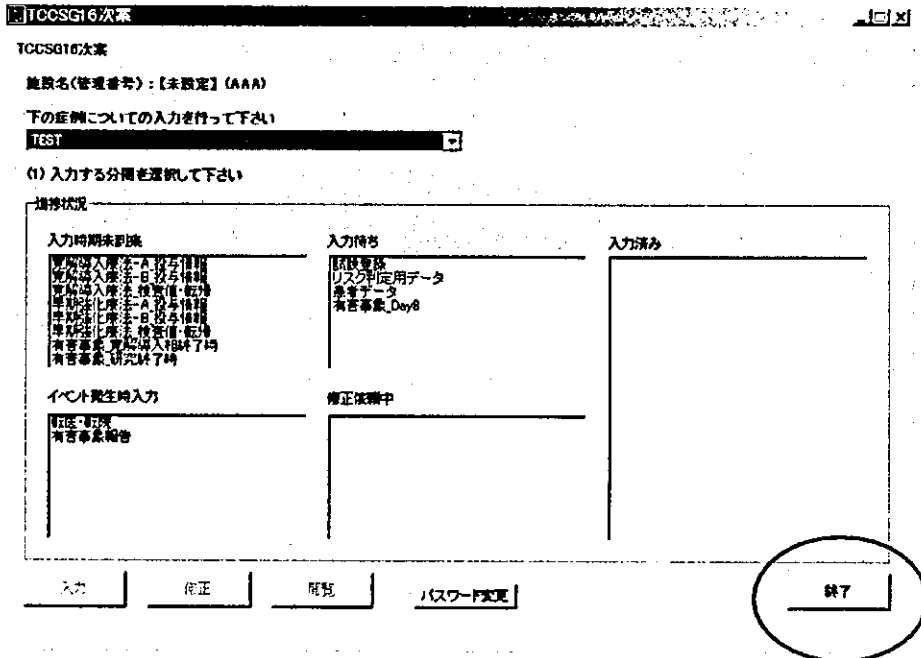


6. Shuttle の終了方法

6. 1 [Shuttle]の終了

メイン画面の終了ボタンをクリックすることで[Shuttle]は終了します。



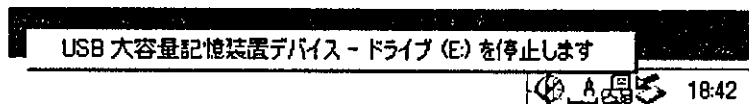
6. 2 USB メディアの抜き取り

・Windows98SE の場合

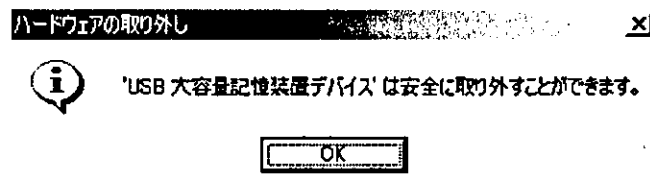
エクスプローラで、[Shuttle]のドライブのアイコンを右クリックし、取り外し（または取り出し）を選択します。USB メディアのランプが点滅していないこと確認してから、抜き取ってください。

・Windows2000,XP の場合

ご使用のパソコンのタスクバー（通常ディスプレイ右下）から、「ハードウェアを取り外すか取り出す」アイコンをクリックして、デバイスを停止させます。



複数のデバイスが表示された場合には、[Shuttle]に対応するドライブを選択してください。



上のメッセージが出力されたことを確認した後、[Shuttle]を抜き取ってください。

(注) USB メディア内のデータが破損する原因は、そのほとんどが不適切な抜き取り方法です。USB メディアの抜き取りには細心の注意を払って頂きますようお願い致します。

7. その他ご使用にあたっての注意

7. 1 パスワードの管理

[Shuttle]において、パスワードは、個人、または、権限を有する集合を判別する最重要の手段となっています。

パスワードを臨床試験のデータを閲覧する権限、もしくは入力する権限のない第三者に漏洩させることは絶対に避けて下さい。パスワードが第三者に漏洩されたことが確認された場合、そのパスワードを使用した全ての臨床試験データは、全く信頼性のないものとなってしまいます。

7. 2 データの改ざんなどに対する対処

万一、悪意のある第三者によって、データが改ざんされた場合には、データが改ざんされた旨のメッセージが出力されます。この時は至急、データセンターまでご連絡して頂きますようお願い致します。

**アップデート中、ごく稀にエラーが生じ、『症例リストが不正です』といった旨のメッセージが出ます。この場合、至急データセンターまでご連絡ください。

7. 3 [Shuttle]の紛失など

[Shuttle]は CRF 原本という位置づけです。[Shuttle]を紛失されないようご注意ください。万一紛失された場合には、至急、データセンターまでご連絡下さい。

7. 4 USB メディアの中身について

[Shuttle](USB メディア)の中には、臨床試験データの他に様々な設定ファイルが格納されていますが、これらの変更や削除は絶対に行わないで下さい。[Shuttle]が起動しなくなります。

Ⅲ. 研究成果の刊行に関する一覧表

【書籍】

著者氏名	タイトル	編集者名	書籍名	出版社名	出版地	出版年	ページ
崎山美知代 牧本 敦	移植治療における関連 毒性の評価基準	高上洋一	細胞医療	医薬ジャー ナル社	東京	2004	106-112

【雑誌】

発表者氏名	タイトル	発表誌名	巻号	ページ	出版年
牧本 敦	臨床試験におけるデータマネージメントと 品質管理・品質保証	小児外科	36 卷 7 号	846-853	2004
牧本 敦	わが国における小児固形腫瘍の本格的臨床 研究確立のための基盤整備	小児科診療	67 卷 4 号	577-582	2004

IV. 研究成果の刊行物・別刷り

第Ⅳ章 同種細胞療法の合併症

3. 移植治療における関連毒性の評価基準

はじめに

同種造血幹細胞移植に伴う毒性のうち、化学療法や放射線療法によるもの、感染症やGVHDに伴うものなどを含む総称を移植関連毒性 (transplant-related toxicity : TRT) と呼び、そのうち移植前治療に直接伴うものを特に治療関連毒性 (regimen-related toxicity : RRT) と呼ぶ。これらは時に重篤となるため、タイムリーかつ適切に評価して迅速に対応することが、移植の臨床成績を左右するといっても過言ではない。また、新しい移植療法の開発や複数の治療法の比較といった学術的な目的においても、客観的な毒性評価法を確立することが必要不可欠である。

① Bearman 基準

従来の同種造血幹細胞移植の前処置である大量化学療法および全身放射線療法においては、血液毒性はほとんどの場合に grade 4 となるため、重症度評価の対象とするのは非現実的である。さらに感染症やGVHDに伴う毒性は別個に評価し、特に非血液毒性だけを選択して評価することが多い。従来この目的のために最も広く用いられてきた Bearman 基準(表1)は、このような移植治療に伴う毒性の特殊性を鑑みて、独自の毒性評価システムとして開発されたものであり、day 28 (例外として肺毒性については day 100)までの治療関連毒性を8つの臓器別に4段階に分けて評価している。Bearman らは、195例の骨髄移植を受けた患者において、前処置関連であると思われる有害事象(RRT)について本評価法を用いて毒性評価を行った¹⁾。その結果を以下に要約する。

① 単変量解析を行うと、非寛解期移植、15.75 Gy の TBI、同種骨髄、特に HLA 不一致移植、シクロスポリン・メトトレキサート併用の GVHD 予防をそれぞれ施された患者において、grade III～IV の RRT が有意に高頻度に認められた。

② ただし、上記の要素は互いに独立していないために多変量解析を行ったところ、同種移植患者においては TBI の線量だけが有意な危険因子であった。

③ 一臓器における最大の RRT grade を用いて層別化した day 100 の生存率は、それぞれ grade I : 84.8%、grade II : 74.2%、grade III : 10.5% であり、特に grade III 以上の RRT を示した患者において著明に低かった。

3. 移植治療における関連毒性の評価基準

表1 BearmanらによるRRT評価システム

	grade I	grade II	grade III
心毒性	治療を必要としない軽度心電図異常、もしくは無症状で見つかった胸部X線での心陰影拡大	治療を必要としかつそれに反応する中等度心電図異常、または治療は必要でないが持続監視を要する中等度心電図異常。またはジギタリス製剤、利尿薬に反応するうつ血性心不全	治療に全くもしくは少ししか反応しない重度の心電図異常。治療に全くもしくは少ししか反応しない心不全。もしくは50%を超える電位の低下
膀胱毒性	最後の化学療法から2日目以降に生じた肉眼的血尿で膀胱炎の自覚症状はなく、感染が原因でないもの	最後の化学療法から7日目以降に生じた肉眼的血尿。もしくは2日目以降に生じた自覚的な膀胱炎症状を有する肉眼的血尿で、感染が原因でないもの	明らかな血液を伴う出血性膀胱炎で、硬化性薬剤の注入や腎臓などの外科的処置による局所治療が必要とされるもの
腎毒性	基準値(通常前処置の開始前の最後の検査値)から2倍までのクレアチニン上昇	基準値の2倍を超えるが透析を必要としないもの	透析が必要となるもの
肺毒性	胸部X線で異常がなく、感染や心不全を原因としない呼吸困難。もしくは胸部X線で孤立性浸潤影や中等度の間質性変化を示すが、無症状で感染や心不全を原因としないもの	呼吸困難を伴う広範な局所浸潤影や中等度の間質性変化で、感染や心不全を原因としないもの。人工呼吸もしくは酸素マスク(>50%)を必要としないPaO ₂ (基準値より>10%)の低下で、感染や心不全を原因としないもの	人工呼吸による補助もしくは酸素マスク(>50%)を必要とする間質性変化で、感染や心不全を原因としないもの
肝毒性	ビリルビン値 \geq 2.0mg%から \leq 6.0mg%の軽度肝障害。基準値から体重>2.5%から<5%の増加で心臓が原因でないもの。もしくは前処置開始前最低値から2倍を超え5倍未満のsGOT増加	ビリルビン値>6mg%から<20mg%の中等度肝障害。前処置開始前から5倍を超えるsGOT増加。臨床的な腹水が画像診断された100mLを超える腹水。もしくは心臓を原因としない基準値から5%を超える体重増	ビリルビン値>20mg%を超える重度肝障害。肝性脳症。呼吸機能に障害を及ぼす腹水
中枢神経毒性	嗜眠傾向だが、簡単に覚醒し、覚醒後は見当識があるもの	嗜眠傾向で覚醒後も混乱を来しているもの。もしくは他の新たな客観的中枢神経症状が出現し、意識消失はなく、他の薬剤、出血、中枢神経感染では簡単に説明できないもの	痙攣発作または昏睡で、他の薬剤、出血、中枢神経感染では簡単に説明、証明できないもの
口腔粘膜毒性	麻薬性鎮痛剤の持続静注を必要としない疼痛、潰瘍	麻薬性鎮痛剤の持続静注(モルヒネ点滴)を必要とする疼痛、潰瘍	予防的挿管を必要とする重度の潰瘍または粘膜炎。もしくは挿管のあるなしにかかわらず明らかな誤嚥性肺炎に至るもの
胃腸毒性	連日、500mLを超えるが2,000mL未満の水様性下痢で、感染に関連しないもの	連日2,000mLを超える水様性下痢で、感染に関連しないもの。心血管系に影響を与えない肉眼的血便で、感染が原因でないもの。サブイレウスで感染に関連しないもの	鼻胃管による吸引もしくは外科的治療を要するイレウスで、感染に関連しないもの。もしくは心血管系に影響を与え輸血を必要とする出血性腸管炎

※ grade IV治療関連毒性は致死的なものと定義する。

④ 複数の臓器におけるそれぞれの最大gradeを総じた値(cumulative toxicities)を用いても、予後の層別化が可能であった。

⑤ 3臓器以上でgrade II以上の毒性を認めた患者は、そうでない患者に比して有意にday 100の生存率が低かった。

表2 NCI-CTC (ver. 2.0) – Bearman 基準の項目に従い抜粋

有害事象	0	1	2	3	4
左室機能	正常	症状はなく、安静時駆出率が治療前値から $\geq 10\%$ 、 $< 20\%$ の低下；短縮率が $\geq 24\%$ 、 $< 30\%$	症状はなく、安静時駆出率がLLN以下または安静時駆出率が治療前値より $\geq 20\%$ の低下；短縮率が $< 24\%$	治療に反応するCHF	重症または難治性CHFまたは挿管の必要あり
排尿痛	なし	軽い症状で処置を要さない	症状があるが治療により軽快する	症状があり治療によっても軽快しない	—
血尿	なし	顕微鏡的血尿のみ	時折の肉眼的出血；凝血塊なし	持続する肉眼的出血または凝血塊；カテーテルや器具の挿入また輸血を要する	緊急手術を要する大出血
クレアチニン	WNL	$> ULN \sim 1.5 \times ULN$	$> 1.5 \sim 3.0 \times ULN$	$> 3.0 \sim 6.0 \times ULN$	$> 6.0 \times ULN$
腎不全	なし	—	—	透析を要するが可逆性	透析を要し不可逆性
成人呼吸窮迫症候群 (ARDS)	なし	—	—	—	あり
呼吸困難	正常	—	労作時呼吸困難	通常の活動レベルでの呼吸困難	安静時呼吸困難また人工呼吸器を要する
低酸素血症	正常	—	労作時の酸素飽和度の低下	安静時の酸素飽和度の低下、酸素吸入を要する	陽圧呼吸補助 (CPAP) または補助換気を要する酸素飽和度の低下
肺炎・肺浸潤	なし	X線上的変化はあるが症状がないまたは症状はあるがステロイドを要さない	X線上的変化がありステロイドまたは利尿剤を要する	X線上的変化があり酸素吸入を要する	X線上的変化があり補助換気を要する
ALP	WNL	$> ULN \sim 2.5 \times ULN$	$> 2.5 \sim 5.0 \times ULN$	$> 5.0 \sim 20.0 \times ULN$	$> 20.0 \times ULN$
ビリルビン	WNL	$> ULN \sim 1.5 \times ULN$	$> 1.5 \sim 3.0 \times ULN$	$> 3.0 \sim 10.0 \times ULN$	$> 10.0 \times ULN$
GOT	WNL	$> ULN \sim 2.6 \times ULN$	$> 2.5 \sim 5.0 \times ULN$	$> 5.0 \sim 20.0 \times ULN$	$> 20.0 \times ULN$
GPT	WNL	$> ULN \sim 2.5 \times ULN$	$> 2.5 \sim 5.0 \times ULN$	$> 5.0 \sim 20.0 \times ULN$	$> 20.0 \times ULN$
体重増加	$< 5.0\%$	5.0～9.9%	10.0～19.9%	$\geq 20\%$	—
腹水	なし	症状がない腹水	症状があり利尿剤を要する	症状があり治療的穿刺を要する	生命を脅かす病態
意識レベル低下	正常	傾眠または鎮静（意識清明でない）；機能障害なし	傾眠または鎮静（意識清明でない）；機能障害はあるが日常生活には支障なし	感覚鈍麻（刺激に対する反応低下）または昏迷；覚醒困難；日常生活に支障あり	昏睡
口内炎・咽頭炎	なし	疼痛がない潰瘍、紅斑または病変を特定できない軽度の疼痛	疼痛がある紅斑、浮腫、潰瘍、摂食・嚥下可能	疼痛がある紅斑、浮腫、潰瘍、静脈内輸液を要する	重症の潰瘍、経管栄養、経静脈または予防的挿管を要する
下痢	なし	治療前に比し < 4 回/日の排便回数増加	治療前に比し4～6回/日の排便回数増加または夜間排便	治療前に比し ≥ 7 回/日の排便回数増加または失禁または脱水に対する静脈内輸液を要する	集中治療を要する病態または循環動態の虚脱

CHF：慢性心不全

以上より、従来の骨髄破壊的な方法による同種造血細胞移植では、Bearman 分類は移植関連死亡の予測に有効であることが証明された。

② NCI-CTC

一方近年、従来の移植に比べて毒性としての臓器障害が少ないといわれる RIST が確立され、高齢者や臓器合併症をもった患者に対する積極的な応用が期待されている。この RIST を受けた患者における毒性評価に、通常の骨髄破壊的移植を想定した尺度として開発された Bearman 基準を用いる妥当性は確立していない。そこで、Bearman 基準に比べて軽症な毒性を評価するのに優れていると考えられ、通常の化学療法領域で普及している National Cancer Institute の Common Toxicity Criteria (NCI-CTC [ver.2.0])(表 2)²⁾ を用いて、RIST の毒性評価を試みた結果を紹介する³⁾。

国立がんセンター中央病院において 1999 年 9 月から 2002 年 4 月までに RIST を施行した 86 例において、前処置開始から移植後 28 日までを対象期間とし、Bearman 基準と NCI-CTC (ver. 2.0) による毒性評価を retrospective に行い、患者予後との相関を解析した。表 3 に患者背景を示す。Bearman 基準を用いた治療関連毒性では、grade III 以上の毒性は中枢神経 2 例、肺 4 例、膀胱 2 例と非常に少なく、grade II 以上の毒性は口内炎、肝毒性において多く認められた(図 1)。一方、NCI-CTC (ver. 2.0) では、grade 3 以上の毒性は腸管 (8 例)、肝 (27 例)、肺

表 3 患者背景

総患者数	86 人	disease states	
年齢中央値(範囲)	51 (4 ~ 67) 歳	low risk / high risk	53 / 33 人
男性 / 女性	57 / 29 人	診断名	
conditioning regimen		ALL	2 人
Flu-based / 2-CdA-based	64 / 22 人	AML	26 人
ATG + / -	49 / 37 人	CML	5 人
TBI + / -	3 / 83 人	TLBL	2 人
GVHD 予防		ATL	3 人
MTX あり / MTX なし	16 / 70 人	MDS	11 人
ドナーの背景		NHL	16 人
related / unrelated	81 / 5 人	その他	3 人
matched / mismatched	69 / 17 人	固形腫瘍	18 人

注：disease state は移植にあたっての risk と定義し、low risk は MDS の RA, ALL, AML の第 1, 第 2 寛解期, CML の CP 1,2, リンパ腫の第 1, 第 2 寛解期, 部分寛解期, 固形腫瘍とした。

Flu：フルダラビン, 2-CdA：クラドリピン, ATG：抗胸腺細胞グロブリン, MTX：メトトレキサート, ALL：急性リンパ性白血病, AML：急性骨髄性白血病, ATL：成人 T 細胞白血病, CML：慢性骨髄性白血病, MDS：骨髄異形成症候群, NHL：非 Hodgkin リンパ腫, TLBL：T 細胞性リンパ芽球性リンパ腫

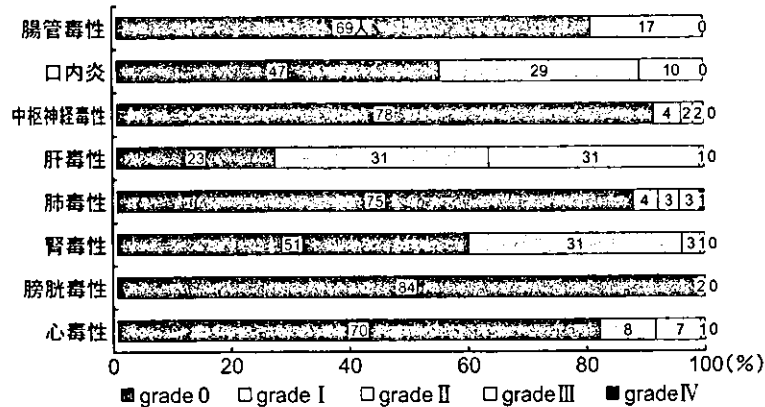


図1 Bearman 基準を用いた RRT の評価

grade III以上の毒性は少なく, grade II以上は口内炎, 肝毒性において多く認められた。

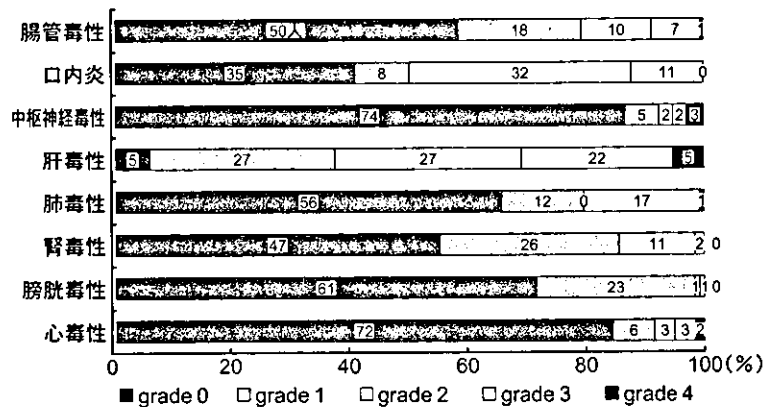


図2 NCI-CTC (ver. 2.0) を用いた RRT の評価

grade 3以上の毒性は腸管毒性, 肝毒性, 肺毒性において多く認められた。

(18例)において多く認められた(図2)。grade 3以上の毒性を生じた患者は, Bearman 基準では5例(6%)であるのに対して NCI-CTC (ver 2.0)では43例(50%)であった。つまり, 同一患者集団では Bearman 基準において毒性が低く評価される傾向が判明した。毒性別に生存率をみると, 両基準とも移植関連毒性の重症度と予後との間に有意な相関を認めている(図3, 4)。

図5では, 移植関連死亡を呈した11例についての検討結果を示した。Bearman 基準で grade III以上を示した症例は11例中3例であったが, NCI-CTC (ver. 2.0)で grade 3以上の毒性は11例中10例であり, NCI-CTCの方が移植関連死亡をより鋭敏に予測し得ることが示唆された。

さらに生存期間に影響を及ぼす因子を調べたところ, ドナー背景と移植関連毒性の重症度が

3. 移植治療における関連毒性の評価基準

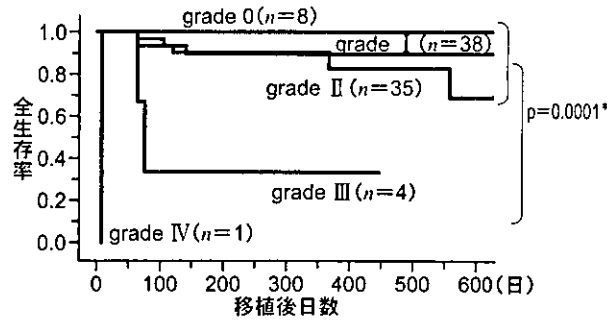


図3 Bearman 基準を用いた毒性別の生存率
毒性の grade と予後との間に有意な相関を認めた。

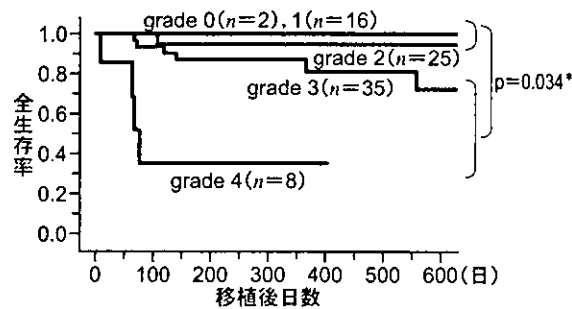


図4 NCI-CTC (ver. 2.0) を用いた毒性別の生存率
毒性の grade と予後との間に有意な相関を認めた。

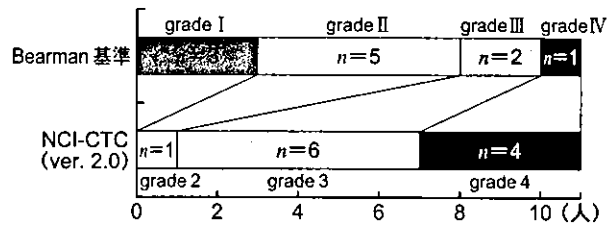


図5 移植関連死亡を呈した 11 例における毒性評価
移植関連死亡 11 例のうち、Bearman 基準と NCI-CTC (ver. 2.0) で grade 3 以上の毒性はそれぞれ 3 例と 10 例であった。

統計学的に有意に関連していた (表 4)。

移植関連死亡の予測という点から毒性評価の閾値を検討すると、従来臨床的に問題とされている grade III 以上で、Bearman 基準は特異度が高く感度は低かった。逆に NCI-CTC (ver. 2.0) の grade 3 以上では、感度は高いものの特異度は低かった (表 5)。

おわりに

本研究における RIST の移植関連死亡は 12.8% (86 人中 11 人) であり、従来型の骨髄破壊的

表4 生存期間に影響する因子(比例ハザードモデル, 多変量解析)

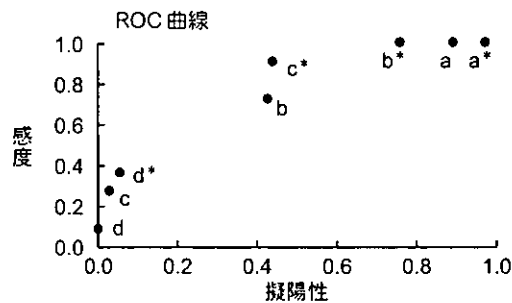
	ハザード比	95%信頼区間	p値
血縁 or 非血縁	7.502	1.715 ~ 32.814	0.0074 *
HLA 一致 or 不一致	3.838	1.143 ~ 12.885	0.0295 *
毒性 (NCI-CTC grade 3 ~ 4)	2.975	1.208 ~ 7.326	0.0177 *

ドナー背景と毒性の grade が有意に生存期間に影響していた。

*: 統計学的有意

表5 移植関連死亡の予測における毒性評価基準の閾値の検討

		毒性	感度	特異度
Bearman 基準	a	grade I 以上	1.000	0.107
	b	grade II 以上	0.727	0.573
	c	grade III 以上	0.273	0.973
	d	grade IV	0.090	1.000
NCI-CTC (ver. 2.0)	a*	grade 1 以上	1.000	0.027
	b*	grade 2 以上	1.000	0.240
	c*	grade 3 以上	0.909	0.560
	d*	grade 4	0.364	0.947



移植と比べて低値であったものの, RIST のさらなる治療成績改善のためには, より優れた毒性評価法に基づいた治療法の改善が求められる。今後, RIST を含めて多様化する同種造血幹細胞移植の移植関連毒性の評価には, それぞれの移植手技に応じた, 高感度かつ擬陽性が少ない新しい評価方法の開発が必須であると考えられる。

(崎山美知代, 牧本 敦)

文献

- 1) Scott I, Bearman et al : Regimen-Related Toxicity in Patients Undergoing Bone Marrow Transplantation. J Clin Oncol 6 (10) : 1562-1568, 1988
- 2) 日本臨床腫瘍研究グループ: NCI-CTC の日本語訳 JCOG 版について. 癌と化学療法 26:1084-1144, 1999
- 3) Sakiyama M, et al : Regimen-related toxicity following reduced-intensity stem-cell transplantation (RIST) : comparison between Seattle criteria and National Cancer Center Common Toxicity Criteria (NCI-CTC) version 2.0. Bone Marrow Transplantation 34 : 787-794, 2004

■ 特集 小児科・小児外科領域における臨床試験

臨床試験におけるデータマネジメントと品質管理, 品質保証

牧本 敦*

はじめに

わが国の臨床試験グループにおいて、もっとも確立したデータマネジメントシステムを有する日本臨床腫瘍研究グループ(JCOG)データセンター長、福田治彦氏は、臨床試験におけるデータマネジメントを次のように定義する。すなわち、「研究において、データエラーを低くコントロールし、研究を科学的、倫理的、効率的に行って正しい結論を導くための技術体系」であり、臨床医学、生物統計学、情報工学の3つの学問に基礎を置く応用化学である、¹⁾。本稿では、このような新しく、かつ複雑な学問体系であるデータマネジメントが、実際の臨床試験において、どのくらい重要で、どのように役立っているのか、また、どうあるべきなのか、について、品質管理・品質保証の観点から解説する。

I. 何故データマネジメントが必要か？

この答えは、医学者であれば容易に想像できると考えられる。データマネジメントが不正確であれば臨床試験の結果が歪められ、試験治療の意義や妥当性が不正確になり、試験結果を根拠(エビデンス)とした「根拠に基づく医療(EBM)」が歪められ、将来の患者に不適切な治療を施す原因をつくりだすと同時に、当該試験結果をベースにした将来の治療開発が誤った方向へ進み、結果的に数多くの将来の患者に不適切な治療を施すことにつながりかねないのである(図1)。

この事実をより具体的に示すために、データの

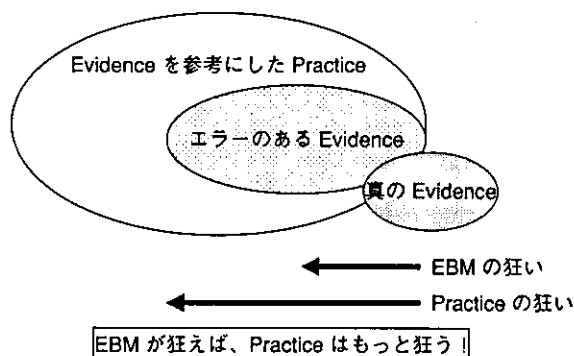


図1 EBMと臨床実践(Practice)

エビデンスを導き出した臨床試験の結果そのものがエラーを含んでいるため、その取り扱いを誤ると、間違った臨床実践につながる危険がある。

欠損や誤差が、どのように結果に影響するかを、臨床試験の第I相、第II相、第III相のそれぞれの評価項目(エンドポイント)について示す。なお、本稿では悪性腫瘍に対する抗がん剤の臨床試験で汎用される試験デザインを例に挙げて説明するため、他分野の臨床試験に当てはめることが必ずしも適切でない場合があることを付記しておく。

1. 第I相試験(用量設定試験)

通常、悪性腫瘍に対する抗がん剤の第I相臨床試験では、3例コホート法が用いられる。この方法では、十分に安全な用量を開始用量と設定して、その用量に最初の3例を登録する。3例とも重篤な有害事象(用量規制毒性:DLT)の出現なく治療を遂行できるか、または3例中1例にDLTが出現しても、さらに3例を追加した6例中1例にしかDLTが出現しなかった場合に、次の3例は増量した用量で治療を行う。1つの用量において、6例中2例(試験によっては3例)以上がDLTを

* 国立がんセンター中央病院小児科
〔〒104-0045 東京都中央区築地5-1-1〕

示すまで増量を繰り返して、最大耐用量 (MTD) と推奨用量を決定する。

この時、有害事象データにエラー (誤差) が生じるとどうなるであろうか? 例えば、よく起こると想像されるエラーは、有害事象の過小評価、である。抗癌剤の臨床試験における毒性評価には米国国立癌研究所 (NCI) の共通毒性基準 (common toxicity criteria [CTC] version 2²⁾ または最新の common terminology criteria for adverse events version 3) が用いられるが、臨床現場で常にこの規準を用いている医師は数少ないため、特別かつ細心の注意を払わなければ正確な判断はできない。このため、臨床現場では、患者に DLT となりうる重篤な有害事象が生じていたにもかかわらず、主治医がそれと認識しないうちに自然回復し、カルテにも事実が記載されないことはしばしばある。このようなことが臨床試験のセッティングで起こると、症例報告書 (CRF) にも記載されず、報告されなかったために次の増量が行われ、推奨用量が過大評価されることにつながり、非常に危険である。欧米でも日本でも、第 I 相試験を行う施設の選定が重要となるのは常識である。

2. 第 II 相試験 (有効性・安全性試験)

通常、悪性腫瘍に対する抗がん剤の第 II 相臨床試験では、第 I 相試験で決定された推奨用量を用いて、限定された対象疾患に対する有効性と安全性を試験する。悪性固形腫瘍の場合は、奏効率 (完全寛解と部分寛解を合わせた率) を有効性のエンドポイントとすることが多い。この時、試験計画書 (プロトコール) には、CT または MRI の評価時期が示されており、さらに進行病変 (PD) と判断されれば、もはや有効ではない試験治療を中止する、という規定も示されていることが多い。

この CT などによる腫瘍径評価およびその測定のための検査にエラーが生じた場合の実例を、Green ら³⁾ の教科書 "Clinical Trials in Oncology, 2nd edition" から引用する。初回評価で、長径 2.0 cm だった肺病変が、2 回目の CT ではいったん縮んで 0.5 cm となり、3 回目の評価でも 0.5 cm、4 回目の評価で再び増大して 1.5 cm となった場合を仮定する。すべての検査が正確にされれば、こ

の評価は 2 回目の評価時点から部分寛解 (PR) が始まり、3 回目の評価で確認され、4 回目の評価で PD が確認されて、試験治療は終了する。すなわち、4 回目の評価時点で、本患者は試験治療からの利益を受けていない、と判断されるのである。もし、4 回目の評価 CT が何らかの理由で撮影できなかった場合、本患者は少なくとも 5 回目の検査まで無効な治療を継続されることになる。また、2 回目と 3 回目の検査が行えなかった場合は、さらに真実を歪め、今度は 4 回目の評価時点で部分寛解 (PR) と評価してしまうのである。これらのエラーは奏効率の誤りとして現れるだけではなく、エンドポイントに無増悪生存割合などを含む場合には、その数値の誤りにもなる。結果、有効でない治療を有効と判断したり (α 誤差)、有効な治療を無効と判断したり (β 誤差) することになる。

3. 第 III 相試験

第 III 相試験では、疾患によっても異なるが、3 年または 5 年の生存割合がエンドポイントとなることが多い。この解析のために最も汎用されているのが Kaplan-Meier 法による生存曲線であることは周知のとおりである。

前述の Green ら³⁾ の教科書の例を図 2 に示す。試験開始後 4 年時の解析で、10 例の患者のうち 5 例が追跡不能で打ち切りされた生存曲線 (図 2 C) を挙げ、5 例が打ち切りの直後に死亡した場合の生存曲線 (図 2 D) と、5 例が 4 年後まで生存していた場合の生存曲線 (図 2 A) との違いを図示している。グラフから読み取れる 4 年生存割合は、A 50%、C 19%、D 0%、となり、生存期間中央値は A 4 年以上、C 33 カ月、D 22 カ月、となる。これは極端な例ではあるが、追跡データの誤りが結果に与える影響を考慮すると、1 例の追跡データであっても決して侮れないことが理解されるであろう。また、同書³⁾ には、切除可能なリンパ節転移のある乳がんに対する術後化学療法の臨床試験、Southwest Oncology Group (SWOG) 研究 7436 の実例が挙げられ⁴⁾、素データの解析では認められた統計学的有意差 (logrank $p=0.03$) が、適格性の補正と未報告イベントの追加というデータクリーニングにより消失 (logrank $p=0.08$) し、そ

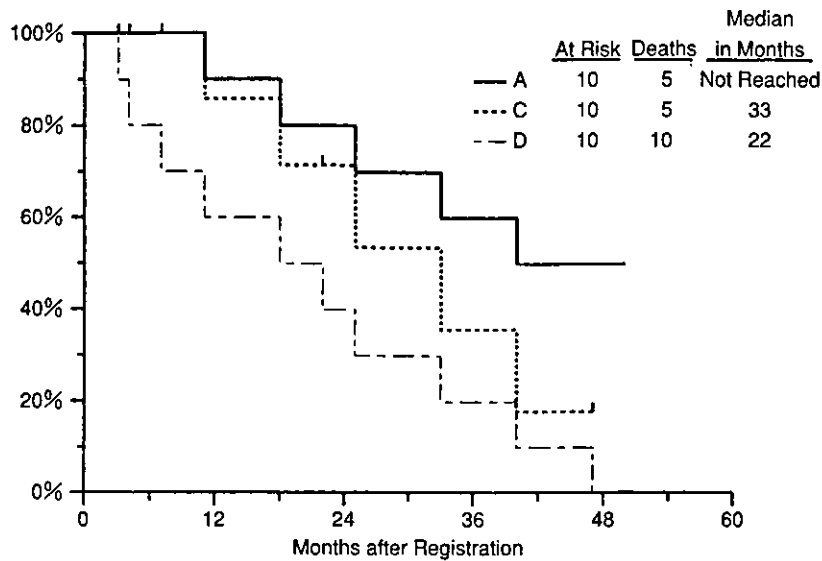


図 2 半数の患者が追跡不能のため打ち切りになった生存曲線のバイアス (Green ら, 2002 より引用, 一部改変)⁹⁾

A : 5 人の追跡不能患者が, 実は全員生存していた場合の曲線

C : 5 人の追跡不能患者が, 最終観察日で打ち切りとされた場合の曲線

D : 5 人の追跡不能患者が, 最終観察日の直後で全員死亡した場合の曲線

の後のフォローアップデータの追加によりさらに素データとの相違が増大した (logrank $p=0.17$) 事実により, データの品質管理 (QC:クオリティコントロール) と品質保証 (QA:クオリティアシユアランス) の重要性が説明されている。

II. 臨床試験の品質管理・品質保証とその法的要件

1. Good Clinical Practice (GCP) と品質管理・品質保証

これまでの議論で明らかのように, データマネジメントが不正確であれば臨床試験の結果が歪められ, 結果的には, 将来の患者に不適切な治療を施す原因をつくると同時に, 当該試験結果をベースにした将来の治療開発が誤った方向へ進む。このようなことが起こらないように, 臨床試験のデータマネジメントにおいては適切な品質管理と品質保証が求められる。これは, 三極合同国際ハーモナイゼーション (ICH)-GCP⁵⁾ の第 2 章「ICH-GCP の原則」第 13 項に明確に示されており, 臨床試験を行う組織が, その組織に見合った適切な品

質保証システムを構築・維持することを保証し, これによって, 臨床試験に参加する被験者の人権保護を確実にを行うよう定められている。ICH-GCP において, 品質管理と品質保証にかかわる記載をまとめたものを表に示す。これを見ても, ICH-GCP がいかに品質管理・品質保証を大切に考えて作られているか, が理解されよう。

2. 治験と臨床試験における法的要件の相違

わが国において, 薬剤の承認申請のデータを提供するための臨床試験である「治験」は, 法的要件として「医薬品の臨床試験の実施の基準に関する省令 (平成 9 年厚生省令第 28 号)」および「医薬品の臨床試験の実施の基準に関する省令の一部を改正する省令 (平成 15 年厚生労働省令第 106 号)」, すなわち省令 GCP に従うことが必要であり, これら「治験」における品質管理・品質保証は製薬会社主導で厳格に行われてきた。わが国の省令 GCP は ICH-GCP の単純な翻訳ではなく, 製薬会社主導の治験に適用する目的で作られたため, 欧米で医師の行う自主臨床試験にも適用される ICH-GCP に比べるとさらに縛りが強い記載

表 ICH-GCP における品質管理・品質保証に係る記載

2. 原則	2.10 臨床試験に関するすべての情報は、正確な報告、解釈および検証が可能ないように記録し、取り扱い、かつ保存されるべきである
	2.13 試験のあらゆる局面の質を保証する一定の手順を持つシステムが、実行されるべきである
5. 試験依頼者	5.1 試験の品質保証および品質管理
	5.2 開発業務受託機関 (CRO)
	5.5 試験の管理、データの取り扱い、および記録の保存
	5.18 モニタリング
	5.19 監査

となっている。

一方で医師の行う自主臨床試験（研究者主導臨床試験と呼ぶこともある）に関しては、わが国に拘束力を持つ法的要件は存在せず、そのために厳格な「治験」とは似ても似つかぬレベルの臨床試験が多々行われ、科学的に正しい結論を出せないばかりか、被験者である患者の人権問題にまで発展したケースもあった⁹⁾。2003年7月に「臨床研究に関する倫理指針（平成15年厚生労働省告示第255号）」が施行されたが、かなりの内容をヘルシンキ宣言と共有する粗いガイドラインに過ぎず、当面、自主研究の品質管理・品質保証はそれを行う研究者の自助努力に頼らざるを得ない。将来の患者のために適切な医療を提供するためのデータを蓄積し、エビデンスを創成する行為という点で、研究者は、治験と臨床試験の間に基本的な相違はないことを知り、臨床試験においてはできるだけエラーの少ない正確な結果を残すように、また、そのための組織およびシステム構築を行うように、最大限の努力を払うべきである、という共通認識を形成して欲しい。言い換えれば、治験であるなしにかかわらず、臨床試験を行う研究者は、ヘルシンキ宣言と臨床研究に関する倫理指針を遵守し、ICH-GCPの精神を理解し、それらを最大限尊重した試験を実行するよう努力すべきなのである。

3. GCPとISO 9000

製造業等に対する品質保証システムの国際規格であるISO 9000は、ICH-GCPの基礎となった欧州の統一GCPに大きな影響を与えた⁷⁾。この歴史から、現行のICH-GCPおよび省令GCPで記載

されている品質管理・品質保証の根源はISO 9000にあると考えられ、ISO 9000の要素をどのように臨床試験の品質保証システムの中に取り入れるかという議論はわが国でも盛んに行われた^{8,9)}。

ISO 9000において、品質管理・品質保証の目的は良い「製品」を顧客に届けること、または確実に仕事を行うことである。「製品」とは、「活動またはプロセスの結果」と定義され、臨床試験において「製品」はデータそのものであるといえる。確実な仕事をして、確実な製品を提供する手段として、どのような組織およびシステムを作って、どのようにそれを動かすか、ということが大切であり、これがISO 9000のいう「品質システム (quality management system)」にほかならない¹⁰⁾。

ISO 9000における品質システムは以下の5つのレベルに集約できる。すなわち、W. Edwards Demingの唱えたPDCAサイクル(plan, do, check, act)を管理の基本とした枠組みである。さらに、このような循環がうまく機能していることを実証することによって、顧客に対して信頼感を供する、このための体系的な活動の総体が品質システムである。

- (1) 用いているプロセスの文書化。
- (2) 文書化した手順に基づくプロセスの実行。
- (3) 実行されたことの記録。
- (4) 事前に決めた品質基準と結果（成果物）の比較・プロセスの有効性の検討。
- (5) さらなるプロセスの改善。

実はこのようなプロセスは、医師・医療従事者にとって日々の患者診療においても生かされるべ

きものであろう。国立大学病院の独立行政法人化に伴って、医療分野一般における品質システム導入も加速される見込みであり、病院 QC サークルや ISO 委員会等が活動していると聞いている。これらの具体的活動に関しては本稿の守備範囲を逸脱するので割愛するが、臨床試験において「製品」であるデータは、試験計画に基づく実地診療の積み重ねによって得られることを考えれば、これらの活動の品質を高めることが、将来的には医学・医療を飛躍的に発展させる原動力になるものと期待している。

III. データマネジメントと品質管理・品質保証

1. 治験の場合

まず、薬剤の承認申請のデータを提供するための臨床試験である「治験」においてその品質管理・品質保証を考えると、ISO 9000 に基づけば、その責任は治験依頼者である製薬会社にある。ところが、実際に品質システムとしてみた場合には、「製品」を作り出す医療機関と品質システム責任者の製薬会社（またはデータマネジメントを委託された第三者）を一体の構造として考える必要が生じる。責任者である製薬会社は、全く別団体である医療機関の品質を、モニタリングという活動によってコントロールしようとしている⁹⁾。治験においては、個々の医療機関における実地臨床の品質を上げる医学的努力と、製薬会社等が行うモニタリング活動と医療機関の協力によるデータの品質管理を併せたものが品質管理 (QC) であり、このような全体のシステムがうまく働いていることを第三者的に審査し、保証する活動 (監査) が品質保証 (QA) だと考えると理解しやすい。

2. 医師の自主臨床試験(研究者主導臨床試験)の場合

一方で、自主臨床試験の場合は、品質管理・品質保証の責任はすべて医師・研究者にある。この状況では、日常診療に多忙な医師がすべての試験においてその品質管理・品質保証を行うことなど、現実的には不可能である。これが、わが国の臨床試験における諸問題の根源だと筆者は考えている。臨床試験を行う研究者が、ヘルシンキ宣言と臨床研究に関する倫理指針を遵守し、ICH-GCP

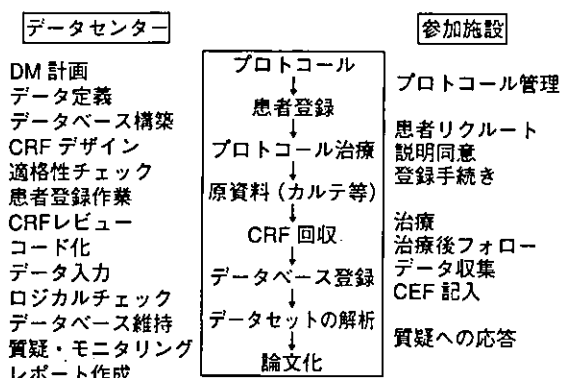


図3 データマネジメント・プロセスとデータセンターの役割 (福田, 2003より引用, 一部改変)¹¹⁾

データマネジメントの各段階において、データセンターと参加施設との共同作業により、試験データの品質管理を行いながら、臨床試験が進行する。

を最大限尊重した臨床試験を実行するためには、臨床試験の品質管理・品質保証に責任を持つ組織とシステムが不可欠である。この考えによって、平成14年度厚生労働科学研究補助金(効果的医療技術の確立推進臨床研究事業:小児分野)「小児肉腫に対する至適治療法確立のための臨床試験とその基盤整備に関する研究(主任研究者:牧本敦)」の1プロジェクトとして、国立がんセンター中央病院内に「小児がんデータセンター」を設立した。これは、冒頭に紹介した JCOG データセンターのリソースが、成人の各種がんの臨床試験で飽和状態であったため、上記補助金を用いて小児がん分野で可能な範囲から活動を開始しようという、小児専門医の自助努力の事始めである^{12,13)}。

3. データマネジメントの各段階と品質管理・品質保証

上述の小児がんデータセンターにおけるデータマネジメントの手法は、その大部分を JCOG データセンターから直接、間接に学び取ったものである。JCOG データセンターのデータマネジメントの概要については文献1)に、小児がんデータセンターに関しては文献13)に記載されているので、それらを参照されたい。この項では、図3に示したデータマネジメント・プロセスの各段階について、その品質管理の実際について説明する。

1) プロトコールおよび症例報告書 (CRF) 作成と施設への説明

この段階は、今までの自主臨床試験ではもっとも軽視されていたと言わざるを得ない。試験のデザインと仮説、プロトコール治療とエンドポイントの明確化、評価の標準化、等について、計画の初期段階から、データセンター、生物統計家および研究事務局が協力して、プロトコールを練り上げることが品質管理の第一歩だと考えている。プロトコール確定後に、データセンターのデータマネージャー (DM) が中心となって CRF を作成する。DM はデータマネジメント計画書を初めとした各種手順書を整備し、試験開始に備える。全準備が整った後に、全参加施設を集めてキックオフミーティングを行い、試験計画、CRF への記入方法や送付方法等について説明を行う。

2) 症例登録

臨床試験プロトコールが、各施設倫理委員会の承認を得た後に、患者登録が開始される。施設医師は、患者がプロトコール内にある「適格規準」を満たし「除外規準」に当てはまらないことを確認してから、「症例登録票」に記載し、データセンターへ Fax 送付する。データセンターでは、判読できない文字や不足項目がないかをチェックした後、データベース上の症例登録システムを使用して、プログラムによる適格性判定を行う。これによって、不適格症例が登録されるというエラーを最小化する。

3) プロトコールに基づく治療、被験者の観察、データ収集

患者登録がされた後は、プロトコール治療が開始されるが、通常、医師は臨床試験の被験者とそれ以外の患者を区別して扱うことは少ないと考えられる。そのため、タイムリーな観察やそれに基づく CRF の記載と回収を行うためには、データセンターによって進捗管理がなされることが理想的である。小児がんデータセンターでは、進捗管理プログラムを採用し、化学療法の治療コースを単位とした visit 型の CRF を定期的に送付し、その内容のチェックと施設へのフィードバックをこまめに行うことにより、治療の初期からデータの品質管理を行っていく。回収された CRF の記載

が正しいか、記載内容に矛盾はないか、有害事象などの追加情報は必要ないか、等を目視チェックとロジカルチェックで行う。これらは前もって定めたチェックリストに基づいて行われる。記載内容の不明点や矛盾点は、発生源である施設の医師・研究者に問い合わせの上、修正を行う。

4) データ入力

あらかじめ当該試験用に構築したデータベース上に、あらかじめ定めた入力マニュアルに基づいてデータ入力を実施されるが、正確性を高めるために、2名の独立した入力者によるダブルエントリーを行う。さらに、別の DM によって入力データの一致性が確認される。

5) モニタリング

JCOG では、データセンターが電話による問い合わせ形式で行う「中央モニタリング」を行っている。これに対し、小児がんデータセンターでは、施設の第一登録症例に関しては施設訪問による「施設訪問モニタリング」を併用することを考えて、現在手順を整備中である。また、試験の進行中に問題のあるケースが発生した場合には、監査とモニタリングを兼ねた施設訪問を行う予定である。

6) 監査

上記 1)～5) の各段階において、品質管理 (QC) が行われる。これらの各品質管理がシステムとして有効に機能しているかどうか、を第三者評価することが監査、であり品質保証 (QA) である。ところが、具体的な監査として何を行うか、という細かな規定は GCP 上にはなく、監査の方法は、治験依頼者や臨床試験の責任者が、自己の責任で決定する事項であると考えられる¹⁴⁾。小児がんデータセンターにおける監査は、当面 JCOG 監査マニュアル¹⁵⁾を参考に行う予定であるが、今後、医師主導治験プロジェクト等の経験を通じて、より効率的かつ効果的な監査の方法を研究し、模索していくことになるであろう。

おわりに

臨床試験におけるデータマネジメントの重要性を、具体例と GCP や ISO 等の法的要件を絡めながら解説した。本稿の執筆初期には、中央データ

センターの品質管理・品質保証のみに着眼していたが、ISO 9000 等に関する論文を精読すると、医療および臨床試験における品質管理・品質保証は、決して中央だけの努力では成し遂げられないことに改めて気づかされた。真に医学の進歩に貢献しうる臨床試験を行うためには、その主たる担い手である医師・研究者が、単に臨床試験のデータのみならず、日々の臨床実践の内容そのものの品質を再度考え直し、その改善のために真摯に努力していく姿勢が必要だと強く感じる。わが国の臨床試験の基盤整備を進めるためには、中央データセンターにおいて品質管理・品質保証システムを確立すると同時に、臨床試験に参加する施設・医師・研究者が、自身の臨床実践の品質管理を行うという二重の自助努力が必要となるであろう。

文 献

- 1) 福田治彦：研究者主導のがん多施設共同臨床試験におけるデータマネージメント。小児臨床薬理学会雑誌 16：76-82, 2003
- 2) 日本臨床腫瘍研究グループ (訳)：National Cancer Institute-Common Toxicity Criteria version 2.0～日本語訳 JCOG 版・第2版～。http://www.jcog.jp/SHIRYOU/CTCJ_v20revised2001_09.pdf
- 3) Green S, Benedetti J, Crowley J：Clinical Trials in Oncology 2nd ed. Chapman & Hall/CRC, Boca Raton, USA, 123-163, 2002
- 4) Rivkin SE, Green S, Metch B, et al：Adjuvant CMFVP versus melphalan for operable breast cancer with positive axillary nodes：10-year results of a Southwest Oncology Group study. J Clin Oncol 7：1229-1238, 1989
- 5) International Conference of Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH)：Guideline for Good Clinical Practice. http://www.nihs.go.jp/dig/ich/efficacy/e6/e6step4.pdf
- 6) 仲正昌樹, 打出喜義, 仁木恒夫：「人体実験」と患者の人格権—金沢大学付属病院無断臨床試験訴訟をめぐって—, 御茶の水書房, 東京, 2003
- 7) 伊藤秀晴：ICH-GCP に要求される治験の品質保証とは。技術情報協会 (編)：信頼性向上をめざした治験における QC 業務と QA 対応, 3-13, 技術情報協会, 東京, 2002
- 8) 臨床評価刊行会：ISO 9000 について—矢野友三郎氏との討論—。臨床評価 28：5-31, 2000
- 9) 藤田 剛：ISO 9000 と治験管理の新しい試み。臨床評価 28：51-72, 2000
- 10) 加藤洋一：ISO 9000 による品質保証の基本的な考え方。臨床評価 28：33-50, 2000
- 11) International Organization for Standardization：Guidance on the process approach to quality management systems. http://www.iso.ch/iso/en/iso9000-14000/iso9000/2000rev9.html
- 12) 牧本 敦：小児悪性腫瘍領域における取り組み。小児臨床薬理学会雑誌 16：59-63, 2003
- 13) 牧本 敦：わが国における小児固形腫瘍の本格的臨床研究確立のための基盤整備。小児科診療 67：577-582, 2004
- 14) 畑 誠一：治験の信頼性を確保する QA 監査の留意点。技術情報協会 (編)：信頼性向上をめざした治験における QC 業務と QA 対応, 159-221, 技術情報協会, 東京, 2002
- 15) JCOG 監査委員会 (編)：JCOG 監査マニュアル version 3.1. http://www.jcog.jp/SHIRYOU/JCOG_auditmanu_ver3.1.pdf

Quality Control and Quality Assurance for Data Management in Clinical Trials

ATSUSHI MAKIMOTO

Division of Pediatric Oncology, National Cancer Center Hospital

Key words：Quality control, Quality assurance, Data management, Clinical trials, Good clinical practice (GCP).
Jpn. J. Pediatr. Surg., 36(7)：846～854, 2004.

Quality control (QC) and quality assurance (QA) of data in clinical trials are so important because potential bias could change not only the results of a trial but also the future direction of development of therapies. Guidelines for good clinical practice

(GCP) include many regulations about QC/QA, which were transferred from the International Organization for Standardization (ISO). Improving procedures in data management in accordance with these guidelines is important for both central data centers and local institutes that perform qualified clinical trials which could produce firm evidence and could finally contribute to improved medical practice.

用語解説

cohort study

何らかの共通要因, 例えば同じ疾患, 同じ曝露要因など, を持った集団 (コホート) を追跡し, 死亡, 疾病の発生などのイベントの発生を観察し, イベントと要因の関連を明らかにする疫学の研究方法の一つ。前向きコホート研究では, 対象者の曝露要因を研究者が調べるところから研究が始まる。これに対して後ろ向きコホート研究では, すでに曝露がおこってしまった後で, 研究者が後ろ向きにその状況を調べ, さらにその集団を追跡調査することで, 疾病の発生を確認する。コホート研究は, 通常は観察研究の範疇である。しかし, 十分な介入試験計画や統計学的評価計画を立てずして行ってきた旧来の臨床研究 (または治療研究) は, 一定の治療を施された一定の疾患群を観察している, と考えることもできるため, コホート研究の一種ととらえることができる。

prospective study

研究計画がデータの発生よりも前に行われ, すべてのデータを前向きに収集していく研究方法の総称。観察研究の中では, 前向きコホート研究が単に「前向き研究」と呼ばれることがある。これに対する後ろ向き研究は, 症例対照研究である。前向きコホート研究では, 最初に健康な人の生活習慣 (喫煙・飲酒・食生活) などを調査し, この集団を「前向き」に追跡調査して, 後から発生する疾病を確認する。これに対して, 症例対照研究では, 最初に疾病にかかった人 (症例) を選び, 次にその人達と性別や年齢などのそろった健康人 (対照) を選んで, 両者の生活習慣などを過去にさかのぼって「後ろ向き」に調査する。EBMの偏重によって, 「前向き研究」という用語から, 介入的研究である臨床試験をイメージする臨床医が多いが, 必ずしもそうではない。「前向き研究」という用語のとらえ方は研究者によって個人差があるため, より詳細な用語を用いるよう注意が必要である。

モニタリングと監査

ICH-GCPによると, 広義のモニタリングは, 試験の進行状況を監督し, 試験がプロトコール, 標準業務手順書, GCP およびその他の規制要件に従って, 実施, 記録および報告されることを保証する行為, とされる。狭義のモニタリングは, 臨床試験のデータ管理において, 定期的にデータの収集状況やその内容についてチェックを行うことである。具体的には, 実際にはなされていない報告内容の抽出, 原資料との記載不整合の訂正, 脱落例や有害事象の発生パターンに基づくプロトコール修正, 症例集積率に基づく研究期間の見直し等, さまざまな役割がある。監査には, 施設監査, プロセス監査, 記録監査, コンピュータバリデーション監査, ドキュメント監査など, さまざまな種類のものを含むが, 一般には臨床試験において行われているデータ管理, モニタリングを含めた品質管理プロセスが全体として妥当で, 体系的に信頼できることを第三者的に確認する行為である。

臨床試験における品質管理と品質保証

臨床試験において「製品」はデータであり、この品質を管理するために、データセンターにおいては、プロトコル作成から、患者登録、症例報告書の管理、システム管理、統計解析、報告書作成など、多段階の活動を行っている。これらひとつひとつの活動は標準業務手順書に従って行われ、各作業は記録され、記録は保管され、モニタリングの対象となってチェックされ、データ品質が保持される。これらの活動のひとつひとつが「品質管理」である。一方、この「品質管理」を含むすべてのプロセスが全体として正しく行われ、信頼できるデータおよび正しい結論を導いている、ということを確認、保証する活動が「品質保証」である。通常、第三者による監査の結果に基づいて「品質保証」が確立される。

効果安全性評価委員会

試験期間中に統計センターで定期的に行われる中間解析結果を第三者の立場から審査し、臨床試験が適正に行われることを監視する委員会。役割は以下のとおり。多段階試験法による臨床試験や、比較的長時間を要する臨床試験において、試験終了前に判定できる有意な効果をできるだけ早期に発見し、被験者に不利な治療を不必要に続けないようにすること。治療を中断するに値する重大で予期し得なかった副作用が有意な頻度でみられるかどうかをチェックすること。臨床試験がプロトコルに則って適性に行われているかどうかについて評価すること。現在の日本臨床腫瘍研究グループ(JCOG)の効果安全性評価委員会では次の4つの役割を担っている。(1) 中間解析結果の審査、(2) 有害事象の審査、(3) モニタリングレポート審査、(4) 上記(1)~(3)を受けて、プロトコルの中止・修正を決定すること。独立データモニタリング委員会、と呼ばれることもある。

〈牧本 敦〉