

Daichi-Sankyo

臨床試験における検定の多重性
～ガイダンス・ガイドラインと実際～
諸外国のガイダンス・ガイドライン

第一三共株式会社
小山 暢之

諸外国のガイダンス・ガイドライン | CONFIDENTIAL

Daichi-Sankyo

目的

- 臨床試験で生じる多重性の問題とその対処方法について諸外国のガイダンス・ガイドラインでの議論を通じて紹介する

Multiple Endpoints in
Clinical Trials

Guidance for Industry

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

EUROPEAN MEDICINES AGENCY
SCIENCE. MEDICINES. HEALTH.

1. 15 December 2016
2. EMA/CHMP/147175/2017
3. Committee for Human Medicinal Products (CHMP)

4. Guideline on multiplicity issues in clinical trials
5. Draft

Draft agreed by Medicines Working Party (MWP)	November 2016
Adopted by CHMP for release for consultation	15 December 2016
Start of public consultation	01 April 2017
End of consultation (deadline for comments)	30 June 2017

諸外国のガイダンス・ガイドライン | CONFIDENTIAL

Daichi-Sankyo

イントロダクション ～臨床試験における多重性の問題～

諸外国のガイダンス・ガイドライン | CONFIDENTIAL

Daichi-Sankyo

多重性の問題とは？

- 複数の帰無仮説をそれぞれ有意水準 α で検定すれば、何れか一つの正しい帰無仮説を棄却してしまう確率は α よりも大きくなること
 - 「下手な鉄砲も数打ちゃ当たる！」



諸外国のガイダンス・ガイドライン | CONFIDENTIAL

なぜ臨床試験で多重性が問題となるのか？

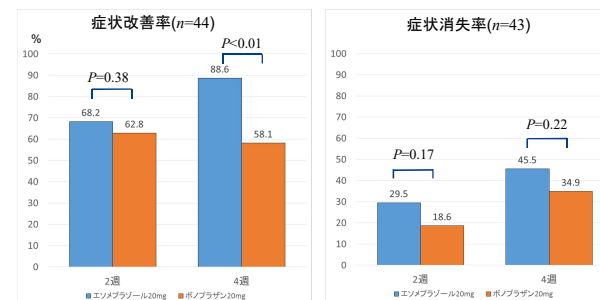
- 薬剤の効果は仮説検定で評価することが多い
 - ある治療の効果を調べるためには他の治療との相対的な比較が必要
 - 薬剤の効果は多面的に評価される
 - 一つの指標・時点・比較対照で薬剤の効果は十分に測れない
- ☞ 通常、複数の評価指標、評価時点、解析対象集団があり、数多くの仮説検定が実施される。

海外版のガイダンス・ガイドライン | CONFIDENTIAL

5

臨床試験でよく見られる解析例

プロトンポンプ阻害剤とカリウムイオン競合型アシッドブロッカーの胃食道逆流症状（GERD症状）に対する効果の無作為化比較試験結果（石原 2017）



主要評価は4週時の症状改善だが、仮説検定（Fisherの直接確率法）は症状改善率と消失率の2週時と4週時に対して行われている。

海外版のガイダンス・ガイドライン | CONFIDENTIAL

6

一つの疾患に対して様々な指標で評価される

- 抗リウマチ薬のACRコアセット
 - ✓ 疼痛及び腫脹関節
 - ✓ 急性期反応物質
 - ✓ 患者の疼痛評価
 - ✓ 患者の全般的活動性評価
 - ✓ 医師の全般的活動性評価
 - ✓ 身体機能評価
 - ✓ 画像診断
- 抗インフルエンザ薬
 - ✓ 罹病期間（解熱や症状消失までの時間）
 - ✓ 臨床症状改善
 - ・ 熱
 - ・ 咳
 - ・ 鼻水・鼻づまり
 - ・ 頭痛
 - ・ のどの炎症
 - ・ 倦怠感 など
 - ✓ ウィルス力価



海外版のガイダンス・ガイドライン | CONFIDENTIAL

7

なぜ臨床試験で多重性が問題になるのか？

- 多重性を考慮せずに試験の結果を解釈してしまうと効果のない薬剤を効果ありと判断してしまう確率が高まる
 - 規制当局としては自国の患者がなるべく不利益を被らないようにプロテクトする必要がある。
- ☞ 多重性の問題は昔から様々な議論がなされている

海外版のガイダンス・ガイドライン | CONFIDENTIAL

8

多重性に関する最近のガイドライン



- Multiple Endpoints in Clinical Trials Guidance for Industry. (FDA, Jan. 2017, draft)
 - 評価項目と多重性の問題との関連及び多重性の調整法を示したガイダンス
- Guideline on multiplicity issues in clinical trials. (EMA, Dec.2016)
 - 臨床試験で生じる多重性の問題を広範囲に取り上げているガイドライン

※ いずれも頻度論の仮説検定による意思決定について議論しているため、本日のお話も頻度論を前提にお話します。

海外薬のガイダンス・ガイドライン | CONFIDENTIAL

9

The family-wise Type I error rate (FWER)



- 他のendpointsに対する効果の有無に関わらず、少なくとも一つのendpointで誤って統計的に有意な治療効果が認められる確率
 - The probability of erroneously finding a statistically-significant treatment effect in at least one endpoint regardless of the presence or absence of treatment effect in the other endpoints within the family.(FDAガイダンス)
- ☞ EMAも同じ定義。一般に、この確率を5%未満(両側)に抑えることが推奨されている

海外薬のガイダンス・ガイドライン | CONFIDENTIAL

10

多重性が生じる主な要因



- 臨床試験で多重性が問題となる主な要因は以下の通り。どの要因が多重性を生じさせるかは、研究の目的や試験デザインに依る
 - Endpoints (評価項目)
 - 評価時点
 - 解析手法
 - 比較群 (用量群)
 - 解析対象集団 (部分集団) など

☞ FDAガイダンスは評価項目に焦点を当てて議論しているが、多重性の基本的な考え方は他の要因の場合も同じ

海外薬のガイダンス・ガイドライン | CONFIDENTIAL

11

The family-wise Type I error rate (FWER)



- 他のendpointsに対する効果の有無に関わらず、少なくとも一つのendpointで誤って統計的に有意な治療効果が認められる確率 (FDA)
 - The probability of erroneously finding a statistically-significant treatment effect in at least one endpoint regardless of the presence or absence of treatment effect in the other endpoints within the family.
- ☞ 上の定義はendpointを解析時点や比較群など、他の要因に置き換えられる

海外薬のガイダンス・ガイドライン | CONFIDENTIAL

12

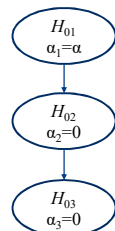
多重性の調整方法の基本的考え方

• FWERを有意水準 α 未満にするには・・・

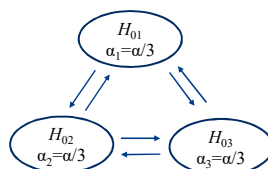
1. 検定の順番をつけて階層的に検定する
2. α を分割して、各検定に割り振る

仮説が棄却されると矢印に沿って、 α が再分配される

階層手順
(Fixed Sequence Method)



Holm法 (α を分割)



杉谷、森川(2017)

海外薬のガイダンス・ガイドライン | CONFIDENTIAL

13

FDAガイダンスで紹介されている多重性の調整法

1. Bonferroni法
2. Holm法
3. Hochberg法
4. α 分割法
5. 固定階層法
6. Fallback法
7. ゲートキーピング法
8. Truncated Holm and Hochberg手順
9. 多枝型ゲートキーピング法
10. リサンプリング法
11. グラフィカルアプローチ (Appendix)

各手法の詳細はこの後の坂巻先生の講演でお聞きください・・・

海外薬のガイダンス・ガイドライン | CONFIDENTIAL

14

多重性の調整と検出力

• 多重性を調整すると検出力は低下する

- FWERの調整が厳しくなるほど検出力は低下するため、必要例数が多くなる
- FWERの調整が複雑になると検出力の計算も難しくなる (計算にはシミュレーションが必要)

☞ 検出力への影響を考慮して、FWERの調整が厳しかったり、複雑な試験計画は避けるべき

海外薬のガイダンス・ガイドライン | CONFIDENTIAL

15

複数の評価項目(Endpoints)と多重性

海外薬のガイダンス・ガイドライン | CONFIDENTIAL

16

複数のEndpointsが設定される理由

1. 複数の重要なEndpointsがある場合
2. 試験目的に対してどのEndpointがベストかのコンセンサスがない場合
3. 何れか一つのEndpointで効果が示されれば臨床的な意義がある場合

☞ 複数のEndpointを一つにまとめられるケースもあるが、個々の指標に対する治療効果を主張する場合は多重性の問題が生じる

海外版のガイダンス・ガイドライン | CONFIDENTIAL

17

評価項目の分類

一般に評価項目は階層的に定義される

1. Primary endpoint(s)
 - 当該薬の有効性（/安全性）を立証するための outcome(s)
2. Secondary endpoint(s)
 - Primary endpoint(s)で評価した効果に対する補足情報を提供するもの
 - Primary endpoint(s)での立証が成功した後、追加の効果を示すもの
3. All other endpoint(s)
 - 上記以外のすべての評価指標

海外版のガイダンス・ガイドライン | CONFIDENTIAL

18

多重性の調整が不要なケース

- 試験が成功（被験薬の効果が検証された）と判断されるケースが1通りしかない場合
 - すべてのEndpointsが統計的に有意にならなければいけない場合（Co-Primary Endpoints）
 - 複数の指標を1つのendpointにまとめた場合
 - ✓ Composite Endpoint
 - ✓ 合計スコア、平均スコア、レスポンスの有無など
 - 事前に定めた順番で複数のEndpointsを階層的に検定する場合
 - ✓ 統計的に有意でないEndpointがあれば、それ以降のEndpoint(s)はすべて有意でないとする

海外版のガイダンス・ガイドライン | CONFIDENTIAL

19

多重性の調整が不要なケース

- 試験が成功（被験薬の効果が検証された）と判断されるケースが1通りしかない場合
 - すべてのEndpointsが統計的に有意にならなければいけない場合（Co-Primary Endpoints）
 - 事前に決めたPrimary endpoint(s)の主解析で治療効果が有意になった後で同一のendpointに対して他の手法で解析しても多重性の問題は生じない（感度分析は多重性を考慮する必要がない）
 - 事前に定めた順番で複数のEndpointsを階層的に検定する場合
 - ✓ 統計的に有意でないEndpointがあれば、それ以降のEndpoint(s)はすべて有意でないとする

海外版のガイダンス・ガイドライン | CONFIDENTIAL

20

多重性の調整が必要なケース

- 試験が成功（被験薬の効果が検証された）と判断されるケースが複数ある場合

- 複数のEndpointsの中の少なくとも一つで効果が示されれば良い場合
 - 結果の良し悪しに関係なく個々のEndpointの結果を全体的なBenefit/risk評価では考慮すべき
- Secondary Endpoint(s)で別の治療効果を主張したい場合

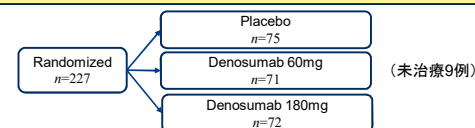
Endpointを解析時点、用量群、解析手法、対象集団などに置き換えても同様

諸外国のガイダンス・ガイドライン | CONFIDENTIAL

21

実際の臨床試験での多重性の調整例 1-(1)

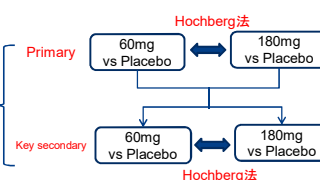
- リウマチ患者を対象としたDenosumabの第II相比較試験
 - Primary EndpointはMRI erosion scoreの6ヵ月後の変化量
 - Key Secondary EndpointはTotal modified Sharp scoreの12ヵ月後の変化量



多重性の調整法

- 投与群間の比較（60mg vs Placebo, 180mg vs Placebo）はHochberg法
- Primaryが有意になった場合のみKey Secondaryを検定

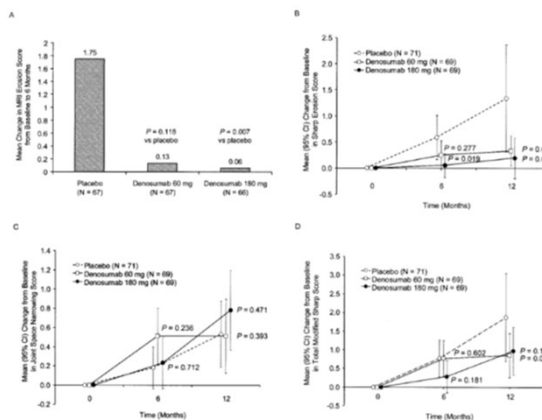
段階的に検定



諸外国のガイダンス・ガイドライン | CONFIDENTIAL

(Cohenら 2008)

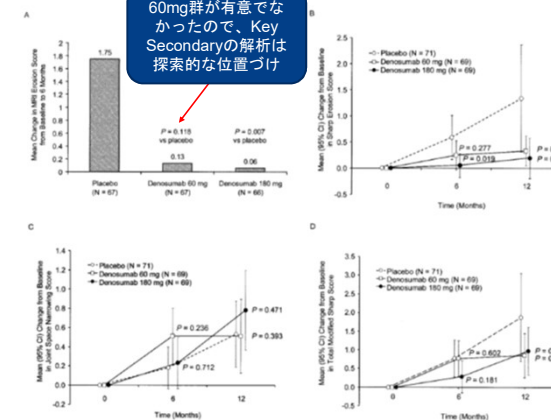
実際の臨床試験での多重性の調整例 1-(2)



諸外国のガイダンス・ガイドライン | CONFIDENTIAL

23

実際の臨床試験での多重性の調整例 1-(2)



諸外国のガイダンス・ガイドライン | CONFIDENTIAL

24

Composite Endpoint

- 事前に定めた複数の臨床イベントの何れか一つの発現をカウントして評価することをComposite Endpoint(以下、Compo EP)という。

Table 1. Decomposition of Endpoint Events in RENAAL*

Endpoint	Losartan (N=751)	Placebo (N=762)	Hazard ratio= (95% CI)	p-value
Primary endpoint				
Doubling of serum creatinine, ESRD, or death	327	359	0.84 (0.72, 0.97)	0.022
Decomposition of the primary endpoint				
Doubling of serum Creatinine	162	198	0.75	
ESRD	64	65	0.93	
Death	101	96	0.98	
Any occurrence of individual components				
Doubling of serum Creatinine	162	198	0.75 (0.61, 0.92)	
ESRD	147	194	0.71 (0.57, 0.89)	
Death	158	155	1.02 (0.81, 1.27)	

ESRD: end-stage renal disease

諸外国のガイダンス・ガイドライン | CONFIDENTIAL

25

Composite Endpoint

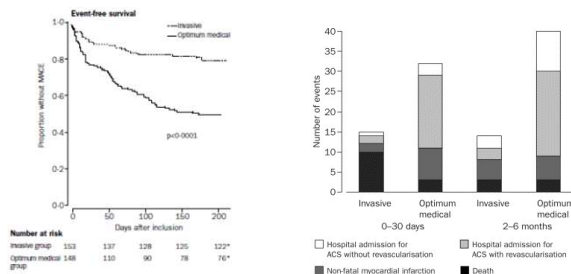
- Compo EPは通常、Primary Endpointとなる
- Compo EPに含まれるイベントより臨床的に重いイベントはCompo EPに含まれているべき
- Compo EPに含まれる個々の指標の評価も重要
 - Compo EP全体がPositiveでもその中に含まれる一部の指標がnegativeだと解釈が難しい
 - 治療効果が認められない指標を含めるとばらつきが大きくなり、非劣性試験の場合には特に問題となる

諸外国のガイダンス・ガイドライン | CONFIDENTIAL

26

Composite Endpointの解釈が問題となった例

Trial of invasive versus medical therapy in elderly patients with chronic symptomatic coronary-artery disease (TIME): a randomized trial (The TIME investigators 2001)



Composite endpointはinvasive群が統計的に有意に優っているが、死亡例はinvasive群の方が多い。(Hospital admissionで大きな差があるため)

諸外国のガイダンス・ガイドライン | CONFIDENTIAL

27

Secondary Endpoint(s)と多重性 (FDA)

- Secondary Endpoint(s)と多重性 (FDA)
- Secondary Endpoint(s) (以下、2nd EP) 内及びPrimaryと2nd EP間に対しても α エラーを制御すべき
- 2nd EPのPositiveな評価はPrimary Endpoint(s)が検証された場合のみ可能
- 2nd EPの検証が重要なら、必要例数など試験デザイン時に考慮すべき
 - 一般に2nd EPに対する検出力は高くない
- 2nd EPの数は少なくすべき
 - 数が多いと多重性の調整が厳しくなり、有意になりにくい
- 新たな仮説作りのための評価項目は2nd EPではなく exploratory endpointにする

諸外国のガイダンス・ガイドライン | CONFIDENTIAL

28

Secondary endpointsと多重性（EMA）



- 2nd EPが主たる結果をサポートする場合
 - 検証的な結果を主張しないので、多重性を考慮は不要
 - 多重性を調整しない場合でもkey secondaryを設定すべき
- 2nd EPの結果を検証的に扱う場合
 - 主の結果が検証された場合のみ追加の主張が可能
 - 階層的に検定するなど多重性を考慮して解析せよ
 - 複雑な方法で多重性を調整する場合は要当局相談
- 臨床的に重要だが、検出力が確保できないために2nd EPにした場合
 - 予期せぬ程の大きな効果が認められてもPrimaryが検証できなければ追加試験での検討が必要
 - Primaryは検証できてもこの2nd EPが逆効果なら承認不可の可能性もある

海外版のガイダンス・ガイドライン | CONFIDENTIAL

29

安全性の評価指標と多重性（EMA）



- 安全性の指標が承認や効能に関連する場合には有効性の指標と同等に扱うべきだが、検出しようとしている差が安全性上の問題を生じる方向の場合は除く
 - ✓ 有意でない結果が安全性に関して懸念がないあるいは対照と同等を意味するわけではない。
 - ✓ リスクのシグナルを検討する上で、多重性の調整は逆効果に働く
 - ✓ （安全性上）意義のある結果と判断するには薬物動態の知識や更なる追加の検討が必要になる可能性がある

FDAガイダンスは安全性の評価に関する多重性については対象外としている

海外版のガイダンス・ガイドライン | CONFIDENTIAL

30

複数の解析対象集団に対する多重性



海外版のガイダンス・ガイドライン | CONFIDENTIAL

31

解析対象集団と多重性



- 主たる解析対象集団を決めている場合は多重性の調整は不要
 - その他の解析対象集団での解析結果は主たる解析対象集団での結果を補足したり、強化するためのものであり、試験の結論に影響しないため
- ただし、特定の部分集団に対して検証的な結果を主張する場合には多重性の問題が生じる

海外版のガイダンス・ガイドライン | CONFIDENTIAL

32

部分集団解析の解釈

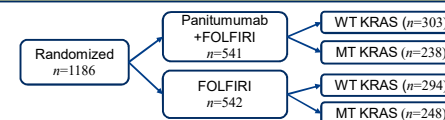
- 全体と部分集団のどちらかがpositiveならばOK
なら多重性の調整が必要
 - ✓ 事前規定する部分集団の数は少なくすべき
 - ✓ 検出力の考察をプロトコル記載
 - ✓ 比較可能性のため、重要なcovariateに関して層別無作為化が必要
- 以下の場合承認が制限される可能性がある
 - ✓ 重要な部分集団の結果に説明できない異質性がある場合
 - ✓ 異質性が想定されても重要な部分集団に対して十分に評価できない場合

海外薬のガイダンス・ガイドライン | CONFIDENTIAL

33

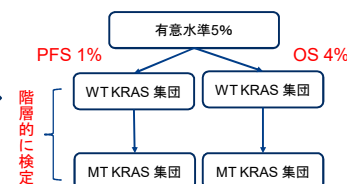
実際の臨床試験での多重性の調整例

- 転移性大腸癌を対象としたpanitumumab + FOLFIRI vs FOLFIRIの非盲検比較試験
 - 評価項目はPFSとOS
 - KRAS遺伝子変異の有無で被験者を分けて評価



多重性の調整法

- 5%の有意水準をPFSとOSに分配
- WT KRASが有意になった場合のみMT KRASを検定



海外薬のガイダンス・ガイドライン | CONFIDENTIAL

多群比較で生じる多重性

海外薬のガイダンス・ガイドライン | CONFIDENTIAL

35

多群比較試験と多重性

- 被験薬、対照薬、プラセボの3群比較試験
 - ✓ プラセボに対する優越性と対照薬に対する非劣性が同時に検証される必要があるため、多重性の問題は生じない
 - ✓ ただし、プラセボに対する優越性が主目的で、対照薬との非劣性は副次的な目的なら階層的に検証する必要がある
- 固定用量の組合せ試験
 - ✓ 単独の効果と組合せの相加/相乗効果が同時に検証される必要があるので多重性の問題は生じない
 - ✓ ただし、組合せの中からベストなものを選択する場合は多重性の問題が生じる

海外薬のガイダンス・ガイドライン | CONFIDENTIAL

36

多群比較試験と多重性



• 用量反応試験

- ✓ Ph2の段階であれば検定より推定ベースでの判断が重要であり、必ずしも多重性を考慮した検定は必要でない
- ✓ ただし、検証的な試験でベストあるいはベターな用量選択をするのであれば多重性の調整は必須
 - 調整方法は試験の目的とデザインに即して選択する

海外版のガイダンス・ガイドライン | CONFIDENTIAL

37

多重性と推定



海外版のガイダンス・ガイドライン | CONFIDENTIAL

38

推定における多重性の調整



- 多重性がある場合、信頼区間も名義的な被覆確率よりも低下する
 - しかしながら、多重性を調整した検定と対応した推定法は存在しないか、存在しても計算が難しかったり、有用でないことが多い
- 棄却域が仮説検定と信頼区間で異なる場合もある
 - 意思決定は仮説検定の結果を用いるべき
 - Bonferroni法で修正した信頼区間のような単純で保守的な信頼区間を用いると良い

海外版のガイダンス・ガイドライン | CONFIDENTIAL

39

まとめ



- 2つのガイドラインは議論の対象範囲は異なるものの、基本的な考え方は類似している
 - 検証対象、解析方法など事前規定が重要
 - 複雑な多重性が生じる設定は調整も解釈も難しくなるので、避けるべき
 - Primary EPが検証された後、2nd EPで別の効果を主張をする場合は多重性の調整必要
- 検出力については詳しく述べられていない
 - 多重性の調整法はいろいろあるが、調整による検出力低下への対応は悩ましい問題

海外版のガイダンス・ガイドライン | CONFIDENTIAL

40

参考文献



1. Multiple Endpoints in Clinical Trials Guidance for Industry (draft) . FDA, Jan. 2017.
2. Guideline on multiplicity issues in clinical trials, EMA, Dec.2016.
3. エソメプラゾール20mgとボノブラザン20mgのGERD症状改善効果の検討. Therapeutic Research, Vol.38, 973-,978, 2017.
4. 階層構造化された試験目的を伴う臨床試験での多重比較法ーゲートキーピング法とグラフィカルアプローチ. 杉谷&森川, 計量生物学 Vol.38 41-78, 2017.
5. Denosumab Treatment Effects on Structural Damage Bone Mineral Density, and Bone Turnover in Rheumatoid Arthritis A Twelve-Month, Multicenter, Randomized, Double-Blind, Placebo-Controlled, Phase II Clinical Trial. Cohen, et al. Arthritis & Rheumatism Vol.58, 1299-1309, 2008.
6. Randomized Phase III Study of Panitumumab With Fluorouracil, Leucovorin, and Irinotecan (FOLFIRI) Compared With FOLFIRI Alone As Second-Line Treatment in Patients With Metastatic Colorectal Cancer. Peeters, et al. Journal of Clin Onco, Vol.28, 4668-4697, 2010.

海外版のガイダンス・ガイドライン | CONFIDENTIAL

41