

データベースの正規化

正規化とは、これ以上分解できないところまで
テーブルを分割すること

37

データベースの正規化

メリット

- データ管理の容易化
- データの部品化
- データ容量の削減 など

デメリット

テーブルの数が増え、パフォーマンスが低下 など

正規化はデータベース設計には欠かせない概念です。
しかし、正規化すれば必ず優れたデータベースになるとは
限りません。

正規化することにより生ずるメリットと、デメリットの
バランスを考えて正規化することが大切です。

38

参考資料

- 有限会社 パースのWebページ
初級シスアド講座

http://www.pursue.ne.jp/jouhousyo/sysad/kaihatu/water_fall.html

- Webアプリ開発エンジニアのための技術情報サイト「テックスコア」

http://www.techscore.com/tech/sql/16_02.html

- Microsoft Office ACCESS 2003
日経BPソフトプレス

39

DMC研修 : 一斉調査

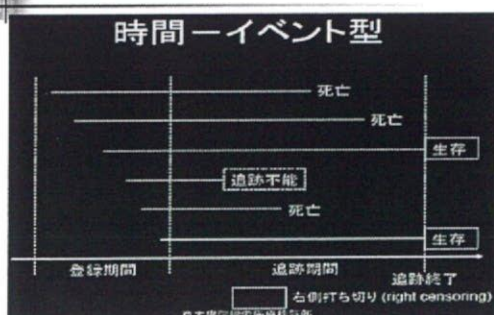
臨床研究データマネジメントセンター
2007/01/18 小野敏子

一斉調査について

- 一斉調査が必要な研究
- GCP と Central Monitoring
- 一斉調査の目的

一斉調査について②

一斉調査が必要な研究:時間-イベント型



目次

- 一斉調査について
→一斉調査の目的とは?
- JPPP試験第一回一斉調査を振り返って

一斉調査について①

一斉調査が必要な研究:エンドポイント

研究デザインのエンドポイントの種類と深い関係

■連続型

臨床検査値、身体的特徴など

■分類型

腫瘍反応、感染の有無など

■時間-イベント型

治療開始、診断などから重要な臨床イベントまでの時間
連続型との決定的な違いは**打ち切り(Sensroing)**の存在

一斉調査について③

～GCP (Good Clinical Practice)～

- 1997年3月27日厚生省令第28号に制定、1997年4月1日施行。
- 倫理的な配慮のもとに(=被験者の人権と安全性の確保)科学的に(=臨床試験のデータの信頼性の確保)適正な臨床試験が実施されることを目的に制定された。

一斉調査について④
～GCP:8-1-22-2 13)～

- i) 治験実施計画書が要求するデータが症例報告書に正確に記載され、それらが原資料と一致していること
- ii) 用量又は治療法の変更があった場合は、その全てが各々の被験者について記録されていること
- iii) 有害事象、併用療法及び併発症が治験実施計画書に従って症例報告書に記載されていること
- iv) 被験者が規定どおりに来院しなかった日、実施されなかった試験及び検査が症例報告書に明確に記載されていること
- v) 登録された被験者の全ての中止例、脱落例が症例報告書に記載され、その理由等が説明されていること。

一斉調査について⑥
一斉調査の目的

モニタリング業務の一環として

イベントを追いかけるタイプの臨床試験において、登録された全ての症例のイベント発生状況を把握する。

一斉調査について⑤
～Central Monitoring～

医療機関などへの訪問によるモニタリングが困難である場合に用いられる方法。
医師や臨床試験協力者に対する訓練や手順書の提供、統計学的にコントロールされた方法でのデータの抽出と検証、医師等との電話・FAXなどによる交信などの手段を組み合わせて臨床試験の実施状況を調査し把握すること。
医師主導臨床研究ではセントラルモニタリングが主流。

JPPP試験第一回一斉調査を振り返って

- スケジュール
- 具体的な作業
- 一斉調査からイベント判定まで
- イベント判定後
- 反省と課題

JPPP試験第一回一斉調査を振り返って①
一斉調査スケジュール

2006年6月・7月
追跡登録画面開発 入力マニュアル作成
8月
各施設へ一斉調査の告知(患者リスト同封)
各施設へ追跡調査票送付
9月
一斉調査開始 第一回データ督促(中旬)
10月
第二回データ督促
11月
データクリーニング 問い合わせ
12月
イベント判定委員会 ステアリングコミッティ開催
2007年1月
反省会

一斉調査データ入力期間
当初の予定
9月1日～10月31日
実際は
8月21日～12月末

JPPP試験第一回一斉調査を振り返って②
具体的な作業

- 1 一斉調査対象症例の確認
- 2 対象患者リストの作成
重複登録等リストからはずれる症例の確認
性別、生年月等割付調整因子の修正がある症例の確認
- 3 一斉調査入力画面入力マニュアル作成・送付(事務局)
- 4 追跡調査票の送付
- 5 追跡調査データの回収(FAX・WEB)
- 6 データ督促(事前に事務局とスケジュール・方法を確認)
- 7 データクリーニング イベント・有害事象症例確認
- 8 イベント判定委員会用資料作成(事務局)

JPPP試験第一回一斉調査を振り返って③

一斉調査からイベント・有害事象判定まで

■ データクリーニング

不適格症例

プロトコール治療中止症例

イベント・有害事象発症例

→ 抽出し、必要に応じて問い合わせ

追跡不能症例

■ イベント・有害事象一覧作成

→ 判定資料へ

JPPP試験第一回一斉調査を振り返って④

イベント判定後

- イベント発生率を算出し、登録症例数の見直しを行った。

石塚先生指示のもとプライマリエvent発生状況一覧を作成



プライマリエvent一覧

Event No.	原因	発症日	判定	登録日	死亡	追跡状況
02-002-006	なし		なし	2006/8/26		
02-002-007	なし		なし	2006/8/26		
02-002-008	なし		なし	2006/8/28		
02-002-009	なし		なし	2006/8/28		
02-002-010	なし		なし	2006/8/28		
02-002-011	なし		なし	2006/8/28		
02-002-012	なし		なし	2006/8/28		
02-002-013	なし		なし	2006/8/30		
02-002-014	なし		なし	2006/8/30		
02-002-015	なし		なし	2006/8/30		
02-002-016	あり	2006/4/11	なし	2006/8/30	死亡	死亡中上
02-002-017	なし		なし	2006/8/30		
02-002-018	なし		なし	2006/8/30		
02-002-019	なし		なし	2006/8/30		
02-002-020	なし		なし	2006/8/30		
02-002-021	プライマリエvent		イベントなしが最後に確認された日 = 死亡日			
02-002-022	による死亡日					
02-002-023	なし		なし	2006/8/30		
02-002-024	なし		なし	2006/8/31		
02-002-025	なし		なし	2006/8/31		
02-002-026	なし		なし	2006/8/31		
02-002-027	なし		なし	2006/8/31		
02-002-028	なし		なし	2006/8/31		
02-002-029	なし		なし	2006/8/31		

JPPP試験第一回一斉調査を振り返って⑤

反省と課題

- 一斉調査にあたり、担当者間(事務局・システム担当・DMC他)のスケジュールや業務手順の共通認識が重要。
- 一斉調査を要する研究のシステム仕様を統一しておくこと今後の開発・運用がスムーズでは。

参考文献

臨床試験の質向上を目指したClinicalDataManagement

吉本 正, 高橋 井久麻, 野村 寿 /編纂

臨床試験データマネジメント-データ管理の役割と重要性-

大橋 靖雄 /監修, 辻井 教 /著

京都大学医学部付属病院 探索医療センター検証部

臨床試験入門コース資料

データベース

高橋由光
国立国際医療センター研究所
国際臨床研究センター 医療情報解析研究部

2007/2/1 @JCRAC DMC

DMCセミナー 課題一覧

発表順	課題	担当者
1	プロトコル作成	鈴木
2	日々のモニタリング	佐々木
3	CRF作成	佐藤
4	モニタリングレポート作成	安原
5	個人情報保護	内田
6	システム開発支援	福田
7	一斉調査	小野
8	IT環境整備 → データベース	高橋
9	ID・パスワード・ROSTER管理	岩本

目次

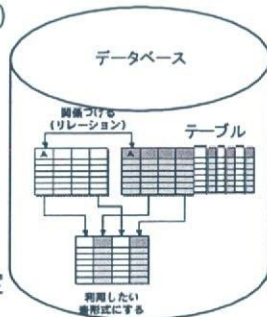
- リレーショナルデータベース
- 正規化
- データ型
- DB設計書
- SQL

リレーショナルデータベース

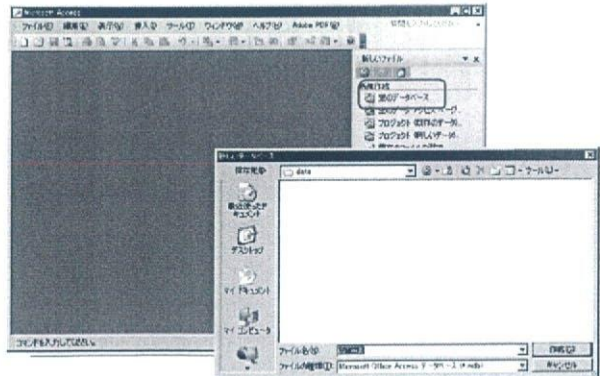
- リレーショナルデータベース(RDB)
 - 複数の表のデータを関係(リレーション)付けることで、すべてのデータをひとつの巨大なデータベースとして活用
 - 目的に合わせてデータを連結させ、求める表を得ることができる
- リレーショナルデータベース管理システム(RDBMS)
 - Oracle(オラクル), MySQL, PostgreSQL(ポストgreSQL), DB2, FileMaker, SQL Server, Microsoft Office Access など

リレーショナルデータベース

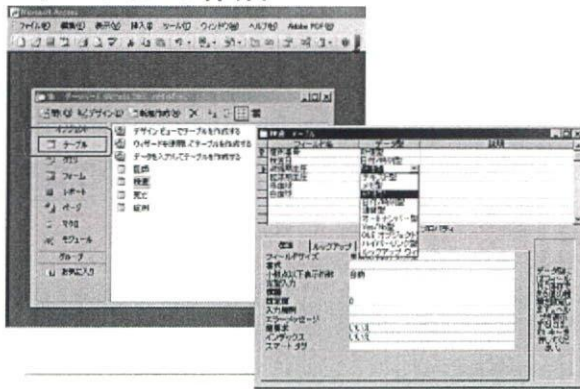
- データベース(Database)
- テーブル(Table)、表
 - 行(レコード)と列(項目)
 - データをもっている
- ビュー(View)、クエリ
 - 仮想的なテーブル
 - データはもっていない
- リレーションシップの設定



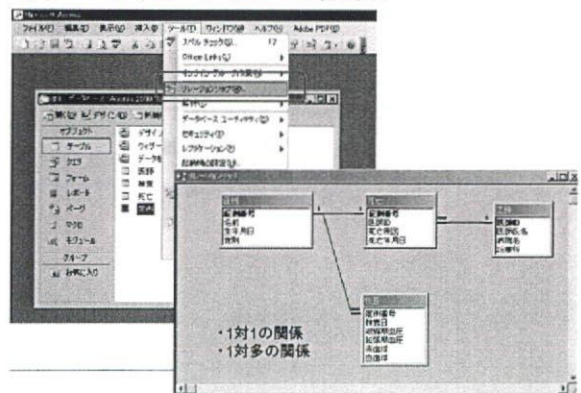
データベースの作成



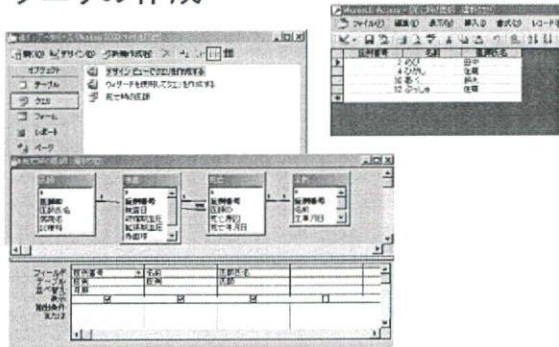
テーブルの作成



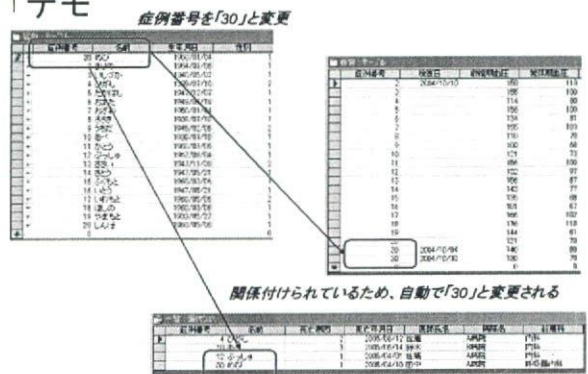
リレーションシップの設定



クエリの作成



デモ

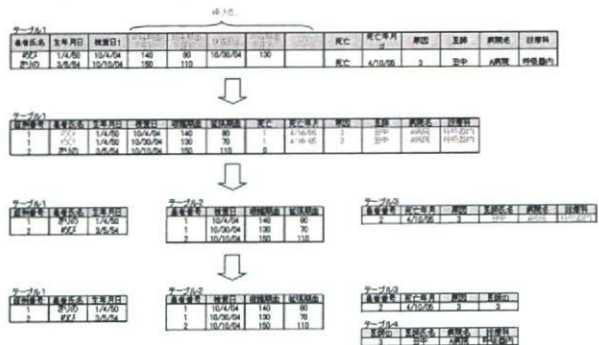


正規化

- 目的
 - 重複データがないように、データの関係(リレーションシップ)を整理 = データの冗長性の減少
- 正規形の種類
 - 非正規形
 - 第一正規形
 - 繰り返す項目を独立したレコードとして分割
 - 第二正規形
 - 主キーの一部の項目だけで決定できる項目を別のレコードとして分割
 - 第三正規形
 - 主キー以外の項目によって決定できる項目を別のレコードとして分割
 - 集計、計算などで導ける項目を削除
 - 第四正規形、第五正規形(省略)
- 主キー/プライマリーキー(PK: Primary key)
 - テーブル内で行を識別するための重複しない値をもつ列

冗長性が減る

正規化の例



非正規形				
得意先コード	得意先名	得意先番号	得意先住所	得意先支店
101	A(株)	01	製造業	aaa
102	B(株)	02	小売業	bbb
103	C(株)	02	小売業	bbb
104	D(株)	03	卸売業	ccc

第一正規形				
得意先コード	得意先名	得意先番号	得意先住所	得意先支店
101	A(株)	01	製造業	aaa
101	A(株)	01	製造業	bbb
102	B(株)	02	小売業	bbb
102	B(株)	02	小売業	bbb
103	C(株)	02	小売業	bbb
103	C(株)	02	小売業	bbb
104	D(株)	03	卸売業	ccc

第二正規形				
得意先コード	得意先名	得意先番号	得意先住所	得意先支店
101	A(株)	01	製造業	aaa
102	B(株)	02	小売業	bbb
103	C(株)	02	小売業	bbb
104	D(株)	03	卸売業	ccc

第三正規形				
得意先コード	得意先名	得意先番号	得意先住所	得意先支店
101	A(株)	01	製造業	aaa
102	B(株)	02	小売業	bbb
103	C(株)	02	小売業	bbb
104	D(株)	03	卸売業	ccc

第一正規形: 繰り返し項目である「販売商品」を独立したレコードとして分割
 第二正規形: 主キーは「得意先コード」。繰り返しのある「販売商品」を独立させ別の表に分割
 第三正規形: 「業種名」の行の重複をさける為別の表に分割

データ型

- 文字
 - CHAR (CHARACTER)
 - テキストをあらわす 固定長
 - VARCHAR (CHARACTER VARYING)
 - テキストをあらわす 可変長
- 数字
 - INT (INTEGER)
 - 整数をあらわす
 - REAL/DOUBLE/DECIMAL
 - 小数をあらわす
- 日時
 - DATE
 - 年、月、日をあらわす
 - TIME
 - 時、分、秒をあらわす
 - TIMESTAMP
 - DATE+TIME

ただし、CapToolは、
すべてVARCHAR

データ型

データ型	記憶領域サイズ	内容
BINARY	1文字につき 1 バイト	このデータ型のフィールドには、どの型のデータでも格納できます。テキスト型 (Text) など、他のデータ型への変換は行われませんが、フィールドに入力された形式のまま、データが出力されます。
BIT	1 バイト	Yes または No の値、あるいは、2 つの値のうちのどちらかしか格納できないフィールド。
TINYINT	1 バイト	0 ~ 255 の整数値。
MONEY	8 バイト	-922,337,203,685,477,5808 ~ 922,337,203,685,477,5807 の範囲の整数。
DATETIME	8 バイト	100 ~ 9999 年の日付または時刻の値。
UNIQUEIDENTIFIER	128 ビット	リモート プロシージャ コールで使用される一意な識別番号。
REAL	4 バイト	-3.402823E38 ~ -1.401298E-45 の負の値、1.401298E-45 ~ 3.402823E38 の正の値、および 0 の単精度浮動小数点値。
FLOAT	8 バイト	-1.79769313486232E308 ~ -4.94065645841247E-324 の負の値、4.94065645841247E-324 ~ 1.79769313486232E308 の正の値、および 0 の倍精度浮動小数点値。
SMALLINT	2 バイト	-32,768 ~ 32,767 の整数値
INTEGER	4 バイト	-2,147,483,648 ~ 2,147,483,647 の整数値
DECIMAL	17 バイト	-1028.1 ~ 1028.1 の値を格納する数値データ型。精度 (1 ~ 28) と桁数 (0 ~ 精度の数値) の両方を定義できます。既定の精度は 18、桁数は 0 です。
TEXT	1文字につき 2 バイト	0 ~ 2.14 GB の文字列。
IMAGE	可変	0 ~ 2.14 GB のデータ。OLE オブジェクトに使用します。
CHARACTER	1文字につき 2 バイト	0 ~ 255 バイトの文字列

Access のヘルプより RDBMS により違いがあります

DB設計書

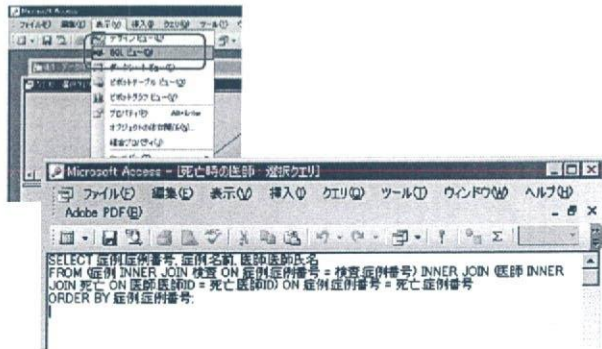
データの型は? 必須項目? 主キーは?

項目名 (日本語)	項目名 (英字)	PK	NULL	属性	桁数
1 医師番号	list_no	●	●	VARCHAR	10
2 レコード登録日	insert_line		●	DATE TIME	
3 レコード更新日	update_line		●	DATE TIME	
4 最新更新者	last_update_user		●	VARCHAR	20
5 最新更新ID	last_update_user_id			VARCHAR	10
6 診療科名	kamei			VARCHAR	60
7 医師名	doctor_name			VARCHAR	60
8 死亡原因	death_cause			VARCHAR	1
9 死亡年月日	death_date			VARCHAR	12
10 コメント	death_comment			VARCHAR	4000
11 患者番号	akushi_shoumu_name			VARCHAR	120
12 外傷の詳細	saisho_detail			VARCHAR	1
13 外傷の詳細の発症詳細	saisho_detail_senota			VARCHAR	120
14 原因その他	cause_senota			VARCHAR	120
15 連番	sequence			INT	11

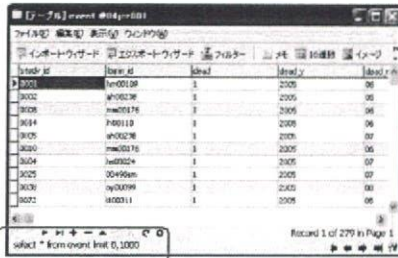
SQL

- SQL (Structured Query Language)
 - データベースの定義や操作などを実現するためのデータベース言語
- ステートメント
 - データの照会
 - SELECT <列名> FROM <テーブル名> WHERE <検索条件>
 - データの挿入、変更、削除
 - INSERT INTO <テーブル名> VALUES (<値>, ...)
 - UPDATE <テーブル名> SET <列名> = <値>, ...
 - DELETE FROM <テーブル名> WHERE <列名> = <値>

SQLビュー



Navicat



event_id	base_id	kind	start_y	end_y
3001	h00100	1	2005	06
3002	h00100	1	2005	06
3006	m00176	3	2005	06
3014	h00110	3	2005	06
3025	h00128	1	2005	07
3030	m00176	3	2005	06
3034	h0020+	1	2005	07
3025	00496m	1	2005	07
3026	h00099	3	2005	06
3022	h00011	1	2005	06

ログファイル C:\Program Files\PremiumSoft Navicat\LogHistory.txt

[2007-01-31 15:25:16][26874] show columns from cptl_applinfo
[2007-01-31 15:27:09][26875] select * from event limit 0,1000

19

参考文献など

- アイテック情報技術教育研究所
『コンピューターシステムの基礎』
- データベースの正規化の基礎
 - <http://support.microsoft.com/kb/283878/>
- Microsoft Office Online トレーニング
 - <http://office.microsoft.com/ja-jp/training/>
- データベースとAccessの基本知識と活用の仕方
 - <http://www.microsoft.com/japan/office/previous/xp/suminak/a/access/database/database0.htm>

20

DMCセミナー

パスワードとROSTER管理

2007年2月8日
臨床研究データマネジメントセンター
岩本智子

1

◆DMCセミナースケジュール

発表順	課題	担当者
1	プロトコール作成	鈴木
2	日々のモニタリング	佐々木
3	CRF作成	佐藤
4	モニタリングレポート作成	安原
5	個人情報保護	内田
6	システム開発支援	福田
7	一斉調査	小野
8	IT環境整備	高橋
9	ID・パスワード・ROSTER管理	岩本

2

◆本日の内容

■パスワードとROSTER管理の一般論

- ◆一般的な情報セキュリティはどうなっているか
- ◆なぜパスワード・ROSTER管理が必要なのか
- ◆パスワード管理基準
- ◆ROSTER管理

■パスワードとROSTER管理DMCの現状

- ◆存在するパスワード、ROSTER
- ◆ID・パスワード発行体系
- ◆ID・パスワード発行体系 問題①
- ◆ID・パスワード発行体系 問題②
- ◆ROSTER管理

3

パスワードとROSTER管理の一般論

4

■一般的な情報セキュリティはどうなっているか

- ◆入室管理
- ◆ファイアウォール

物理的対策



- ◆情報セキュリティポリシー
＝決まりを作り、全員が決まりを守る。

情報セキュリティポリシーの一例

「経済産業省 情報セキュリティ管理基準」

- 資産の分類及び管理
- 物理的及び環境的セキュリティ
- アクセス制御
- 人的セキュリティ
- 通信及び運用管理
- システムの開発及び保守

etc. . .

5

■なぜパスワード・ROSTER管理が必要なのか

経済産業省 情報セキュリティ管理基準

7.2 利用者のアクセス管理

目的: 情報システムへの認可されていないアクセスを防止するため

7.3 利用者の責任

目的: 認可されていない利用者のアクセスを防止するため

7.4 ネットワークのアクセス制御

目的: ネットワークを介したサービスの保護のため

6

■パスワード管理基準

経済産業省 情報セキュリティ管理基準

7.3 利用者の責任

7.3.1.3 すべての利用者に、システム又はパスワードに対する危険の兆候が見られる場合は、パスワードを変更するように助言すること

7.3.1.4 すべての利用者に、最短6文字の質の良いパスワードを選択すること

7.3.1.5 すべての利用者に、パスワードは定期的に、又はアクセス回数に基づいて変更するように助言すること

7

■パスワード管理基準

◆「質の良いパスワード」とは

・通常8文字以上

・数字のみ、英文字のみではなく、記号や大文字小文字を組み合わせる。

・ユーザIDと同じものや、名前、電話番号、誕生日などを避ける。

・辞書に載っているような単語は避ける。

・メモしない。メモしなければ覚えられないようなものは不適切である。他人に教えない。システム管理者であってもパスワードを尋ねることはない。

(IPA 独立行政法人 情報処理推進機構ホームページ Q&Aより抜粋)

9

■ROSTER管理

◆「ROSTER」=名簿…情報資産

- ▶ 氏名・電話番号などの個人情報
- ▶ ID・パスワード…個人情報ではないが、氏名と同じ表になっていたら、個人情報(内田さんのセミナーより)

◆情報資産の分類とは

- ▶ 重要性による分類
- ▶ 機密性、完全性、可用性の3つの側面から重要性を検討し、情報資産を分類する。

9

■ROSTER管理

機密性

アクセス権を持つ者だけが、情報にアクセスできることを確実にすること。
⇒機密事項が漏れないようにする

完全性

情報および処理方法が正確であることおよび完全であることを保護すること。
⇒不正な改ざんがあってはいけないが、適正な更新は必要

可用性

認可された利用者が、必要ときに、情報および関連する資産にアクセスできることを確実にすること。
⇒サービス停止しないこと

10

パスワードとROSTER管理DMCの現状

11

■存在するパスワード

◆メールアドレスのパスワード

- ▶ 研究・試験アカウント(25)
- ▶ 個人アカウント

◆パソコンのユーザーアカウントのパスワード

◆研究・試験のEDC (Electrical Data Capturing 電子的臨床試験データ収集システム) 画面ログインパスワード

- ▶ 管理者権限
- ▶ システム管理者権限
- ▶ 研究・試験参加者権限

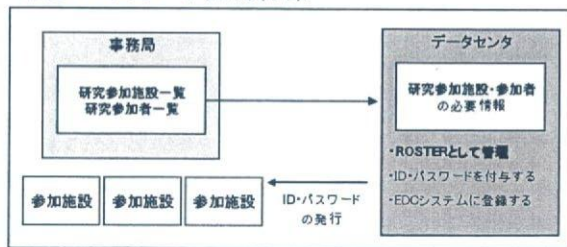
⇒研究・試験参加者へ、DMCが発行するパスワードなど

12

■存在するROSTER(例)

- ◆ DMC職員名簿
 - ▶ 氏名
 - ▶ 電話番号
- ◆ 研究・試験参加者名簿
 - ▶ 参加医師氏名
 - ▶ 参加医師の所属
 - ▶ EDCへのログインIDやパスワード
- ◆ 研究・試験参加施設一覧
 - ▶ 施設名称
 - ▶ 住所
 - ▶ 電話番号
 - ▶ 研究・試験代表者氏名

■ID・パスワード発行体系



- ◆ 重複したID・パスワード作成しないこと
- ◆ 複数の人に同一のID・パスワードを発行しないこと
- ◆ ID・パスワードは確実に本人に届くこと

■ID・パスワード発行体系

国立病院機構
EBM推進研究での例

- ◆ 重複したID・パスワードを作成しないこと
- ◆ 複数の人に同一のID・パスワードを発行しないこと

●個人パスワード
フリーソフト(『パスワード
クリエーター』)を使用し、
重複のない設定で大量に
作成する。機構EBM研究の
個人パスワードは3000
作成

●個人ID
管理をしやすくするため
の連番と二文字アルファ
ベットの組み合わせ

すでに付与したものは色
をつける。

■ID・パスワード発行体系

- ◆ ID・パスワードは確実に本人に届くこと

①今回発行者にチェックをし、
クエリ抽出する。

②『エクセルシート』にコピー
する。

■ID・パスワード発行体系

- ◆ ID・パスワードは確実に本人に届くこと

③『エクセルシート』は、自動メール
送信のフリーソフトである。「SB」～
「SH」の内容を自動的に文章に挿
入し、A列のアドレスに送信する。

JPPP試験・日循などでも使用

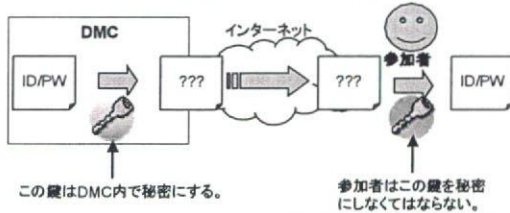
■ID・パスワード発行体系

- ◆ パスワードは確実に本人に届くこと

研究・試験名	パスワード発行方法
J-SAP	郵送(配達記録)
J-CHF	CDを郵送(書留)
J-DHF	メール
J-ELAN	メール
JPPP	メール
機構研究	メール
日循	メール・郵送
ACTION-J	郵送(配達記録)

■ ID・パスワード発行体系 問題①

- ◆ メールで発行する場合の安全性
 - ▶ 現状はメール文面に入れており、漏洩・不正利用の可能性がある。(応急処置:メールを分ける)
- ◆ 対策…例:「公開鍵方式」の暗号化



この鍵はDMC内で秘密にする。

参加者はこの鍵を秘密にしなければならない。

ID/PW発行は基本的に1回限りなので、不向き
暗号化を実行するには、運用する参加者の高い意識・理解が必要

19

■ ID・パスワード発行体系 問題②

◆ パスワード変更頻度

大分類	詳細	変更状況
メールアカウントPW	個人	個人責任
	試験・研究	不定期
パソコンの ログインPW	個人使用	個人責任
	出張用	使用の毎に変更
試験・研究EDC 画面ログイン用	研究参加者 (DMCより発行されたPWの使用者)	個人責任 ^(*)
	試験・研究管理者権限	不定期
	システム管理権限	不定期
	二重認証の共通PW (例えば 機構研究の施設PWなど)	1年に1回 (2007年は未実施)

「不定期」については決まりを作らなくてはならない。

「*」については変更の必要性を、周知してもらう。

20

■ ROSTER管理

- ◆ 研究・試験参加者名簿
- ◆ 研究・試験参加施設一覧

「DMC職員名簿」
については省略

- ▶ 機密性…保たれている。全てDMCのファイルサーバ内にデータで保管されており、アクセスを許可されたユーザーアカウントしか入れない。
- ▶ 完全性…「適正な更新」が保たれていないものもある。
⇒ 国立病院機構EBM推進研究の参加者マスタ
- ▶ 可用性…保たれている。

問題点

21

■ ROSTER管理

- ◆ 現状の問題点
 - ▶ 「国立病院機構 EBM研究参加者マスタ」が適正に更新されない。
- ◆ 原因
 - ▶ 変更や削除の情報を得られないことがある。
 - ▶ 変更や削除の申請手順が煩雑
- ◆ 対策案
 - ▶ 変更や削除の手続きを、ID・パスワード保有者に周知させる。
 - ▶ 機構本部と連携しての対策が必要

22

■ 参考文献

- ◆ 「インターネット・ネットワーク入門」
監修:水岡祥二 株式会社アイテック 情報処理技術者教育センター
- ◆ 「臨床研究の進め方」
編纂:大橋靖雄・荒川義弘 南江堂

ご清聴
ありがとうございました。

■ 参照ホームページ

- ◆ 財団法人 日本情報処理開発協会
<http://www.jpdec.jp/>
- ◆ IPA独立行政法人 情報処理推進機構
<http://www.ipa.go.jp/index.html>
- ◆ 総務省 国民のための情報セキュリティサイト
http://www.soumu.go.jp/joho_tsusin/security/index.htm
- ◆ 経済産業省 情報セキュリティに関する政策・緊急情報
<http://www.meti.go.jp/policy/netsecurity/audit.htm>

23

臨床研究のデザイン

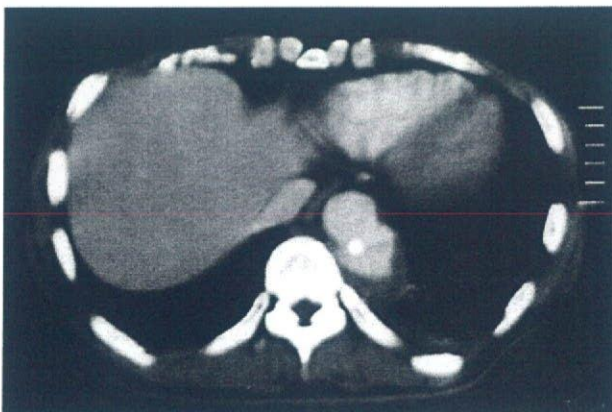
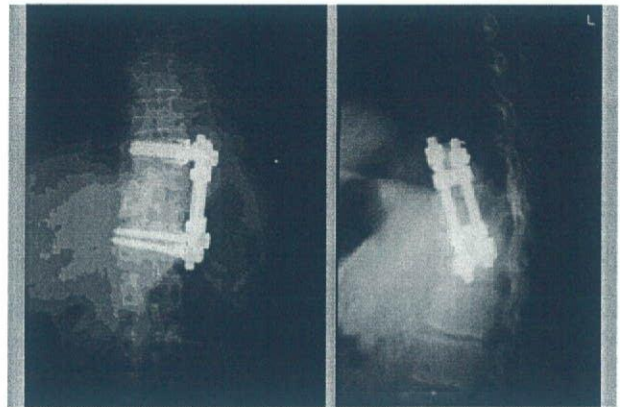
国立国際医療センター研究所
医療情報解析研究部
新保卓郎

臨床研究の目標:

- 1) 記述的研究: 問題の大きさの把握。
有病率、罹患率、死亡率などの評価。
- 2) 分析的研究: 関連(Association)の検討。
予測、因果関係の検討。
変数間の関連を検討する方法
平均値の比較
比率の比較
相関
速度の比較

臨床研究の種類

- 観察研究
 - 記述的研究
 - 1) 症例報告
 - 2) 症例集積
 - 分析的研究
 - 1) 症例対照研究
 - 2) 横断研究
 - 3) コホート研究
- 介入研究
 - ランダム化比較試験
- データ統合型研究
 - 系統的レビュー・メタ分析
 - 決断分析
 - 医療経済分析
- その他
 - 質的研究
 - 生態学的研究
 - 検査特性に関わる研究
 - 診療ガイドライン



診断: 仮性大動脈瘤

- 脊椎固定術に使用したピンが原因(?)
- 症例報告では、仮説を提示できるが、検証ができない。

症例集積(case series)

- 1956年Crowlは多発性骨髄腫に末梢神経炎を合併した2例を報告した。
- 1981年にPneumocystis carinii肺炎の5例が報告された。彼らは男性同性愛者であった。

症例対照研究(case control study)



方法

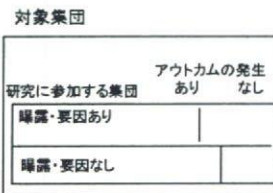
- 1) 疾患をもつ対象集団から研究に参加する疾患群を選択する
- 2) 疾患をもたない対象集団から研究に参加する非疾患群を選択する
- 3) 両群で曝露・要因を測定し、比較する。

症例対照研究における関連の検討

		アウトカム	
		あり	なし
曝露・検査・治療	あり	a	b
	なし	c	d

Odds ratio = a:c/b:d = ad/bc
 関連なしではOdds ratio = 1

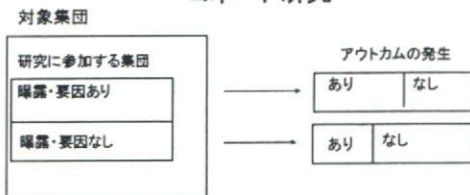
横断研究(Cross sectional study)



方法

- 1) 対象集団から研究に参加する集団を選択する
- 2) 曝露・要因とアウトカムの発生を測定する
- 3) 関連を検討する

コホート研究



方法

- 1) 対象集団から研究に参加する集団を選択する
- 2) 曝露・要因を測定する
- 3) コホートを追跡する
- 4) アウトカムの発生を測定し、比較する

コホート研究における関連の検討

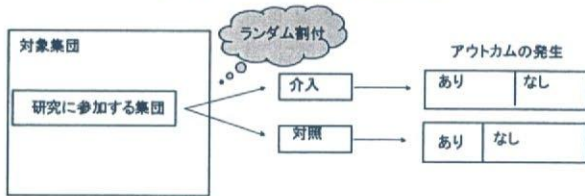
		アウトカム		
		あり	なし	
曝露・治療など	あり	a	b	n
	なし	c	d	m

RR (Relative Risk) = (a/n)/(c/m)

		アウトカム		観察人年
		あり	なし	
曝露・治療など	あり	a	b	PY _e
	なし	c	d	PY _c

IRR (Incidence rate ratio) = (a/PY_e)/(c/PY_c)

ランダム化比較試験:RCT



方法

- 1) 対象集団から研究に参加する集団を選択する
- 2) ランダムに介入群と対照群に割り付ける
- 3) 追跡する
- 4) アウトカムの発生を測定し、比較する

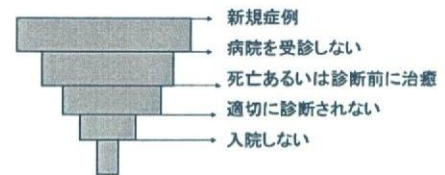
臨床研究の3つの問題点

- Chance (偶然)
 - 研究参加者が集団の中で、偶々そのような人達であったのではという可能性。
 - 統計学的に対応 (推定、検定を行う)
- Bias (selection bias, information bias)
 - 測定値が真実から系統的に異なっている
- Confounding (交絡因子)
 - 共変量の影響で、関連がゆがんでみえる

Selection bias (選択バイアス)

- 市中肺炎発症の原因に腎機能障害がなっているかを検討するため、入院患者を対象として症例対照研究を行った。症例群は入院した肺炎患者、対照群は入院した尿路以外の癌患者。
- その他の選択バイアス
- Spectrum bias
- Publication bias (出版バイアス)

症例対照研究に参加した症例群がもとの集団を代表していない原因の例



選択のされ方が曝露・要因と関連があり対照群と異なるとバイアスとなる

Information bias (情報バイアス)

- アトピー性皮膚炎の増悪に、日常生活習慣が関係するのではと医師が疑い、重症患者と軽症患者で、日常生活習慣を比較した。両群の患者に、食事、睡眠、ストレス、飲酒・喫煙などについて医師が質問し記録した。
- 新生児の先天異常の原因を検討するため、疾患児を出産した母と、健常児を出産した母に、妊娠中の特定食品の摂取について質問した。(Recall bias)
- 耐糖能異常者からの糖尿病発症を糖尿病教室が予防できるか5年間に渡るRCTで検討された。糖尿病教室6回群と2回群に割り付けられ、糖尿病教室時にGTTで糖尿病の発症が確認された。

情報バイアスにおける2つのMisclassification (誤分類)

- Non-differential misclassification (無差別的誤分類)
- (例) 虚血性心疾患の既往歴の有無により、予後に違いがあるか比較したいが、虚血性心疾患の既往歴の情報が予後の状態によらず同様に不正確
- バイアスは関連を薄める方向に働く。バイアスの方向が予測できる。
- Differential misclassification (差別的誤分類)
- (例) 重症の患者にのみ丁寧に既往歴を聴取する。
- バイアスは関連を小さくしたり大きくする方向に働く

無差別的誤分類の例

真実	測定結果	
	重症	軽症
既往歴あり	90	60
既往歴なし	10	40

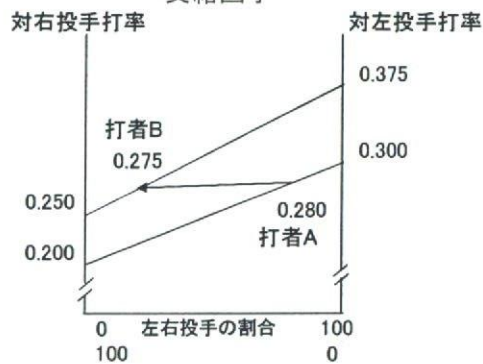
RR=6.0 RR=2.8

重症群、軽症群の両群共に、既往歴ありの20%を誤分類した場合

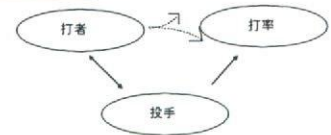
	重症	軽症
既往歴あり	90	60
既往歴なし	18	12

- 野球で、打率が2割8分のバッター(A)は、2割7分のバッター(B)より打率が高い。(?)

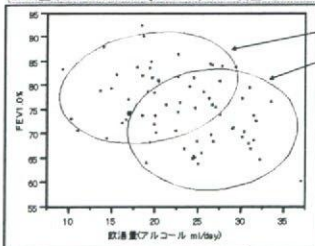
交絡因子



全打席	ヒット	凡打	打数	打率
打者A	280	720	1000	0.280
打者B	275	725	1000	0.275
対左				
打者A	240	560	800	0.300
打者B	75	125	200	0.375
対右				
打者A	40	160	200	0.200
打者B	200	600	800	0.250



飲酒量(アルコール ml/day)とFEV1.0%の二変量の関係



非喫煙者
喫煙者
それぞれの群単独では飲酒量とFEV1.0%の間に相関はないが、一緒に分析すると相関があるようにみえてしまう。

二変量正規検定 P<0.000
二変量正規検定 P<0.000 喫煙状態=0
二変量正規検定 P<0.000 喫煙状態=1

喫煙者で飲酒量が多い
喫煙者でFEV1.0%がさがる

交絡因子に対する対応

デザインの段階で考慮

- 対象者の限定
 - Matching
- 分析の段階で考慮
- 層化(Stratified analysis)
 - 統計的補正

症例対照研究の利点と欠点

- 利点
 - まれな疾患・アウトカムの検討に適する
 - 短期間で行える
 - 多くの曝露・要因について検討できる
 - 小さい研究費で行える
 - 小規模で行える
- 欠点
 - バイアスの危険性が高い(選択、情報バイアス)
 - 時間の前後関係が不明確
 - 一つの疾患・アウトカムについてしか検討できない。

コホート研究の利点と欠点

- 利点
 - 時間の前後関係が明確
 - 曝露・要因、アウトカムの測定が正確
 - 複数のアウトカムについて検討できる
 - 罹患率を検討できる
- 欠点
 - 大きな参加対象者が必要
 - 研究期間が長い
 - 費用が高い
 - 交絡因子の可能性が残る

RCTの利点と欠点

- 利点
 - コホート研究の利点をもつ
 - 交絡因子の影響を小さくできるため、因果関係について説明しやすい。
- 欠点
 - 時間と費用が大きい。
 - 倫理的に実施不可能なCQがある。
 - 標準化された介入は実地診療と異なる可能性がある
 - 研究参加者が極度に限定されており、結果の一般化に注意がいる。

診断：“検査特性”に関わる研究

検査の有用性や限界を、定量的に表現する。
その結果、意思決定、予測、コミュニケーション
や教育に寄与できる。

	疾患あり	疾患なし
検査陽性	真陽性	偽陽性
検査陰性	偽陰性	真陰性

感度 = 真陽性 / 疾患あり 特異度 = 真陰性 / 疾患なし

関連と因果関係

- 関連から因果関係を推測するための基準 (Hill, 1965)
- 1) Strength: 関連の強さが大きい
- 2) Consistency: 同様の研究結果の報告
- 3) Specificity: 1つの原因が1つの結果を生じる (批判大)
- 4) Temporality: 原因が結果の前
- 5) Biologic gradient: 量反応関係がある
- 6) Plausibility: 生物学的にもっともらしい
- 7) Coherence: ほかの知識と一致
- 8) Experimental evidence: 実験的に原因を除くと結果が起きない
- 9) Analogy: ほかの因果関係の事例との類似性

臨床研究デザインとClinical questionの領域

	有病率	検査	治療・介入	予後	有害事象	因果関係
観察研究						
症例報告・集積			△	△	△	
症例対照研究		○	○		○	○
横断研究	○	○				
コホート研究	○	○	○	○	○	○
介入研究						
RCT			○	○	○	○
データ統合型研究						
メタ分析		○	○		○	○
決断分析・費用効果分析			○			

本日のまとめ

- Clinical questionに答えるために、いろいろな研究デザインを用いる。得失を理解して選択する必要がある。
- 選択されたデザインの中で、Chance, Bias, Confoundingに影響されないよう、研究計画を立てる必要がある。
- 実際にやってみるのが、理解するための優れた方法である。



医学統計：基礎中の基礎 誤用してるor誤解してる？

国立国際医療センター
研究所地域保健医療研究部
予防医学研究室長
石塚直樹

1



医学統計とは？



- 計量生物学 ⇨ Biometrics
- 生物統計、バイオ統計学 ⇨ Biostatistics
- 医学統計 ⇨ Medical Statistics
- 臨床生物統計 ⇨ Clinical Biostatistics
- 他に医薬統計、医療統計
- 統計学を臨床研究に應用ではない
- 臨床研究の問題を統計学的に解決する学問体系
- マーケティングや品質管理で用いられることを前提とした統計学の本は薦められない！

2

臨床研究で頻出する統計関連用語 どれだけ知ってる？

- | | |
|---------------------|------------------------|
| 1. 平均値、分散、標準偏差、標準誤差 | 21. カプラン・マイヤー法 |
| 2. F値 | 22. ロジスティック回帰 |
| 3. 有意水準と検出力 | 23. コックス回帰 |
| 4. 帰無仮説と対立仮説 | 24. Intention To Treat |
| 5. 両側検定と片側検定、信頼区間 | 25. 感度と特異度 |
| 6. カイ二乗検定 | 26. 交互作用と交絡 |
| 7. スチューデントt検定 | 27. リスク比、オッズ比、ハザード比 |
| 8. 対応のあるt検定 | 28. 中央値(median) |
| 9. 分散分析 | 29. 尤度比検定 |
| 10. ウィルコクソン順位検定 | 30. ランダム割付、最小化法 |
| 11. ウィルコクソン符号付順位検定 | 31. 層別解析 |
| 12. マン・ホイットニーU検定 | 32. マクネマー検定 |
| 13. Fisher直接確率検定 | 33. カップ{(n)}係数 |
| 14. マンデル・ハントセル検定 | 34. サイモン2段階デザイン |
| 15. ピアソン相関係数 | 35. 多量性 |
| 16. スピアマン相関係数 | 36. ボンフェローニ修正 |
| 17. 偏相関係数 | 37. ダネット検定 |
| 18. 回帰分析 | 38. 非劣性試験 |
| 19. 共分散分析 | 39. クロスオーバー試験 |
| 20. ログランク検定 | 40. 中間解析 |

3

臨床研究における医学統計の役割 1

- 求めるエビデンス
 - より効果のある治療法は？
 - 同程度の治療成績ならコストや副作用がない治療法は？
- 臨床研究の基本は比較
 - 同じ治療法で効く人もいれば効かない人も
 - 偶然と必然を区別するために統計学が必要
- 比較の方法は統計的仮説検定
 - 有意差あり → 差がある(必然)
 - 有意差なし → 差があるとは言えない(偶然)
- 研究目的と背景から
 - どのようなデータをどのように得るのか？
 - どのように解析するのか？
 - どのような結論が得られるのか？
 - 差があった理由は？



集めたデータで有意差を出すことが仕事ではありません(^_^;) 4

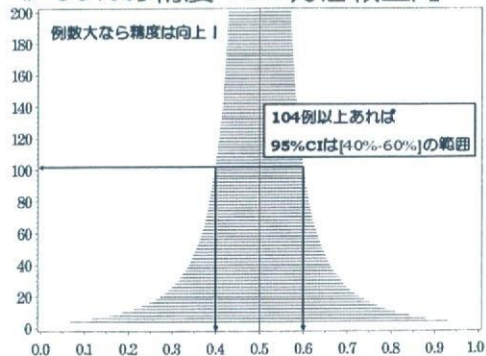
臨床研究における医学統計の役割 2

- 「偶然と必然を区別する」
 - たまたま効いた？
 - たまたま効かなかった？
 - たまたま副作用が出た？
 - たまたま副作用が出なかった？
 - 偶然：帰無仮説
 - 本当に有効な治療法だから治療成績が良いのか？
 - 本当に有効ではない治療法だから治療成績が悪いのか？
 - 必然：対立仮説
- 「デザインと解析」
 - 生物には個体差があって、必ず同じ反応をするとは限らない
 - 反応に対する未知の要因を誤差として扱う
 - 反応に対する既知の要因を管理することも重要
 - 「デザイン」 > 「解析」

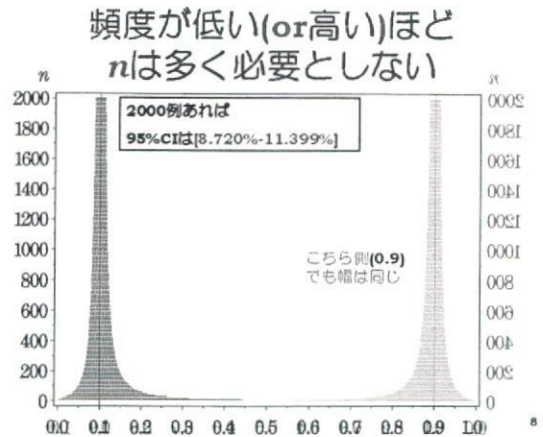
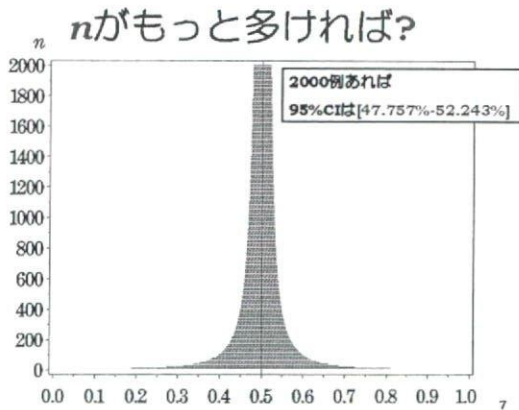
5

2人に1人、効果があっても50%？

50%の精度：95%信頼区間



6



ピロリ菌感染歴、胃がんリスク10倍 厚労省調査 朝日新聞 2006年9月5日

- 人の胃にすみ着く細菌ピロリ菌(ヘリコバクター・ピロリ)に感染したことがある人は、全く感染したことがない人に比べ、胃がんになるリスクが10倍に跳ね上がる。全国4万人余を対象にした厚生労働省研究班の大規模疫学調査でわかった。ピロリ菌と胃がん発症との因果関係が一段と濃厚になってきた。28日から横浜市である日本癌(がん)学会で報告する。
- 90~95年に全国10地点で40~69歳の男性約1万5300人、女性約2万6700人に血液を提供してもらい、経過を追ったところ、04年までに512人が胃がんになった。保存血液でピロリ菌感染を調べ、胃がんにならなかった人と比べて、
- その結果、採血時にピロリ菌陽性の人たちの胃がん発症リスクは、陰性の人たちの5.1倍だった。
- 菌の感染で胃炎が進むと、胃の中の環境が変わり、ピロリ菌がすみやすくなる。こうした場合は陰性に含まれてしまうため、研究班が他の指標も併用して採血時までの感染歴の有無で比べると、感染歴のある人たちが胃がんになるリスクは、感染歴のない人たちの10.2倍に達した。
- ただ、ピロリ菌感染歴があっても、胃がんを発症するのはごく一部とみられる。
- 研究班によると、抗生物質でピロリ菌を除去しても胃がんを防げるかどうか、疫学調査でははっきりしていないという。除菌は現在、胃や十二指腸潰瘍(かいよう)に限って公的保険が適用されている。
- 研究班の桂月静・国立がんセンター予防研究部長は「除菌も選択肢の一つだが、疫学調査では喫煙や塩分の高い食事など、がんの危険を高める生活習慣を改めることがより大切だ」としている。

ピロリ菌除菌で胃がん発症率3分の1 感染者調査で判明 朝日新聞 2003年9月28日

- 胃潰瘍(かいよう)を起すこととされる細菌、ヘリコバクター・ピロリ(ピロリ菌)を薬で除菌すると、胃がんになるリスクが3分の1になるという調査結果を法政正博・北海道大学教授(消化器内科学)らがまとめた。感染者約3400人を調べた結果で、除菌による胃がんの予防効果も明らかになった。27日まで名古屋市で開催された日本癌(がん)学会で発表された。
- 同大学や東北大、信州大など全国23の医療施設が協力し、ピロリ菌感染者で除菌をしていない11233人と、除菌した2186人について、5年以上にわたって胃がん発症の有無を調べた。
- その結果、除菌していない人では3.5%(43人)が胃がんになった。一方、除菌した人で胃がんになったのは1.1%(23人)で、除菌によって胃がんの発症率が3分の1以下に抑えられた。
- ピロリ菌は胃の粘膜に感染し、炎症や潰瘍につながる。胃潰瘍の治療では、薬による除菌が普及している。しかし、除菌で胃がん発症を予防できるかどうかはまだ明らかになっていなかった。
- 胃がんはがんの死因の第2位で、01年には約5万人が死亡した。法政教授は「食生活などの生活習慣を改めるとともに除菌をすることで、胃がんの7.8割を抑えられる可能性がでた」と話している。
- ◆富永拓茂・愛知県がんセンター名誉館長の話
- ◆大規模試験の結果であり、除菌の胃がん予防効果もほぼ証明されたと思う。今後は、発症するリスクが高い人など、どのような人を対象に除菌をするかを究極めなければならない。

ピロリ菌の臨床試験、被験者不足で難航 朝日新聞 99年4月4日

- 日本人の約半数が感染しているといわれる細菌ヘリコバクター・ピロリについて、胃がんの原因かどうかを調べる厚生省の臨床試験が被験者不足で難航している。除菌希望者が多く、約10年間胃に菌を残したまま経過を観察させてくれる被験者が集まりにくいことが原因だ。試験担当の医師たちは「除菌をしないと、副作用が出ることもあるのに」と困惑している。
- ピロリ菌は胃・十二指腸にかいようの原因で、欧米の疫学的研究から、前がん状態の慢性萎縮性胃炎、さらに進んで胃がんを起す、との説が有力だ。しかし、これを否定する調査もあり、胃がんの多い日本の研究が必要とされていた。厚生省の「発がんの高度危険群を対象とした予防研究」班(班長・豊島幸生・国立がんセンター中央病院長)が臨床試験をすることに。1997年1月から、病院を訪れる人などに協力を呼び掛け始めた。計画によると、ピロリ菌に感染した20歳から59歳までの男女を、抗生物質などで菌を殺す「除菌組」とそのまま経過する「感染組」に分け、約10年間検査を受けてもらう。慢性萎縮性胃炎の原因かを調べるには700人、胃がんとの関連まで突き止めるには5000人が必要になる。
- ところが、協力を呼び掛けた人のうち、7割が除菌を希望した。また、「感染組」「感染組」のどちらに入るか自分で決められないため断る人も多く、当初の参加締め切りだった昨年未だに承諾したのは約500人だけだった。
- 推定約6000万人の日本人がピロリ菌に感染しているが、胃がん患者の発生は年24万人で、仮に患者全員が感染者としても0.4%にすぎず、大部分は一生がんを発症しない。一方、除菌者には陣やけが強く出る逆流性食道炎や、胃と食道の境付近のがんが増えるなどの副作用もあることがわかってきた。

仮説、必要なデータ、倫理、結果の解釈 ピロリ菌を例にして

- 4種類のデータ収集：研究デザイン
 - 胃ガンの発症率とピロリ菌の感染率を地域ごとに調べる
 - 胃ガン患者と胃ガンのない人の過去のピロリ菌の感染の有無を調べる
 - 胃ガンでない人を対象にピロリ菌の有無とその後の胃ガンの発症の有無を調べる
 - ピロリ菌に感染している胃ガンでない人を対象に、ランダムにピロリ菌を除去するグループと除去しないグループに分け、その後の胃ガンの発症の有無を調べる
 - 倫理
 - 無理やり感染させるグループ、感染しないグループのランダム化は許されない
 - 統計解析で示される結果
 - 示すべき指標がサンプリングで違う
 - 行・列の関連はカイ二乗>3.84ならP<0.05で有意
 - 同じ仮説でも、データの収集方法(研究デザイン)は複数あり、それに伴って倫理、結果の解釈が異なる
 - エビデンスにレベルの違いがある
- | | | 胃ガン | | |
|------|---|----------------|----------------|----------------|
| | | あり | なし | 計 |
| ピロリ菌 | 無 | a | b | R ₁ |
| | 有 | c | d | R ₂ |
| 計 | | C ₁ | C ₂ | N |
- $$\chi^2 = \frac{(ad - bc)^2 N}{R_1 R_2 C_1 C_2}$$
- http://www.cebm.net/levels_of_evidence.asp