

2003 generate: year=1day/305.25
2004 sbst: year, failure (event, as=1)
2005 sts graph, by (intmatch) dds (Overall survival)
censored (multiple) y000c (Probability) style
(Years after transplant) label (0) (2) 12

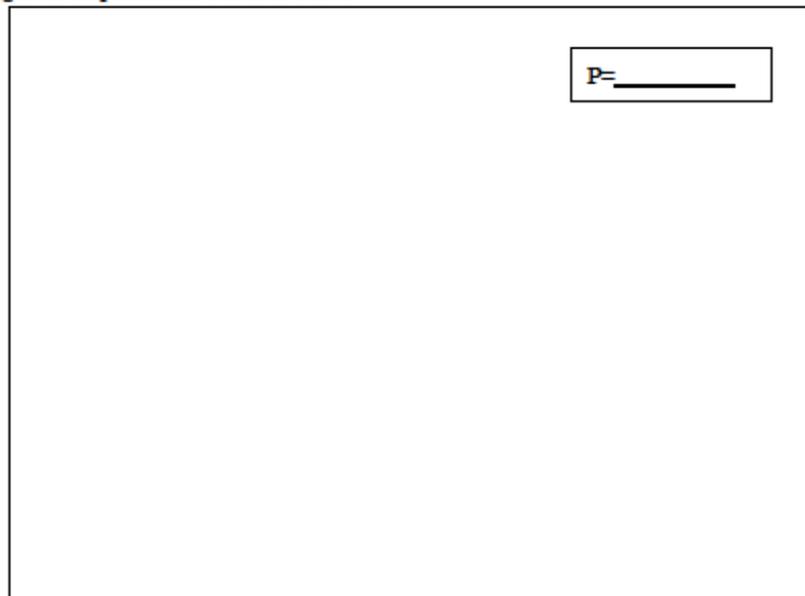


ミニデータセット ワークシート

Table 1. Characteristics of study patients

Characteristic	HLA-matched		HLA-mismatched		P value
	N	(%)	N	(%)	
Total					
Age at transplant, years, median (range)					
		(-)		(-)	
Gender					
Male		(%)		(%)	
Female		(%)		(%)	
Disease					
AML		(%)		(%)	
ALL		(%)		(%)	
Year of transplant					
2000-2005		(%)		(%)	
2006-2010		(%)		(%)	

Figure 1. Kaplan-Meier Survival Curve



/**** 解析データセットの作成 ****/

*Stata を起動して、ファイルの中にある、minidataset_2013.xls を取り込んでください。

*余裕があれば、log file を作成してみてください。

*方法は問いません。

```
*** import excel "C:¥data¥minidataset_2013.xls", firstrow
```

```
*** log using "C:¥data¥mini20131020.smcl"
```

```
*** あるいは、新たなログファイルを"Log Begin"のタブから作成
```

/////////****新たな変数作成の練習****/////////

* 準備運動 *

```
/**generate, replace**/
```

```
generate new_disease=""
```

```
    *新たな変数 new_disease が作成された
```

```
replace new_disease=dx
```

```
replace new_disease="AML" if diagnosis==1
```

```
replace new_disease="ALL" if diagnosis==2
```

```
generate new_disease2=.
```

```
replace new_disease2=1 if new_disease=="AML"
```

```
replace new_disease2=2 if new_disease=="ALL"
```

```
/**症例削除の練習 **/
```

```
/**drop, keep**/
```

```
drop new_disease
```

```
    * 変数 new_disease が削除される。戻らないので注意！
```

```
drop if diagnosis==1
```

```
keep if diagnosis==4
```

*一度 Stata を閉じてください。

* あるいは、clear コマンドを使用。データはすべて消えてしまいます。

/////////****変数作成の練習****/////////

配布資料：ミニデータセット ワークシートの作成

*新たに配布 DVD/USB 内の "Minidata_2013"をダブルクリックして開く

```
*** use "C:¥data¥Minidata_2013.dta", clear のコマンド使用も可能
```

/****データセットを Data Editor で確認 ****/

*DataEditor を使って、データのシートを開いてください。

```
*Variables, Properties, Data editor (Browse)
```

```
*Label, Value Label
```

```

/**変数ラベルの概念についての復習：変数「sex」におけるラベル「性」**/
/**項目ラベルの概念について：変数「diagnosis」の中のデータ「1」に対するラベル「AML」**/

/*症例数確認*/
count

/****不要データの削除****/
*****
* STEP
* 1) AML, ALL, CMLの中からCMLを削除
* 2) survival dataが欠損の症例を削除
*     last_date (最終観察日)が欠損の症例を削除
*****
**** 1) AML, ALL, CMLの中からCMLを削除 ****
tab dx, missing
tab diagnosis, missing
tab diagnosis, missing nolabel
/*2x2 tableを作る*/
tab dx diagnosis, missing
/*AMLとALLのみをkeep*/
* if以下の"等しい"は、"=="
* 文字列の値は、""で囲む。数値は囲まない
* & (and かつ), | (or または), == (equal 等しい), != (not equal 等しくない)
keep if diagnosis==1 | diagnosis==2
* drop if diagnosis==4 でも同じ結果が得られます
count

**** 2) survival dataとlast_dateが欠損の症例を削除 ****
tab survival, missing
tab last_date, missing
drop if survival==" | last_date=="
* keep if survival!=" | last_date!="
count

/**** dataの加工 ****/
*****
* 性別 dataの作成
* 「男」「女」のdataでは後に多変量解析ができないため、数値による変数を作成する必要があります
*
* STEP
* 1) 「男」「女」というデータから、女性が0 男性が1の新たな変数を作り
* 2) それに"Female" "Male"のラベルを付けます

```

* 3) 作成した変数の確認

**** 1) sex の「男」「女」というデータから、女性が0 男性が1 の新たな変数"pt_sex"を作成

```
tab sex, missing
```

```
gen pt_sex = 0 if sex=="女"
```

*変数 "pt_sex" を作成。sex=="女" の症例を0とする

```
replace pt_sex = 1 if sex=="男"
```

*Data Editor を確認。新たな変数 "pt_sex"が作成されている

** 項目にラベルをつけることができる（公開スクリプト 01 を参照）

```
** label var pt_sex "患者性別"
```

**** 2) コードの意味を忘れてしまうこともあるが、以下のラベルが便利

```
label define sex_label 0 "Female" 1 "Male"
```

```
label values pt_sex sex_label
```

*Data editor を確認。pt_sex の項目が Female, Male とラベルされています

**** 3) 新しく作成した変数を確認

```
*tab sex, missing
```

```
tab pt_sex, missing
```

```
tab pt_sex, missing nolabel
```

```
tab sex pt_sex, missing
```

* HLA mismatch あり1, なし0 の data を作成

* STEP

* 1) mismatch locus 数:0個を mismatch=0

* mismatch locus 数: 1個以上が mismatch=1 の変数を作成

* 2) それに"match" "mismatch"のラベルを付けます

* 3) 作成した変数の確認（終了して、余裕のある方は次の年齢の categorical data 作成へ。）

```
tab seromis6abdr2, missing
```

```
generate mismatch=0 if seromis6abdr2==0
```

```
replace mismatch=1 if seromis6abdr2==1 | seromis6abdr2==2
```

** 欠損値がある場合は、欠損値のカテゴリーも作る。

```
** replace mismatch=99 if seromis6abdr2==.
```

```
label define mismatch_label 0 "match" 1 "mismatch"
```

```
label values mismatch mismatch_label
```

```
tab mismatch seromis6abdr2, missing
tab mismatch seromis6abdr2, missing nolabel
```

```
*****
```

- * 年齢の categorical data の作成
- * 0-9 歳、10-19 歳、20-29 歳の変数を作成
- * ラベルをつける

```
*****
```

- * summarize age
- * sum age, detail
- * gen agecat=1 if age>=0 & age<10
- * replace agecat=2 if age>=10 & age<20
- * replace agecat=3 if age>=20 & age<30
- * *Stata の場合欠損値は “ large positive values ” として認識。使用する統計ソフトによって設定が異なる。
- * label define agecat_label 1 "0-9" 2 "10-19" 3 "20-30"
- * label values agecat agecat_label
- * tab agecat, missing
- * tab agecat, missing nolabel

```
*****
```

- * 日付と日数の計算
- * STEP
- * 1) stata 日付データの作成
- * 2) 移植日から移植年の作成
- * 3) 移植年カテゴリーの作成

```
*****
```

*Excel 1900 年 1 月 1 日がシリアル値「 1 」

*Stata 1960 年 1 月 1 日がシリアル値「 1 」

*Excel を Stata にデータを取り込んだ場合には、自動的に Stata 日付データに変換されるため、下記のコマンドは不要になります。

** 1960/1/1 を「1」とした 5 桁の変数「 tdata 」を作成します。

** 10/20/2013 と入力されたデータの場合は"MDY"と指定する必要があります。

```
generate tdate= date(tx_date, "YMD")
```

```
gen tyear=year(tdate)
```

*5 桁の Stata 日付データから「年」が抽出され、新たな変数 tyear を作成

```
tab tyear
```

3) 移植年の categorical data を作る [宿題]

```
gen tyear_cat = 1 if tyear>=2000 & tyear<=2005
```

*あるいは gen tyear_cat = 1 if tyear==2001 | tyear==2002 | tyear==2003 | tyear==2004 | tyear==2005

```
replace tyear_cat = 2 if tyear>=2006 & tyear<=2010
```

```

label var tyear_cat "移植年 category"
label define tyear_cat_label 1 "2000-2005" 2 "2006-2010"
label values tyear_cat tyear_cat_label

tab tyear_cat tyear, missing
*tab tyear_cat tyear, missing nolabel

/**** 背景表(Table 1)の作成 ****/
tab mismatch, missing
sum age if mismatch==0, detail
sum age if mismatch==1, detail
ranksum age, by (mismatch) /*Wilcoxon Rank Sum Test*/

tab pt_sex mismatch, chi2 colum
tab diagnosis mismatch, chi2 colum
** tab agecat mismatch, chi2
tab tyear_cat mismatch, chi2 colum

/**** Survival data の作成 ****/
generate ldate=date(last_date, "YMD") /*Stata 日付データ、5ケタの数字*/
generate lday=ldate-tdate
sum lday, detail /*データの確認、生存日が移植日より前のデータなどないか?など。詳細はTRUMPセッションで。*/

/*生死データ*/
tab survival, missing nolabel
gen event_os =1 if survival == "死亡"
replace event_os =0 if survival == "生存"
*tab survival event_os, missing

stset lday, failure (event_os==1)
sts graph
sts graph if diagnosis==1, by (mismatch) censored(multiple)

generate lyear=lday/365.25
stset lyear, failure (event_os==1)
sts graph, by (mismatch) title (Overall Survival) censored (multiple) /*
*/ ytitle (Probability) xtitle (Years after transplant) xlabel (0 (2) 12)

** Log rank 検定 **
sts test mismatch

```

** 生存解析の結果の表示（点推定値）**

sts list

sts list, by (mismatch)

sts list, by (mismatch) at (1, 3, 5)

模擬プロトコールに基づいたデータマニピュレーション

- ・ **統計解析（単変量解析）の実際**

H25年度 TRUMP統計セミナー
基礎編

模擬プロトコルに基づいた
データマニピュレーション・統計解析の実際

担当: 熱田由子

変数の決定と変数リストの作成

- 変数の種類
 - 患者背景変数
 - 患者年齢・性別
 - 疾患情報
 - 移植関連情報
 - アウトカム変数
 - イベント変数
 - 時間変数
- 研究計画が定まったら、解析に必要な変数リストを作成する

例1: ABO mismatch

データを覗いてみると...

変数の作成

エクセル上で作業をしない！！

- 150-200例未満程度の場合は目視・手入力でも出来るかもしれないが...
- ソートを繰り返していたらデータが無茶苦茶...
- 気付いたらどこかで一行ずれてしまっていた...
- 変数作成のアルゴリズムを忘れてしまうことも...

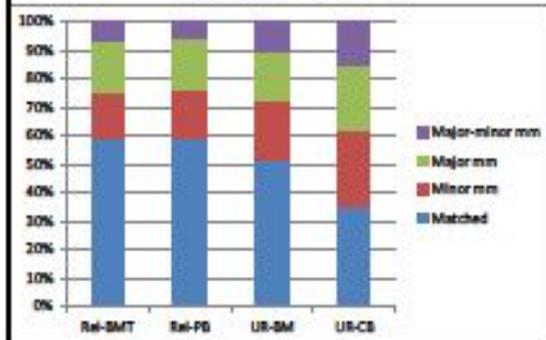
「.....」だらけ

家数一覧(基本家数リスト)
(Stata / EZR共有スクリプト対応)
ダウンロード可能!!

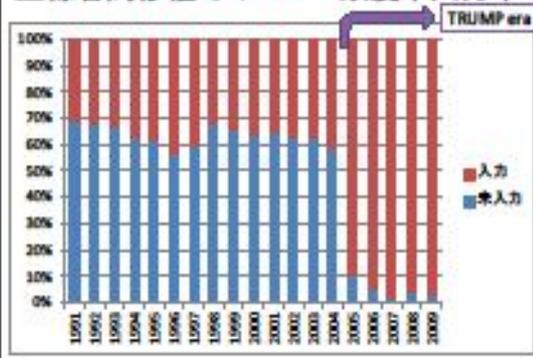
変数作成

- ABO matching
 - Matched
 - Minor mismatch
 - Major mismatch
 - Major-minor mismatch
- ABO major mismatch
 - (0) Matched or minor mismatch
 - (1) Major (major-minor) mismatch

作成変数の確認



血縁者間移植でのABO一致度未入力率



どうしよう?

- 研究課題
- ①. 「血縁者間移植においてABO一致度が生存に与える影響」
- ②. 「血縁者間移植におけるBMとPBの比較」

研究①「ABO一致度の影響」の場合

- 2005年以降の移植に限定した解析を行う
 - 観察期間が短い
- 入力があるものを採用する。その場合、2004年以前に関しては対象に入るのは3-4割程度。
 - 施設でのTRUMP以降の自主的入力
 - 未入力の場合と大きな差が(なんらかのバイアスが)ないか確認しておく必要がある。
- 2004年以前移植に関しては、入力率が7割以上の施設のデータのみを対象

TRUMP 統計解析セミナー（基礎編）出席者名簿

1	新井 康之	京都大学大学院医学研究科
2	青木 淳	駒込病院
3	青木 智広	名古屋第二赤十字病院
4	有馬 靖佳	田附興風会医学研究所北野病院
5	伊豆津 宏二	虎の門病院
6	井上 政弥	島根大学医学部附属病院
7	植木 俊充	長野赤十字病院
8	梅田 雄嗣	京都大学医学部附属病院
9	緒方 正男	大分大学医学部
10	冲中 敬二	国立がん研究センター中央病院
11	籠尾 壽哉	国立病院機構東京医療センター
12	加藤 格	京都大学医学部附属病院
13	加藤 剛二	名古屋第一赤十字病院
14	加藤 実穂	名古屋第一赤十字病院
15	亀井 美智	名古屋市立大学
16	鴨下 園子	名古屋赤十字病院
17	河北 敏郎	国立病院機構熊本医療センター
18	川島 直実	名古屋第一赤十字病院
19	川瀬 有美	京都大学医学部附属病院
20	木田 理子	NTT 東日本関東病院
21	栗田 尚樹	筑波大学
22	近藤 健	北海道大学病院
23	佐々木 裕哉	小倉記念病院
24	住 昌彦	長野赤十字病院
25	杉盛 千春	石川県立中央病院
26	瀬戸 愛花	岡崎市民病院
27	妹尾 紀子	信州大学医学部附属病院
28	高野 久仁子	大分大学医学部附属病院
29	竹田 淳恵	神戸医療センター中央市民病院
30	千原 大	愛知県がんセンター研究所
31	中田 佳世	大阪府立成人病センター
32	長谷川大一郎	兵庫県立こども病院
33	林 邦雄	明和病院
34	半田 寛	群馬大学医学部附属病院
35	平瀬 敏志	神戸大学医学部附属病院
36	深野 玲司	九州がんセンター
37	藤盛 好啓	兵庫医科大学
38	藤原 弘	愛媛大学病院
39	増成 太郎	中国中央病院

40	町田 真一郎	東海大学医学部附属病院
41	水谷 元紀	愛知医科大学医学部
42	三橋 健次郎	東京女子医科大学
43	宮尾 康太郎	安城厚生病院
44	宮村 能子	大阪大学医学部附属病院
45	森 健	神戸大学医学部附属病院
46	森下 喬允	名古屋大学大学院医学系研究科
47	矢内 友子	兵庫県立こども病院
48	横畠 絵美	名古屋赤十字病院
49	吉満 誠	鹿児島大学病院
50	渡邊 健一郎	京都大学大学研究科

TRUMP 統計セミナー資料集 (応用編)

*変数表は省略する

厚生労働科学研究費補助金難治性疾患等克服研究事業(免疫アレルギー疾患等予防・治療研究事業 移植医療研究分野)

「本邦における造血細胞移植一元化登録研究システムの確立」

平成 25 年度 TRUMP 統計セミナー：応用編

日時： 平成 25 年 11 月 23 日(土) 10:00~15:30

場所： 名古屋大学大学院医学系研究科・医学部保健学科 東館 4F 大講堂
〒461-8673 名古屋市東区大幸南一丁目 1 番 20 号

プログラム

9:30~ 受付、資料配布

10:00-10:30

I. 我流 Stata の使い方：自分なりの方法を確立しよう

(名古屋大学大学院医学系研究科 造血細胞移植情報管理・生物統計学
准教授 鈴木 律朗)

10:30-12:30

II. 多変量解析に臨む前に知っておくこと

(九州大学大学院医学研究院 予防医学分野 教授 松尾 恵太郎)

12:30-13:30

昼食

13:30-14:30

III. 競合リスクイベントの扱いの解説と演習

(名古屋大学大学院医学系研究科 造血細胞移植情報管理・生物統計学 熱田 由子)

14:30-15:30

IV. 時間依存性変数の扱いの解説と演習

(自治医科大学附属さいたま医療センター 血液科 諫田 淳也)

我流 Stata の使い方 : 自分なりの方法を確立しよう

前回のおさらい

- `cii` confidence interval immediately
- `tabi` tabulate immediately
- `csi` chi square immediately
- `cci` case control immediately
- `iri` incidence rate immediately
- `sampsi` sample size

信頼区間

急性GVHDの治療を12例にしたら、7例で有効であった。
発表に使うので、有効率の95%信頼区間を求めたい。

以下のコマンドをタイプ。

`cii 12 7`

```

cii 12 7
-----
|         |   %   |   %   | 95% CI | 90% CI |
|-----|-----|-----|-----|-----|
|         | 58.3  | 58.3  | 27.7-84.8 | 31.5-81.9 |

```

58.3% (95% CI: 27.7-84.8%)

90%信頼区間

以下のコマンドをタイプ。

`cii 12 7, level(90)`

```

cii 12 7, level(90)
-----
|         |   %   |   %   | 95% CI | 90% CI |
|-----|-----|-----|-----|-----|
|         | 58.3  | 58.3  | 27.7-84.8 | 31.5-81.9 |

```

58.3% (95% CI: 27.7-84.8%)

58.3% (90% CI: 31.5-81.9%)

helpコマンドと、Stataのマニュアル

helpコマンドと、Stataのマニュアル

Stataには、もともとマニュアルの本が附随して
いました。(有料)

これらは、全部で17冊あります。(ver. 11)



内容は、helpコマンドで参照できます。

最新の ver. 13 では、紙のマニュアル本は廃止されました。



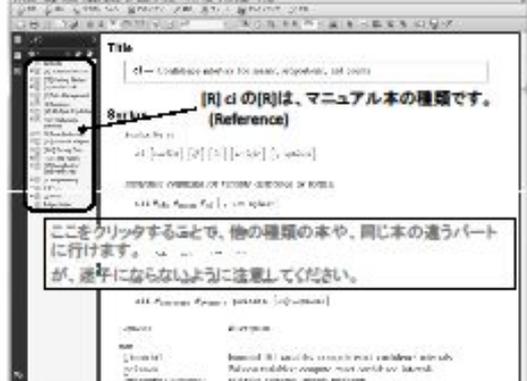
元々、この本は、Stataのインストールパッケージに含まれていました。

help コマンド

help + コマンド名で、コマンドの使い方が参照できます。
 help ci と、タイプしてみてください。




ci は、ci コマンドの immediate オプション。
 同じ ci コマンドでも、使い分けがある。
 使い分けの方法
 下の方に、Example (用例) もありますが...
 [R] をクリックしてください。



[R] ci の [R] は、マニュアル本の種類です。
 (Reference)
 ここをクリックすることで、他の種類の本や、同じ本の違うパートに行けます。
 が、誤字にならないように注意してください。

PDF版のhelp (または本)の方が、結果表示があるので分かりやすいです。

Gridview confidence intervals

Example 1

Gridview confidence intervals for survival, logrank, and other distributions using the confidence interval of the logrank test.

コマンド行 →

```
ci [varname] [varname] [varname] [varname] [varname]
```

結果表示

Variable	HR	95% CI	95% CI	95% CI	95% CI
Age	1.05	1.01	1.09	1.01	1.09
Sex	0.98	0.93	1.03	0.93	1.03

コマンド行 →

```
ci [varname] [varname] [varname] [varname] [varname]
```

結果表示

Variable	HR	95% CI	95% CI	95% CI	95% CI
Age	1.05	1.01	1.09	1.01	1.09
Sex	0.98	0.93	1.03	0.93	1.03

Immediate form

Example 7

Gridview confidence intervals for survival, logrank, and other distributions using the confidence interval of the logrank test.

Variable	HR	95% CI	95% CI	95% CI	95% CI
Age	1.05	1.01	1.09	1.01	1.09
Sex	0.98	0.93	1.03	0.93	1.03

Example 8

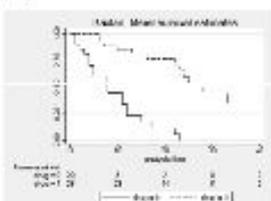
Gridview confidence intervals for survival, logrank, and other distributions using the confidence interval of the logrank test.

Variable	HR	95% CI	95% CI	95% CI	95% CI
Age	1.05	1.01	1.09	1.01	1.09
Sex	0.98	0.93	1.03	0.93	1.03

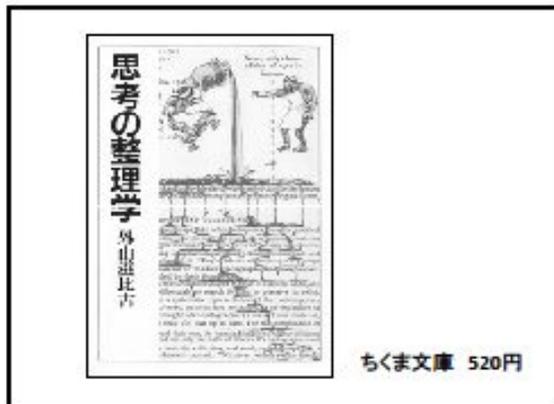
sts graph のようなグラフコマンドだと、PDFでは図が入る。

自分が描きたい図を、目で見て探せるので便利。

sts graph, hybrid, ci, risktable



risktable オプションで、at risk 数を追加できる。



今の社会は、つよい学校信仰ともいうべきものを持っている。

学校はグライダー人間の訓練所である。飛行機人間は作らない。グライダーの練習に、エンジンのついた飛行機などがまじっていては迷惑する。危険だ。学校では、ひっぱられるままに、どこへでもついて行く従順さが尊重される。勝手に飛び上がったりますのは規律違反。たちまちチェックされる。やがてそれぞれに、グライダーらしくなって卒業する。

グライダーとして一流である学生が、卒業間際になって論文を書くことになる。これは、これまでの勉強といささか勝手が違う。(中略)グライダーは途方にくれる。突如としてこれまでと違うことを要求されてもできるわけがない。グライダーとして優秀な学生ほどあわてる。

一般に、学校教育を受けた期間が長ければ長いほど、自力飛行の能力は低下する。グライダーとして飛べるのに、危ない飛行機になりたくないのは当たり前である。



多変量解析に臨む前に知っておくこと



本日の話の大まかな流れ

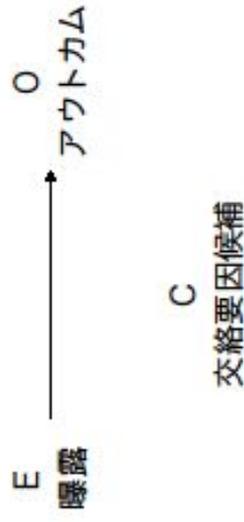
- 解析前の流れ
 - おまけ
- 生存解析
 - モデルの検証
- 多変量解析の仕組
 - 統計パッケージで解析する際の基本原則
- 第3の要因をどう取り扱うか？
 - α 、 β エラー



大まかな流れ

- 研究で答える疑問をただ一つ設定
 - 対象者 (Population)
 - 主たる検討要因 (Exposure)
 - アウトカム (Outcome)
 - その他の関連する要因 (交絡、交互作用の可能性のある要因 Covariates or Confounder)
- どの研究デザインを採用するか？

E, O, and C

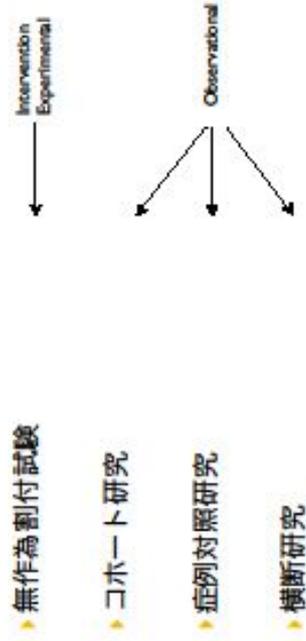


この構図に落とし込めてない状態では解析しない

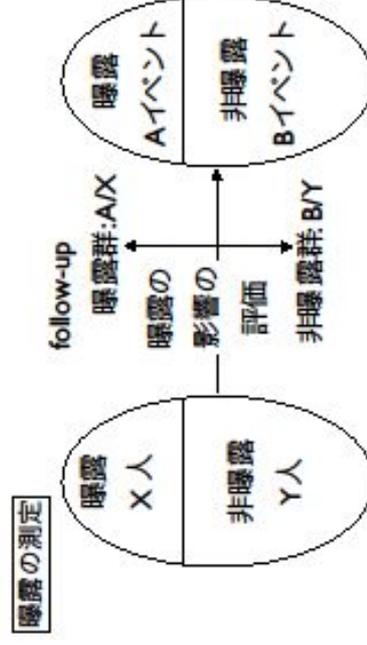
疑問の設定時

- TRUMPのようなデータでは、色々なことが出来てしまいます
- 失敗する研究は何がしたいかが絞れていない研究です

4つの代表的な研究デザイン

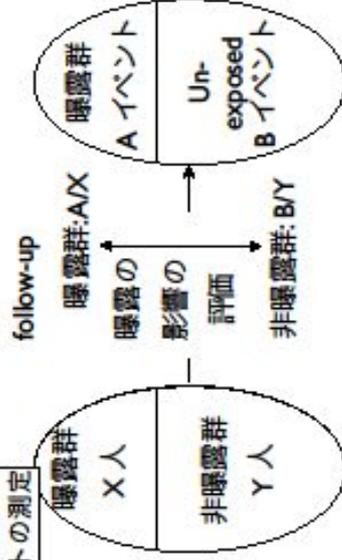


前向きコホート研究



後向きコホート研究

診療録レビュー等
・曝露の測定
・イベントの測定



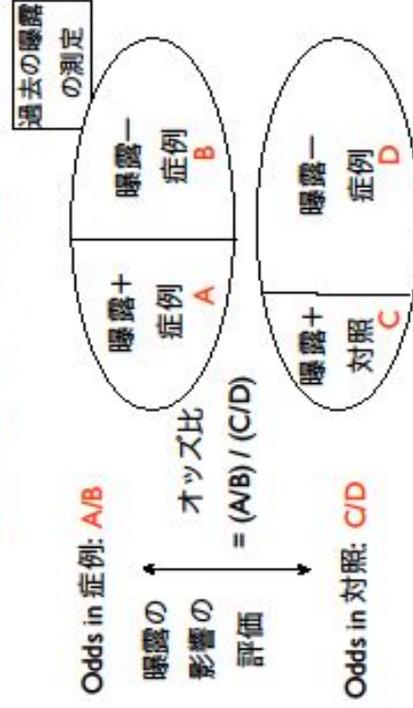
9

コホート研究 (続き)

- Pro
 - イベントがおきる前に曝露要因を測定している、つまり曝露要因に関するエラーが少ない (前向き)
 - 情報の扱い方などにエラーを起こす可能性がある (後向き)
- Con
 - イベント発生までに時間がかかる (前向きの場合)
 - 希なイベントは評価しづらい
 - よく使う解析モデルは比例ハザードモデル
 - TRUMPデータでやる生存解析は前向きコホート?後向きコホート?

10

症例対照研究



11

Case-control study (cont')

- Pro
 - 前向き、後向きコホートよりも時間がかからない
- Con
 - 曝露測定にバイアスが入る可能性
 - 曝露要因次第 (遺伝子とかの場合関係ない)
 - 対照群設定の難しさ
- よく使う解析モデルはロジスティック回帰分析 (マッチング有りの時は条件付きロジスティック回帰分析)

12