説を立ててはいけないのか?

- 欠点
 - 多重性の問題
 - 結論が不明瞭
 - 利点
 - 多くの研究テーマを扱える、効率が良い
 - 似通った仮説を同時に示すことで妥当性がある
- 対策
主仮説(primary hypothesis)を決めておく

SUMMIT研究においても、

Another primary outcome of the study was maternal mortality. On the basis of the above conditions and assuming an estimated maternal mortality ratio of 400 per 100 000 livebirths with an effect size of 30%, the projected sample size was 126 000. However, the data and safety monitoring board indicated in March, 2004, that maternal deaths in the study had fallen to 274 per 100 000 livebirths, which was substantially lower than the ratio for which the trial was designed. This factor and the termination of the study limited conclusions about maternal mortality.

統計学的解析
の第2段落

仮説「MMN投与は、IFA投与よりも胎児死亡・乳児死亡を減らす」以外にも、仮説「MMN投与群の方が、母体死亡率が低い」も立てたけれども、デザイン通り進まず解析を断念した

Data mining (データ発掘)

- 予期しない関連
 - 偶然の産物の可能性

解釈には慎重であるべきである。報告の中心に差し替えてはいけない。
 - 新しい発見
- 新しい発見の可能性は秘めているため、将来の研究テーマになりうる。研究者が興奮するのも当然。「新しい発見の可能性は秘めているが、その発見の妥当性は担保されていない」

まとめ (1)

- サンプルサイズの計算は、分析的研究でも記述的研究のどちらにおいても重要で、早期に見積もるべきである
- 分析的研究、実験的研究では、統計学検定を行う前駆として、予測因子とアウトカムとの関係について仮説を立てておく必要がある (純粋な記述的研究では、比較を伴わないため不要である)
- 良い仮説とは、1) サンプリング方法や測定方法が明確、2) 単純である、3) 研究開始前に立てたものである

まとめ (2)

- 帰無仮説は、予測因子とアウトカムの間には関連が「無い」とする仮説で、統計学的有意性の検定の基礎である。対立仮説とはその反対で、関連が「有る」とする仮説である。仮説検定とは、帰無仮説が否定されるかどうかを検定するものであり、否定されれば、対立仮説を採用する
- 帰無仮説には、片側のもの(関連が一方のみ)と両側のもの(関連が両方向)がある

まとめ (3)

- サンプルサイズとは、与えられた(調べられた)効果量と変動度の持つ関連を、定めた第I種過誤(α エラー, 偽陽性)と第II種過誤(β エラー, 偽陰性)の確率のもとで、統計学的に有意に検出するのに必要な推定対象者数のことを言う
- 研究立案段階で、複数の仮説を立てることは、研究の効率上望ましいことが多いが、研究の焦点を明確にして、サンプルサイズ計算を可能にするために、1つの主仮説を立てておく必要がある。多重仮説検定の結果や予期せぬ仮説検定の結果の解釈には、それが真である事前確率の大きさを考慮する必要がある

本日覚えるべき単語

- 帰無仮説・対立仮説
- α エラー(偽陽性)・有意水準
- β エラー(偽陰性)・検出力
- P値
- 効果量
- 変動度
- 多重検定・事後仮説・事前確率