

医学統計セミナー アドバンスコース
臨床試験における症例数設定とガイドライン

下川 敏雄

和歌山県立医科大学 臨床研究センター

2016年度 医学統計セミナー

■ **ベーシック・コース**

- **基礎統計学** (6月15日・住金棟5F 大研修室)
- **量的データの解析** (7月27日・住金棟5F 大研修室)
- **質的データの解析** (8月24日・住金棟5F 大研修室)
- **共変量調整を伴う解析** (11月2日・病院棟4F 臨床講堂1)
- **生存時間・臨床検査データの解析**(11月16日・住金棟5F 大研修室)

■ **アドバンス・コース**

- **多群・経時データの解析と多重比較**
(11月30日・病院棟4F 臨床講堂1)
- **臨床試験における症例数設定とガイドライン**
(12月28日・住金棟5F 大研修室)
- **アンケート調査データの解析** (2月1日・病院棟4F 臨床講堂1)
- **統計的因果推論と傾向スコア** (2月22日・住金棟5F 大研修室)
- **メタアナリシス** (3月22日・病院棟4F 臨床講堂1)

症例数設定の概要

症例数とp値の関係

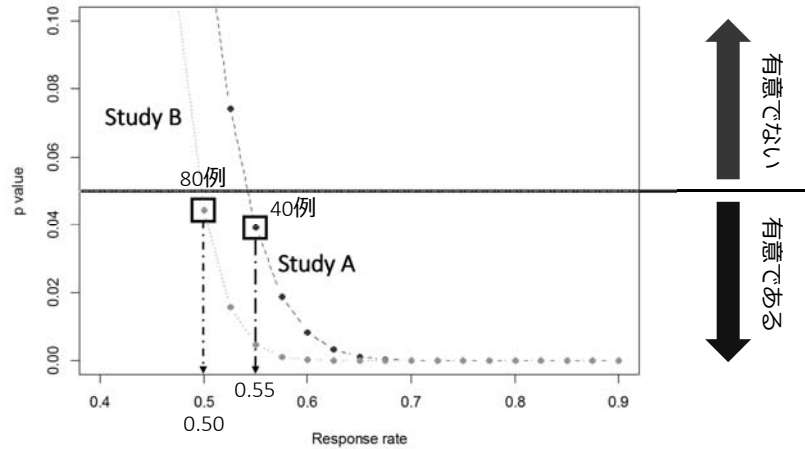
胃がんに対する二つの臨床研究があったとする。これらは、同一対象患者に対して、同一レジメンの術後補助化学療法が実施された。なお、これまでの当該患者に対する奏効率は40%だった。

Study A	Study B
症例数：40例, 奏効例：20例 奏効割合： $\hat{p} = 0.5$ (50%)	症例数：80例, 奏効例：40例 奏効割合： $\hat{p} = 0.5$ (50%)
p 値(閾値0.4)でのp値 = 0.130 (n.s.)	p 値(閾値0.4)でのp値 = 0.045*

全く同じ奏効割合(0.50)であったとしても、症例数の違いから、Study Aではnegative studyとなり、Study Bではpositive studyとなる。

Study AがPositive studyになるにはどのような状況である必要があるか？

Study AとStudy Bで有意になる奏効割合



80例の場合には、奏効割合が0.5 (50%)以上の奏効割合で有意になるが、40例の場合には0.55 (55%)にならないと有意にならない。

臨床試験では、新薬(治療)で期待される効果量を予め規定し、試験の結果、期待される効果が得られたときに、有意差が得られるような最小症例数を計算する。

症例数設定の一例

論文での記述例

Phase II study of trastuzumab in combination with S-1 plus cisplatin in HER2-positive gastric cancer (HERBIS-1)

The required sample size was estimated based on a threshold RR of 35% and an expected RR of 50%, 80% power, and an alpha value of 0.1 (one-sided) using the binomial test. Given 2% of ineligible patients, the target sample size was determined to be at least 50 patients.

(Kurokawa et al., British Journal of Cancer, 110, 1163-68, 2014)

■ 上記記載の要約

試験デザイン：Phase II Study (単アーム)

主要評価項目：奏効割合 (RR: Response Rate)

閾値奏効率：35%， 期待奏効率：50%

臨床的判断

第1種の過誤(type I error, 有意水準)： $\alpha = 0.1$

1 - 第2種の過誤(1-type II error, 検出力)： $1 - \beta = 0.8$ (80%)

統計的判断

必要症例数 = 50

第1種の過誤(α エラー)・第2種の過誤(β エラー)とは

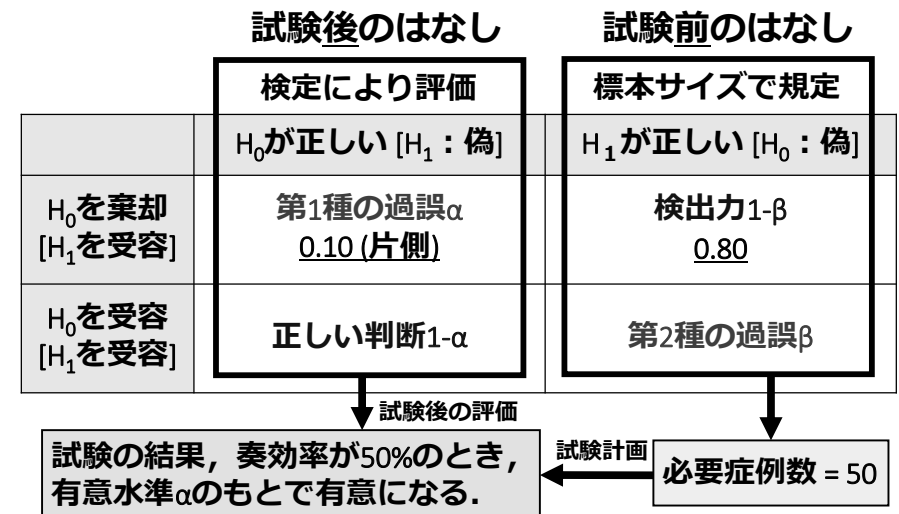
	検定により評価	標本サイズで規定
	H_0 が正しい [H_1 : 偽]	H_1 が正しい [H_0 : 偽]
H_0 を棄却 [H_1 を受容]	第1種の過誤 α (H_0 が正しいのに棄却)	検出力 $1 - \beta$ (H_0 が正しいので受容)
H_0 を受容 [H_1 を受容]	正しい判断 $1 - \alpha$ (H_1 が正しいのに棄却)	第2種の過誤 β (H_1 が正しいのに受容)

■ 第2種の過誤が一定水準未満(β 未満)になるように、標本サイズを規定する (本当はpositiveなのに本試験を通じてnegativeと判断する可能性を低くする)。

■ 統計解析の結果、第1種の過誤が一定水準未満(α 未満)であるか否かを確認する (本当はnegativeなのに本試験結果を通じてpositiveと結論付けるエラーがかなり低いことを確認する)。

先ほどの事例に当てはめる

H_0 ：奏効割合は35%である。 H_1 ：奏効割合は35%を上回る。



試験デザインでの留意点

■ 閾値(帰無仮説)には根拠が必要

単群試験では、閾値に対して臨床的な根拠が必要であり、統計解析は閾値よりもポジティブな結果が否かが評価される。ランダム化比較試験では、帰無仮説は「差がない(or HR=1)」になる。

HERBIS-1では、ToGA試験(Y-J Bang, et al. *The Lancet*, 28, 687 - 697, 2010.)の結果を参考している。ToGA試験では、化学療法群(5-FU /CDDP あるいはCapecitabine /CDDP)での奏効率が35%であった。

■ 期待値(対立仮説)には実現可能性及び臨床的有効性に基づいて設定

期待値は、臨床的に合意されるような有効性(エフェクトサイズ)として設定される。したがって、期待値が本臨床試験のエンドポイントの目標値となる。

HERBIS-1では、ToGA試験では、(1) Trastuzumab併用群の奏効率が47%であること、(2) HER2による選択基準(IHC 3+ or FISH positive + IHC 2+)、(3) 本邦の状況を鑑みて、閾値奏効率を50%と設定している。

必要症例数設定のプロセス

STEP.1 : 主要評価項目(主たるアウトカム)を考える

- ・本研究におけるクリニカルクエスションに応えられるものか(妥当性)?
- ・当該分野においてAgreeされるものか(信頼性)?
- ・エンドポイントを取得できるか(実施可能性)?
 - －例えば、抗がん剤治療において標的病変がなければ奏効率が測定できない等

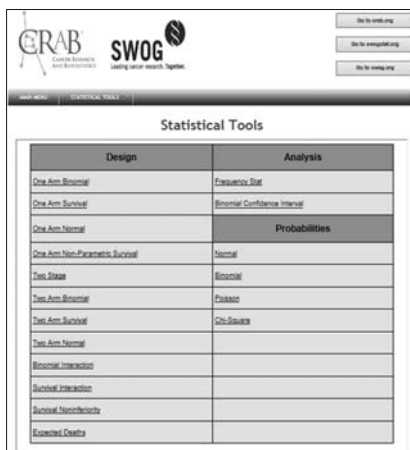
STEP.2 : 仮説を決定する

- ・無作為化比較試験の場合には非劣勢・優越性・同等性を検討する。
- ・単アーム試験の場合には閾値を文献等から決定する。
- ・本プロトコル治療における主要評価項目での有用性を検討する。
 - 量的変数の場合：共通の分散(標準偏差)を決定する(個体差)。
 - 生存時間の場合：受け入れ期間、フォローアップ期間を決定(censoring割合)

STEP.3 : 有意水準・検出力を決定する

単アーム： $\alpha=0.05$ or 0.10 , $1-\beta = 0.8$
 多アーム： $\alpha=0.05$, $1-\beta = 0.80$ or 0.90

必要症例数の計算 : WEB SITE (1/2)



<https://stattools.crab.org/>

単アーム試験

- －母比率 (One Arm Binomial)
- －母平均 (One Arm Normal)
- －生存時間 (One Arm Survival)
- (One Arm Non-parametric / KM推定に基づく)
- －2段階デザイン(母比率のみ, Two Stage)

2アーム試験

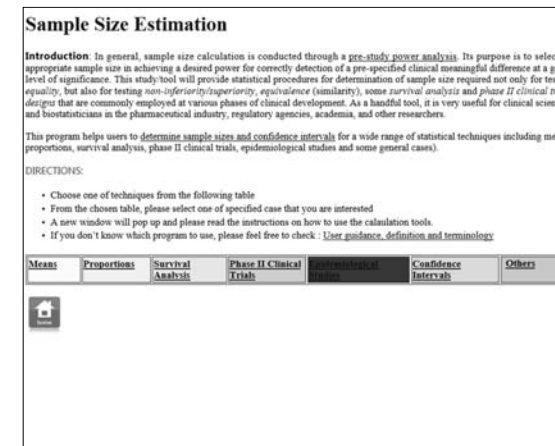
- －母比率 (Two Arm Binomial)
- －母平均 (Two Arm Normal)
- －生存時間 (One Arm Survival)

その他の試験

- －母比率に対する交互作用 (Two Arm Binomial)
- －生存時間に対する交互作用 (Two Arm Normal)
- －生存時間での非劣勢 (Survival Noninferiority)
- －期待死亡数 (Expected Deaths)

必要症例数の計算 : WEB SITE (2/2)

The Chinese University of Hong Kong



<https://www2.ccrb.cuhk.edu.hk/stat/>

あまり知られていないか、比較的豊富な試験デザインが含まれている。

必要症例数の計算：フリーソフトウェア



統計ソフトウェアR(統計学分野のデファクトスタンダード)をR言語が使えない利用者のために修正した統計ツール

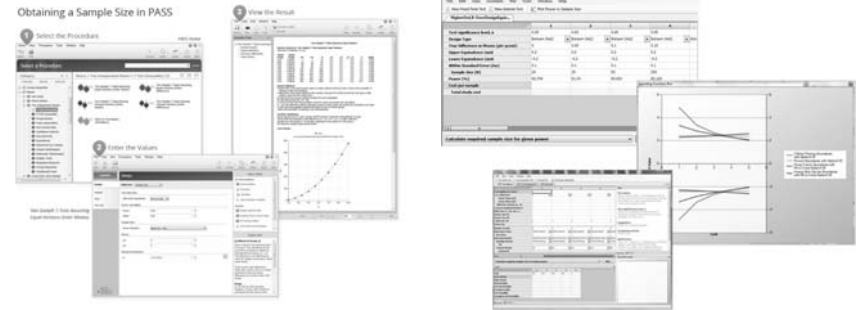
http://www.jjichi.ac.jp/saitama-sct/SaitamaHP_files/statmed.html



ヴァンダービルド大学が開発した標本サイズを計算するためのソフトウェア

<http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize>

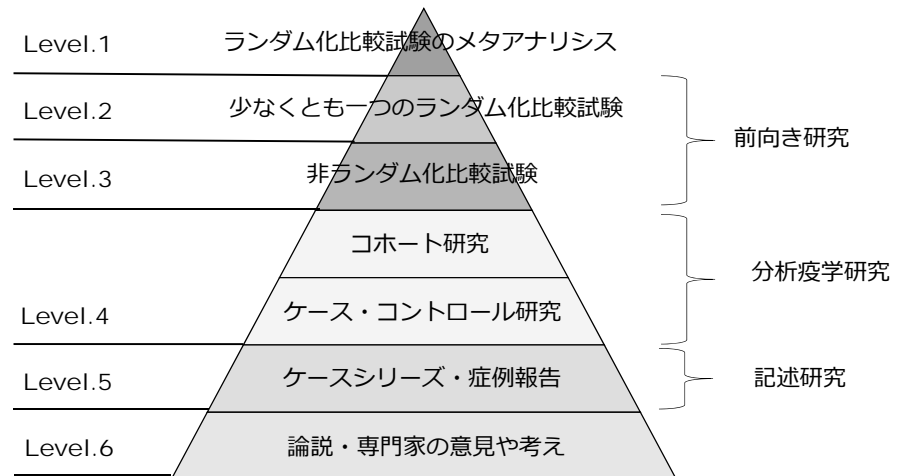
有償の標本サイズを計算するだけのソフトウェア



ガイドラインについて

研究デザインとエビデンス

治療に関する論文のエビデンスレベル (AHRQ:米国医療政策研究局)



ランダム化比較試験を行う上でのガイドライン：CONSORT声明

<http://www.consort-statement.org/>

CONSORT(Consolidated Standards of Reporting Trials：臨床試験報告に関する統合基準)とは、無作為化比較試験の報告の質の向上について記載された基準であり、論文とともにチェックリストを提出することが義務付けられるようになってきている。最新版は「CONSORT2010」

CONSORT2010 Check List

無作為化比較試験を報告する際に記載すべき情報のチェックリストになっている。本チェックリストを埋めるには、適切なプロトコルを作成することが必須。

CONSORT Flow Diagram

被験者の登録・割付・追跡・解析に至るすべての状況において除外された症例とその理由等を記載するためのダイアグラムである。臨床試験の適切な運用が求められる。

CONSORT 2010 checklistの内容(1/4)

タイトル・抄録 (Title and Abstract)

	1a	タイトルにランダム化比較試験であることを記載
	1b	試験デザイン(trial design), 方法(method), 結果(result), 結論(conclusion)の構造化抄録

はじめに (Introduction)

背景・目的 (Background and Objective)	2a	科学的背景と論拠 (rationale)の説明
	2b	特定の目的または仮説 (hypothesis)

方法 (Method)

試験デザイン (Trial Design)	3a	試験デザインの記述 (並行群間, 要因分析など), 割付け比を含む
	3b	試験開始後の方法上の重要な変更(適格基準 eligibility criteriaなど)とその理由
参加者 (Participant)	4a	参加者の適格基準 (eligibility criteria)
	4b	データが収集されたセッティング(setting)と場所
介入(Intervention)	5	再現可能となるような詳細な各群の介入。実際にいつどのように実施されたかを含む

CONSORT 2010 checklistの内容(2/4)

アウトカム (Outcome)	6a	事前に特定され明確に定義された主要・副次的アウトカム評価項目。いつどのように評価されたかを含む。
	6b	試験開始後のアウトカムの変更とその理由。
症例数 (Sample size)	7a	どのように目標症例数が決められたか。
	7b	あてはまる場合には、中間解析と中止基準の説明。

ランダム化 (Randomization)

順番の作成 (Sequence generation)	8a	割振り(allocation)順番を作成(generate)した方法
	8b	割振りのタイプ: 制限の詳細(ブロック化, ブロックサイズなど)
割振りの隠蔽機構 (Allocation concealment mechanism)	9	ランダム割振り順番の実施に用いられた機構(番号付き容器など), 各群の割付けが終了するまで割振り順番が隠蔽されていたかどうかの記述
実施 (Implementation)	10	誰が割振り順番を作成したか, 誰が参加者を組入れ(enrollment)たか, 誰が参加者を各群に割付けた(assign)か
ブラインディング (Blinding)	11a	ブラインド化されていた場合, 介入に割付け後, 誰がどのようにブラインドかされていたか(参加者, 介入実施者, アウトカムの評価者など)
	11b	関連する場合, 介入の類似性の記述
統計学的手法 (Statistical method)	12a	主要・副次的アウトカムの群間比較に用いられた統計学的手法
	12b	サブグループ解析や調整解析のような追加的解析の手法

CONSORT 2010 checklistの内容(3/4)

結果 (Results)

参加者の流れ (Participant flow) (フローチャートを強く推奨)	13a	各群について, ランダム割付けられた人数, 意図された治療を受けた人数, 主要アウトカムの解析に用いられた人数の記述
	13b	各群について, 追跡不能例とランダム化後の除外例を理由とともに記述
募集(Recruitment)	14a	参加者の募集期間と追跡期間を特定する日付
	14b	試験が終了または中止した理由
ベースライン・データ (Baseline data)	15	各群のベースラインにおける人口統計学的(demographic), 臨床的な特性を示す表
解析された人数 (Number analyzed)	16	各群について, 各解析における参加者数(分母), 解析が元の割付け群によるものであるか
アウトカムと推定 (Outcome and estimation)	17a	主要・副次的アウトカムのそれぞれについて, 各群の結果, 介入のエフェクト・サイズの推定とその精度(95%信頼区間など)
	17b	2項アウトカムについては, 絶対エフェクト・サイズと相対エフェクト・サイズの両方を記載することが推奨される
補助的解析 (Ancillary analysis)	18	サブグループ解析や調整解析を含む, 実施した他の解析の結果。事前に特定された解析と探索的解析を区別する
害 (Harm)	19	各群のすべての重要な害(harm)または意図しない効果

CONSORT 2010 checklistの内容(4/4)

考察 (Discussion)		
限界(Limitation)	20	試験の限界, 可能性のあるバイアスや精度低下の原因, 関連する場合は解析の多重性の原因を記載
一般化可能性 (Generalisability)	21	試験結果の一般化可能性 (外的妥当性, 適用性)
解釈(Interpretation)	22	結果の解釈, 有益性と有害性のバランス, 他の関連するエビデンス
その他の情報 (Other information)		
登録(Registration)	23	登録番号と試験登録名
プロトコール (Protocol)	24	可能であれば, 完全なプロトコールの入手方法
資金提供者 (Funding)	25	資金提供者と他の支援者 (薬剤の供給者など), 資金提供者の役割

統合倫理指針を遵守し, きっちりとしたプロトコールを作成した上で臨床試験を実施すれば, 難しいところは何もない. 他方, 不十分なプロトコール上での臨床試験はCONSORT声明非遵守となり, 臨床試験の質を問われる恐れがある.

UMINにおける介入研究デザイン: 基本デザイン

■ 並行群間比較

被験者が, 2つ以上の複数の群の内の1つに割付けられる試験

■ クロスオーバー試験

すべての被験者は, 同一の2つ以上の複数の介入を受けるが, 介入を受ける順序が異なる. 介入を受ける順序の違いで複数の群を規定し, この内1つの群に被験者を割付ける試験.

■ 要因デザイン

複数の介入の, 異なる組み合わせを複数用いて, 2つ以上の介入を同時に評価する試験.

■ 単群

すべての被験者が同一の介入を受ける試験

■ 継続・拡大投与

倫理的・救済的措置として, 本試験の終了後にも引き続き介入が実施されたり, 本試験に参加できなかったが試験薬の投与を希望する被験者に試験薬を投与するために実施される試験.

割付の方法

割付方法	
静的割付 (登録前に予め割付コード表を規定する)	単純無作為化法
	ブロック無作為化(置換ブロック法)
	層別無作為化法
動的割付 (層別因子毎に登録のたびに割付を行う)	最小化法 拡張最小化法(層別因子に連続量を追加)

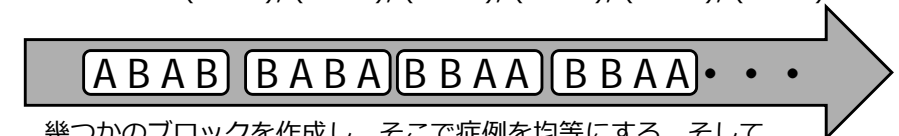
- 静的割付の3種類にはランダム化が含まれているため, 確率的な割付が行われるが, その割付内容は, 試験開始前に決定される.
- 当初の最小化法では, 層別因子により決定論的に決定される. ICH-E9では, 完全な決定論的な割付に対して, 盲検性を保障するための十分な注意が喚起されていることから, 現在ではランダム化を伴う割付が実施されている.
- 欧州ではブロック無作為化法, 本邦では最小化法が一般的である. 最小化法では, ソフトウェア技術及びバリデーションが重要である.

静的割付の方法(1): ブロック無作為化法

割付の例示

群: 2群(治療群A, 治療群B)
層別因子: 層別因子(年齢(≤65, >65), PS(0,1,2))

A群とB群が平等になる順列組み合わせ(ブロック数4):
(AABB), (ABAB), (ABBA), (BAAB), (BABA), (BBAA)



幾つかのブロックを作成し, そこで症例を均等にする. そして, ブロックの並びをランダムに決定する.

症例番号

- 予測不能であるため, 予測可能性を回避できる.
- 試験規模が小さくても不均衡は改善される.
- 層別因子を含められないため, 群ごとに背景因子が不均衡になる.

静的割付の方法(2)：層別無作為化法

割付の例示

群：2群(治療群A, 治療群B)
層別因子：層別因子(年齢($\leq 65, > 65$), PS(0,1,2))



層別因子の組み合わせ毎に無作為化が行われる(ブロック無作為化)

- 層別因子内のバランスがとられる.
- 多くの予後因子を含められない.
- 層毎でバランスを取るため、全体としてそれぞれの層にバランスよく振り分けられない場合に治療群間のバランスが崩れる.

動的割付の方法(1)：Tavesの最小化法(Taves, 1974)

例示：前立腺がん患者に対する例(丹後,2003)

層別因子	年齢		合計
	A剤	B剤	
65歳未満	3	3	6
65歳以上	12	13	25
合計	15	16	31

層別因子	疾病		合計
	A剤	B剤	
A	3	3	6
B	4	4	8
C	8	9	17
合計	15	16	31

層別因子	分化度		合計
	A剤	B剤	
高分化	5	4	9
中分化	8	9	17
低分化	2	3	5
合計	15	16	31

層別因子：年齢, 病期分類, 分化度

参考：丹後俊郎：無作為化臨床試験, 朝倉書店, 2003.

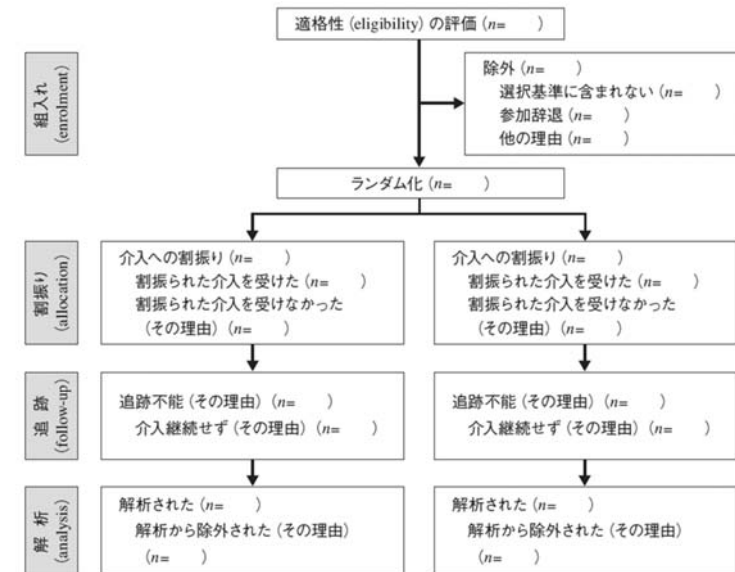
登録された症例の層別因子で分けたときの割り付け症例数の合計を比較する(同数の場合はランダム)

	A剤	B剤	
			「65歳以上, 疾病分類=D, 分化度=高」の場合
年齢	12	13	➔ A剤に割り付けられる
疾病分類	8	9	
分化度	5	4	
合計	25	26	

	A剤	B剤	
			「65歳未満, 疾病分類=D, 分化度=高」の場合
年齢	3	3	➔ 乱数によってランダムに割りつけ
病気分類	8	9	
層別因子	5	4	
合計	16	16	

- 患者が登録される毎に(最初の登録患者は完全無作為化する), それぞれの層別因子毎に症例数の均等化を図る方法である.
- 無作為性を犠牲にする一方で, 層別因子の均等化を登録時点で図りたいというデザインである.

CONSORT Flow Diagramとは



観察・疫学研究を行うためのガイドライン：STROBE

<http://www.strobe-statement.org/>

観察的疫学研究報告の質改善のための声明(STROBE: Strengthening the Reporting of Observational Studies in Epidemiology)とは、観察的疫学研究の研究報告の質の向上と結果の透明性を目指して作られた。

□ STROBE Checklist

コホート研究，ケースコントロール研究，及び横断研究の各デザインについて，論文の中に記載すべき内容に関するチェックリストである。全体で22項目が存在し，そのうちの18項目はすべてのデザインの共通項目になっている。そして，4項目については，上記の3種類のデザインによって異なる。

STROBEの解説と詳細のなかには，具体的な観察的疫学研究の事例とともに，記載内容を明記していることから，研究を整理するうえでも非常に参考になる。

STROBE : チェックリスト(1/4)

タイトル・抄録 (title and abstract)	1	(a) タイトルまたは抄録のなかで，試験デザインを一般に用いられる用語で明示する。 (b) 抄録では，研究で行われたことと明らかにされたことについて，十分な情報を含み，かつバランスの良い要約を記載する。
はじめに(introduction) 背景 (background) 論拠 (rationale)	2	研究の科学的な背景と論拠を説明する。
目的(objective)	3	特定の仮説を含む目的を明記する。
方法 (methods)		
研究デザイン (study design)	4	研究デザインの重要な要素を論文のはじめの部分で示す。
セッティング (setting)	5	セッティング，実施場所ほか，基準となる日付については，登録，暴露，追跡，データ収集の期間を含めて明記する
参加者(participant)	6	(a) ・コホート研究：適格基準，参加者の母集団，選定方法を明記する。追跡の方法についても記述する。 ・ケースコントロール研究：適格基準，参加者の母集団，ケースの選定方法とコントロールの選択方法を示す。ケースとコントロールの選択における論拠を示す。 ・横断研究：適格基準，参加者の母集団，選定方法を示す。 (b) ・コホート研究：マッチング研究の場合，マッチングの基準，暴露群，非暴露群の各人数を記述する。 ・ケースコントロール研究：マッチング研究の場合，マッチングの基準，ケースあたりのコントロールの人数を明記する。

STROBE : チェックリスト(2/4)

変数 (variable)	7	すべてのアウトカム，暴露，予測因子，潜在的交絡因子，潜在的な効果修飾因子を明確に定義する。該当する場合は診断方法を示す。
データ源(data source) /測定方法	8*	関連する各因子に対して，データ源，測定・評価方法の詳細を示す。二つ以上の群がある場合は，測定方法の比較可能性を明記する。
バイアス(bias)	9	潜在的なバイアス源に対応するためにとられた措置があればすべて示す。
研究サイズ (study size)	10	研究サイズ(観察対象者数)がどのように算出されたかを説明する。
量的変数 (quantitative variable)	11	量的変数の分析方法を説明する。該当する場合は，どのグループがなぜ選ばれたかを記載する。
統計・分析方法 (statistical analysis)	12	(a) 交絡因子の調整に用いた方法を含め，すべての統計的方法を示す。 (b) サブグループと交互作用の検証に用いたすべての方法を示す。 (c) 欠損データをどのように扱ったかを説明する。 (d) ・コホート研究：該当する場合は，脱落例をどのように扱ったかを説明する。 ・ケース・コントロール研究：該当する場合は，ケースとコントロールのマッチングをどのように行ったかを説明する。 ・横断研究：該当する場合は，サンプリング方法を考慮した分析法について記述する。 (e) あらゆる感度分析の方法を示す。

STROBE : チェックリスト(3/4)

結果 (result)		
参加者(participant)	13*	(a) 研究の各段階における人数を示す(例：潜在的な適格者数，適格性が調査された数，適格と確認された数，研究に組み入れられた数，フォローアップを完了した数，分析された数)。 (b) 各段階で非参加者の理由を示す。 (c) フローチャートによる記載を考慮する。
記述データ (descriptive data)	14*	(a) 参加者の特徴(例：人口統計的，臨床的，社会的特徴)と暴露や潜在的交絡因子の情報を示す。 (b) それぞれの変数について，データが欠損した参加者数を記載する。 (c) コホート研究：追跡期間を平均および合計で要約する。
アウトカムデータ (Outcome data)	15*	・コホート研究：アウトカム事象の発生数や集約尺度の数値を経時的に示す。 ・ケースコントロール研究：各暴露カテゴリーの数，または暴露の集約尺度を示す。 ・横断研究：アウトカム事象の発生数または集約尺度を示す。
おもな結果 (main result)	16	(a) 調整前の推定値と，該当する場合は交絡因子での調整後の推定値，そしてそれらの精度(例：95%信頼区間)を記述する。どの交絡因子が，なぜ調整されたかを明確にする。 (b) 連続変数がカテゴリー化されているときは，カテゴリー境界を報告する。 (c) 意味のある場合は，相対リスクを意味を持つ期間の絶対リスクに換算することを考慮する。

STROBE : チェックリスト(4/4)

他の解析 (other analysis)	17	その他に行われたすべての分析(例: サブグループと相互作用の解析や感度分析)の結果を報告する。
--------------------------	----	---

考察 (discussion)

鍵となる結果 (key result)	18	研究目的に関しての鍵となる結果を要約する。
限界 (limitation)	19	潜在的なバイアスや精度の問題を考慮して、研究の限界を議論する。潜在的バイアスの方向性と大きさを議論する。
解釈 (interpretation)	20	目的、限界、解析の多重性、同様の研究で得られた結果やその他の関連するエビデンスを考慮し、慎重で総合的な結果の解釈を記載する。
一般可能性 (generalizability)	21	結果の一般可能性を議論する。

その他の情報 (other information)

研究の財源 (funding)	22	研究の資金源、本研究における資金提供者の役割を示す。該当する場合には、現在の研究の元となる研究についても同様に示す。
--------------------	----	--

*ケース・コントロール研究では、ケースとコントロールに分けて記述する。コホート研究と横断研究において該当する場合には、暴露群と非暴露群に分けて記述する。

その他のガイドライン

- PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)**
対象: メタアナリシス, システマティックレビュー

無作為化比較試験のメタアナリシスの報告の質を向上させるためのQUOROM (The Quality of Reporting of Meta-Analyses) 声明を発展させる形式で策定された。PRISMAではQUOROMで想定していなかったシステマティック・レビューを包含するとともに、チェックリストに関しても見直された。

<http://www.prisma-statement.org/>

- SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials)**
対象: プロトコル作成

臨床試験のプロトコルにおいて推奨される最小の記載要件について記載されたガイドラインである。33項目のチェックリストから構成されている。

<http://www.spirit-statement.org/>

ご清聴ありがとうございました

