

多重性調整に関する 現在の実践

吉田征太郎
中外製薬株式会社
東京大学

発表構成

- 臨床試験における多重性調整に関する現状
 - 多重性に関するガイドライン
 - 日本における多重性調整の現状（多群試験）
- 臨床試験における多重性調整の実例
 - 当局と議論になった事例
 - 複雑な仮説構造の事例
- まとめ

2017/11/29

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～

2

臨床試験における多重性調整に関する現状

2017/11/29

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～

3

多重性に関する各規制当局の指針

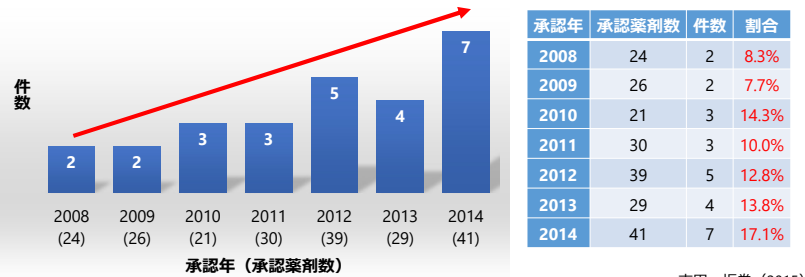
- ICH E9 (1998)
 - 多重性に関しては事実上日本における唯一の行政指針
- CPMP (2002)
 - POINTS TO CONSIDER ON MULTIPLICITY ISSUES IN CLINICAL TRIALS
- CHMP (2017)
 - Guideline on multiplicity issues in clinical trials
- FDA (2017)
 - Multiple Endpoints in Clinical Trials Guidance for Industry

2017/11/29

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～

4

承認薬剤数に占めるGatekeeping法を用いた薬剤の件数 (FDA)



対象：公開されている承認品目一覧（年次 New Molecular Entity (NME) Drug and New Biologic Approvals）のうち、statistical review(s) が公開されている薬剤
 17/11/29 臨床試験における検定の多重性～ガイドライン・ガイドラインと実際～

吉田, 坂巻 (2015)

5

日本における多重性調整の実際 (多群試験)

Current Practice on Multiplicity Adjustment and Sample Size Calculation in Multi-arm Clinical Trials: An Industry Survey in Japan

Kentaro Sakumaki, PhD¹, Toshifumi Kamiura^{1,2}, Yusuke Morita^{1,2}, Kazuhiko Ito^{1,2}, Setaro Yoshida^{1,2}, Akira Wakana, PhD^{1,2}, Satoru Tsuchiya^{1,2}, Satoru Fukuhara, PhD^{1,2}, and Hideki Suganami, PhD^{1,2}

- 製薬協加盟会社66社を対象
- 44社より151試験の回答
 - 内資33社, 外資11社
- 症例数設計時の多重性考慮有無についても調査
- 現在multi endpointについても調査を実施し論文投稿検討中

DIA STATISTICS
INDUSTRY SURVEY

	Confirmatory	Exploratory	Total
Region where the trial was conducted	n = 91	n = 59	n = 150
Japan only	54 (59.3%)	46 (78.0%)	100 (66.7%)
Multi-regional	36 (39.6%)	13 (22.0%)	49 (32.7%)
Nice disclosed	1 (1.1%)	0 (0%)	1 (0.7%)
Trial phase	n = 91	n = 58	n = 149
Ia	0 (0%)	29 (50.0%)	29 (19.5%)
Ib	21 (23.1%)	28 (48.3%)	49 (32.9%)
II/III	14 (15.4%)	0 (0%)	14 (9.4%)
III	56 (61.5%)	0 (0%)	56 (37.6%)
IV	0 (0%)	1 (1.7%)	1 (0.7%)
Disease classification	n = 90	n = 60	n = 150
Diseases of the circulatory system	7 (7.8%)	2 (3.3%)	9 (6.0%)
Neoplasms	6 (6.7%)	4 (6.7%)	10 (6.7%)
Gastrointestinal disease	4 (4.4%)	6 (10.0%)	10 (6.7%)
Endocrine, nutritional, and metabolic diseases	10 (11.1%)	7 (11.7%)	17 (11.3%)
Inflammation, allergy, and immunologic diseases	9 (10.0%)	7 (11.7%)	16 (10.7%)
Diseases of the nervous system	23 (25.6%)	4 (6.7%)	27 (18.0%)
Diseases of the respiratory system	9 (10.0%)	11 (18.3%)	20 (13.3%)
Diseases of the sensory system	15 (16.7%)	4 (6.7%)	19 (12.7%)
Pain	1 (1.1%)	7 (11.7%)	8 (5.3%)
Other	6 (6.7%)	8 (13.3%)	14 (9.3%)

2017/11/29

臨床試験における検定の多重性～ガイドライン・ガイドラインと実際～

Sakumaki et al (2016)⁶

多重性調整有無と主な方法

	Confirmatory	Exploratory	Total
Multiplicity adjustment for multi-arm comparison	n = 91	n = 60	n = 151
Controlled	89 (97.8%)	23 (38.3%)	112 (74.2%)
Not controlled	2 (2.2%)	37 (61.7%)	39 (25.8%)
MCP ^{a,b}	n = 89	n = 23	n = 112
Bonferroni procedure	6 (6.7%)	3 (13.0%)	9 (8.0%)
Dunnnett procedure	11 (12.4%)	5 (21.7%)	16 (14.3%)
Holm procedure	1 (1.1%)	1 (4.3%)	2 (1.8%)
Hochberg procedure	13 (14.6%)	1 (4.3%)	14 (12.5%)
Hommel procedure	1 (1.1%)	0 (0%)	1 (0.9%)
Fixed sequence procedure	34 (38.2%)	6 (26.1%)	40 (35.7%)
Gatekeeping procedure	22 (24.7%)	0 (0%)	22 (19.6%)
Multiple comparison procedure-modeling	0 (0%)	3 (13.0%)	3 (2.7%)
Maximum method			
Other			

複雑な仮説構造に対応した方法や比較的新しい方法の適用事例が存在

- 検証試験ではほぼ全ての試験で多重性を調整
- 探索的試験でも4割近くの試験で多重性を調整
- 検証試験では固定順序法やGatekeeping法の使用が多い
- 探索試験ではMCP-Modの適用事例も存在

2017/11/29

7

臨床試験における多重性調整の実例

2017/11/29

臨床試験における検定の多重性～ガイドライン・ガイドラインと実際～

8

【事例 1】Plegridyの審査事例

- plegridy (pegylated interferon beta-1a) は再発寛解型多発性硬化症の治療薬として2014年7月にEMAに承認
- 主要評価項目に加え、主要副次評価項目までを含めた複数の評価項目間の多重性調整を行い、統計的有意かつ臨床的意義が認められた主要副次評価項目も含め承認申請を実施
- CHMPによるAssessment Reportにおける主要副次評価項目の検定順序についての議論

European Medical Agency, Committee for Medicinal Products for Human Use (CHMP), Plegridy Assessment Report, EMA/339024/2014, Rev. 1.
http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/002827/WC500170303.pdf

2017/11/29

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～

9

【事例 1】Plegridyの審査事例

105MS301試験のデザインの概要

Phase	III
試験デザイン	二重盲検並行群間比較国際共同試験
投与群	1) プラセボ群 2) 125μg 4週間隔投与群 (Q4W群) 3) 125μg 2週間隔投与群 (Q2W群)
主要評価項目	1人年あたりの再発率
主要副次評価項目	1) MRIによるT2新規・拡大病巣数 2) 1年時点での再発患者の割合 3) 身体機能障害の進行が認められた患者の割合

2017/11/29

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～

10

【事例 1】Plegridyの審査事例

多重性に関連した審査上の論点

- 主要評価項目に加え検証を意図した主要副次評価項目が重要度の高いものから3つ設定
- 投与群は3群設定されており、各評価項目に対しQ2W群 vs プラセボ群の検定を行い、統計学的有意差が得られた場合はQ4W群 vs プラセボ群の検定を実施
- 多重性は固定順序法により調整

2017/11/29

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～

11

【事例 1】Plegridyの審査事例

105MS301試験の解析結果

評価項目		Q2W vs プラセボ	Q4W vs プラセボ
主要評価項目 1人年あたりの再発率	プラセボに対する減少率	35.6 %	27.5 %
	p値	0.0007	0.0114
主要副次評価項目1) MRIによるT2新規・ 拡大病巣数	プラセボに対する減少率	67 %	28 %
	p値	0.0001	0.0008
主要副次評価項目2) 1年時点での再発患者の 割合	プラセボに対する減少率	39 %	26 %
	p値	0.0003	0.0200
主要副次評価項目3) 身体機能障害の進行が 認められた患者の割合	プラセボに対する減少率	38 %	38 %
	p値	0.0383	0.0380

2017/11/29

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～

12

【事例 1】Plegridyの審査事例

結果に対するCHMPからの指摘

- 申請者が主要副次評価項目の第3位に位置づけた「身体機能障害の進行が認められた患者の割合」を、審査上は最重要な主要副次評価項目として位置づけたと指摘
- *"Disability (3-month sustained disability progression) was included as a secondary endpoint, however, it was only ranked 3rd among the secondary endpoints. In this assessment, disability was considered to be the most important secondary endpoint."*

2017/11/29

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～

13

【事例 1】Plegridyの審査事例

指摘の背景： 疾患評価ガイドラインの存在

- CHMPからの指摘は多発性硬化症の疾患評価ガイドライン「*Guideline on clinical investigation of medicinal products for the treatment of Multiple Sclerosis*」において下記の記載があることによるものと考えられる
- *"A relapse-based primary endpoint though cannot be taken as a surrogate for disability progression and this would be expressed accordingly in the SmPC. Moreover, if the primary endpoint is based on relapse assessment, progression of disability should be evaluated as key secondary endpoint."*

SmPC: 欧州製品情報概要 (Summary of Product Characteristics)

2017/11/29

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～

14

【事例 1】Plegridyの審査事例

本事例からの教訓： 仮説構造の当局との事前合意の重要性

- 臨床試験における多重性の問題は統計学的な観点だけで解決できるものではない
- 疾患評価ガイドライン等を踏まえ臨床的観点からの検討も必要
- 本事例は承認申請後に当局より仮説構造の変更を指示されたものであり、仮説構造については予め規制当局と合意しておくことが望ましい

2017/11/29

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～

15

【事例 2】Spiolto Respimatの審査事例

- Spiolto Respimat（チオトロピウム臭化物／オロダテロール塩酸塩）は慢性肺閉塞性疾患（COPD）の治療薬として2015年8月に本邦で承認
- 規制当局間で異なる仮説構造を提示された事例

チオトロピウム臭化物水和物・オロダテロール塩酸塩 CTD <http://www.pmda.go.jp/drugs/2015/P20150908001/index.html>

2017/11/29

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～

16

【事例2】 Spiolto RespiMATの審査事例

事例の概要

- プロトコルにEMAとFDA用の仮説構造をそれぞれ設定
 - 日米欧の三極での申請を意図した2つの検証的国際共同治験を計画
 - EMAとFDAとで有効性の主要評価に関する疾患評価ガイドライン上の要件が異なった
- FDA申請用の仮説構造
 - 呼吸機能に関する評価項目のみ仮説構造に設定
- EMA申請用の仮説構造
 - CHMPのCOPD疾患評価ガイドラインに従い、主要評価項目に疾患特異的QOLスコアであるSGRQ総スコア、主要副次評価項目に症状に対するTDI総スコアを設定

SDRG: St. George Respiratory Questionnaire, TDI: Transitional Dyspnea Index

2017/11/29

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～

17

【事例2】 Spiolto RespiMATの審査事例

FDA・EMA用の仮説構造

評価項目	被験者群	対照群	FDA用仮説構造での順位	EMA用仮説構造での順位
主要評価項目				
Mean FEV ₁ AUC _{0-3h}	Tio + Olo 5/5 µg	Olo 5 µg	1	1
		Tio 5 µg	2	2
Mean トラブFEV ₁	Tio + Olo 5/5 µg	Olo 5 µg	3	3
		Tio 5 µg	4	4
Mean SGRQ総スコア (2試験併合)	Tio + Olo 5/5 µg	Olo 5 µg	-	5
		Tio 5 µg	-	6
Mean FEV ₁ AUC _{0-3h}	Tio + Olo 2.5/5 µg	Olo 5 µg	5	7
		Tio 2.5 µg	6	8
Mean トラブFEV ₁	Tio + Olo 2.5/5 µg	Olo 5 µg	7	9
		Tio 2.5 µg	8	10
Mean SGRQ総スコア (2試験併合)	Tio + Olo 2.5/5 µg	Olo 5 µg	-	11
		Tio 2.5 µg	-	12
主要副次評価項目				
TDI総スコア (2試験併合)	Tio + Olo 5/5 µg	Olo 5 µg	-	13
		Tio 5 µg	-	14
	Tio + Olo 2.5/5 µg	Olo 5 µg	-	15
		Tio 2.5 µg	-	16
副次評価項目				
Mean FEV ₁ AUC _{0-3h}	Tio + Olo 2.5/5 µg		9	17
Mean トラブFEV ₁	Tio + Olo 2.5/5 µg		10	18
Mean SGRQ総スコア (2試験併合)	Tio + Olo 2.5/5 µg		-	19
TDI総スコア (2試験併合)	Tio + Olo 2.5/5 µg		-	20

2017/11/29

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～

18

【事例3】 複雑な仮説構造の例[KEYNOTE-10]

- 目的
 - PD-L1陽性進行性非小細胞肺癌患者を対象としたpembrolizumabとdocetaxelの比較
- 概要
 - 試験の相：Ⅱ/Ⅲ
 - 投与群: pembrolizumab 2mg/kg, pembrolizumab 10mg/kg, docetaxel 75mg/m² (計3群)
 - 主要評価項目: PD-L1強発現集団およびPD-L1発現集団におけるProgression free survival (PFS)とOverall survival (OS)
 - 中間解析: 2回 (1回目: Futility, 2回目: Efficacy)

Herbst et al (2016)

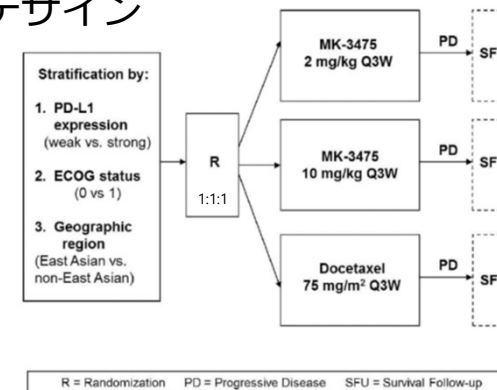
2017/11/29

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～

19

【事例3】 複雑な仮説構造の例: KEYNOTE-10

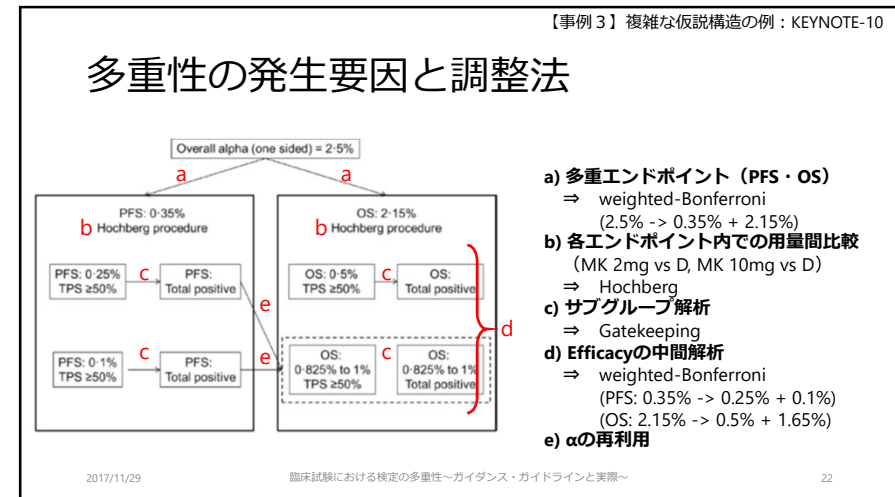
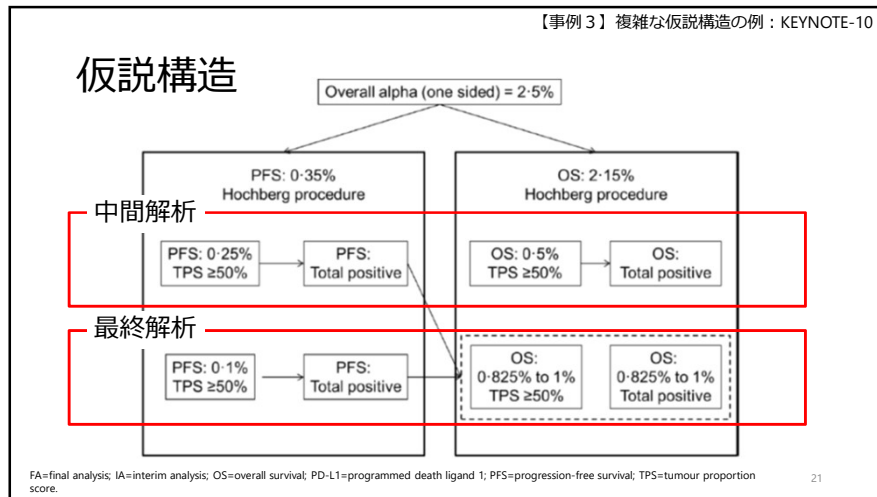
試験デザイン



2017/11/29

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～

20



【事例3】複雑な仮説構造の例：KEYNOTE-10

仮説構造を考慮した有意水準の調整

IA2 (alpha = 0.25% for PFS)	Final analysis (alpha = 0.1% for PFS)	Nominal alpha for OS in PD-L1 strong positive	Both MK arms are superior in OS at FA Empirical hazard ratio ¹ (empirical median improvement in month ²)	Only One MK arm is superior in OS at FA Empirical hazard ratio ¹ (empirical median improvement in month ²)
At least one MK arm does not have superior PFS	At least one MK arm does not have superior PFS	0.825%	0.667 (4.5 m.)	0.640 (5.1 m.)
At least one MK arm does not have superior PFS	Both MK arms have superior PFS	0.875%	0.669 (4.4 m.)	0.642 (5.0 m.)
Both MK arms have superior PFS	At least one MK arm does not have superior PFS	0.95%	0.673 (4.4 m.)	0.645 (5.0 m.)
Both MK arms have superior PFS	Both MK arms have superior PFS	1.00%	0.675 (4.3 m.)	0.647 (4.9 m.)

¹ Observed treatment effect which may lead to statistical significance based on 140 deaths between one MK arm and docetaxel arm in the PD-L1 strongly positive stratum, and half of allocated alpha as a conservative estimate under Hochberg procedure
² Median survival for docetaxel is assumed to be 9 months

【事例3】複雑な仮説構造の例：KEYNOTE-10

仮説構造を考慮した有意水準の調整

Overall alpha (one sided) = 2.5%

最終解析のOSの有意水準

IA2 (alpha = 0.25% for PFS)	Final analysis (alpha = 0.1% for PFS)	Nominal alpha for OS in PD-L1 strong positive
At least one MK arm does not have superior PFS	At least one MK arm does not have superior PFS	0.825%
At least one MK arm does not have superior PFS	Both MK arms have superior PFS	0.875%
Both MK arms have superior PFS	At least one MK arm does not have superior PFS	0.95%
Both MK arms have superior PFS	Both MK arms have superior PFS	1.00%

まとめ

- 仮説構造の複雑化に伴い多重性調整の方法論も発展
 - 臨床試験でもGatekeeping法など複雑な仮説構造に対応した適用事例が増加
- 仮説構造と意思決定の方法の試験計画時の議論が重要
 - 開発の早期段階から関係機能との十分な議論が必要
 - 統計解析担当者は方法論の正確な理解と非統計家にもわかりやすく説明するコミュニケーション能力が求められる
 - 当局とも事前に相談し合意しておく（さもないと申請時に困る）

製薬協より「臨床試験における多重性の諸問題に関する最近の展開と今後の課題について
－非統計解析担当者のための解説書－」（仮題）を公開予定。

2017/11/29

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～

25

謝辞

本発表の一部は日本製薬工業協会医薬品評価委員会データサイエンス部会2017年度タスクフォース4での調査・検討結果に基づきます。

この場をお借りしてタスクフォースメンバーに感謝申し上げます。

菅波 秀規 興和株式会社（推進委員）
土屋 悟 大日本住友株式会社（推進委員）
富金原 悟 小野薬品工業株式会社（推進委員）
伊庭 克拓 大塚製薬株式会社
浦田 正夫 サノフィ株式会社（2017年4月まで）
小川 直之 株式会社三和化学研究所
神浦 俊文 日本新薬株式会社
森田 祐介 杏林製薬株式会社

2017/11/29

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～

26

参考文献

- 吉田征太郎, & 坂巻顕太郎. (2015). 臨床試験における Gatekeeping 法の適用例. 計量生物学, 36(Special_Issue_2), S165-S171.
- Sakamaki, K., Kamiura, T., Morita, Y., Iba, K., Yoshida, S., Wakana, A., ... & Suganami, H. (2016). Current Practice on Multiplicity Adjustment and Sample Size Calculation in Multi-arm Clinical Trials: An Industry Survey in Japan. *Therapeutic Innovation & Regulatory Science*, 50(6), 846-852.
- European Medical Agency. Committee for Medicinal Products for Human Use (CHMP). Plegidry Assessment Report. EMA/339024/2014, Rev. 1. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/002827/WC500170303.pdf
- European Medical Agency. Committee for Medicinal Products for Human Use (CHMP). Guideline on clinical investigation of medicinal products for the treatment of Multiple Sclerosis. EMA/CHMP/771815/2011, Rev. 2. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2015/03/WC500185161.pdf
- チオトロピウム臭化物水和物・オロダテロール塩酸塩 CTD <http://www.pmda.go.jp/drugs/2015/P20150908001/index.html>
- European Medical Agency. Guideline on clinical investigation of medicinal products in the treatment of chronic obstructive pulmonary disease (COPD). 2012. Available online at http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/08/WC500130880.pdf
- Herbst, R. S., Baas, P., Kim, D. W., Felip, E., Pérez-Gracia, J. L., Han, J. Y., ... & Majem, M. (2016). Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *The Lancet*, 387(10027), 1540-1550.

2017/11/29

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～

27

Survey Questions (Sakamaki et al, 2016)

Table 1. Survey Questions.

Q1	Type of company	Multiple choice (select one)
Q2	Internal standard guidelines or position papers on multiplicity (if any)	Free text
Q3-1	Disease classification (therapeutic area)	Multiple choice (select one)
Q3-2	Disease name	Free text
Q3-3	Region where a trial was conducted	Multiple choice (select one)
Q3-4	Trial phase	Multiple choice (select one)
Q3-5	Type of trial (confirmatory or exploratory) ^a	Multiple choice (select one)
Q3-6	Number of arms	Free text
Q3-7	Control arm(s)	Multiple choice (multiple answers allowed)
Q3-8	Special notes about control arm(s)	Free text
Q3-9	Experimental arm(s)	Multiple choice (multiple answers allowed)
Q3-10	Free description about the constitution of arms	Free text
Q3-11	Control arm(s) used for primary analysis	Multiple choice (multiple answers allowed)
Q3-12	Comparisons in primary analysis	Multiple choice (select one)
Q3-13	Statistical method for primary analysis	Free text
Q3-14	Multiplicity adjustment for multi-arm comparison	Multiple choice (select one)
Q3-15	What MCP was used?	Multiple choice (multiple answers allowed)
Q3-16	Sources of multiplicity other than multi-arm	Multiple choice (multiple answer allowed)
Q3-17	Reason for selecting MCP	Multiple choice (multiple answers allowed)
Q3-18	Whether to consider multiplicity adjustment in sample size calculation	Multiple choice (select one)
Q3-19	Considerations for sample size calculation	Multiple choice (multiple answers allowed)
Q3-20	Special topics on this study related to multiplicity	Free text

Abbreviation: MCP, multiple comparison procedure.
^aWe did not explicitly define "confirmatory" or "exploratory."

2017/11/29

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～

28

Sample size calculation

Table 5. Sample Size Calculation

	Confirmatory n = 89	Exploratory n = 58	Total n = 147
Consideration of multiple comparison in sample size calculation			
Considered	61 (68.5%)	15 (25.9%)	76 (51.7%)
Not considered	28 (31.5%)	43 (74.1%)	71 (48.3%)
Considerations for sample size calculation ^a			
No statistical calculation	1 (1.1%)	19 (32.8%)	20 (13.4%)
Probability of statistical significance in most effective treatment vs control comparison	22 (24.2%)	13 (22.4%)	35 (23.5%)
Probability of statistical significance in least effective treatment vs control comparison	17 (18.7%)	3 (5.2%)	20 (13.4%)
Probability of statistical significance in at least one pairwise comparison	32 (35.2%)	9 (15.5%)	41 (27.5%)
Probability of statistical significance in all pairwise comparisons	22 (24.2%)	1 (1.7%)	23 (15.4%)
Dose-response relationship	8 (8.8%)	10 (17.2%)	18 (12.1%)
Unequal allocation ratio	5 (5.5%)	3 (5.2%)	8 (5.4%)
Conservative MCP that is different from planned MCP	7 (7.7%)	4 (6.9%)	11 (7.4%)
Other source of multiplicity	6 (6.6%)	1 (1.7%)	7 (4.7%)
Other	11 (12.1%)	6 (10.3%)	17 (11.4%)

Abbreviation: MCP, multiple comparison procedure.

^aCalculated as number in each category / Number Of effective answers (n) × 100.

こおける検定の多重性～ガイドランス・ガイドラインと実際～

29

- 検証試験では7割がサンプルサイズ時に多重性を考慮
- 探索的試験では3割未満