

研究者主導臨床試験でのデータ入力におけるエラー割合の検討

○諸星晴香¹⁾ 井出和希¹⁾ 原田翔平¹⁾ 藤井貞子²⁾
中原あつ子²⁾ 鈴木剛士²⁾ 田熊規方²⁾ 山田 浩¹⁾

¹⁾静岡県立大学薬学部医薬品情報解析学分野 ²⁾白十字ホーム

【背景】

治験と比較し、研究者主導の臨床研究における信頼性は決して高いとは言えず、データマネジメントの質を確保することが、昨今、重要視されている。

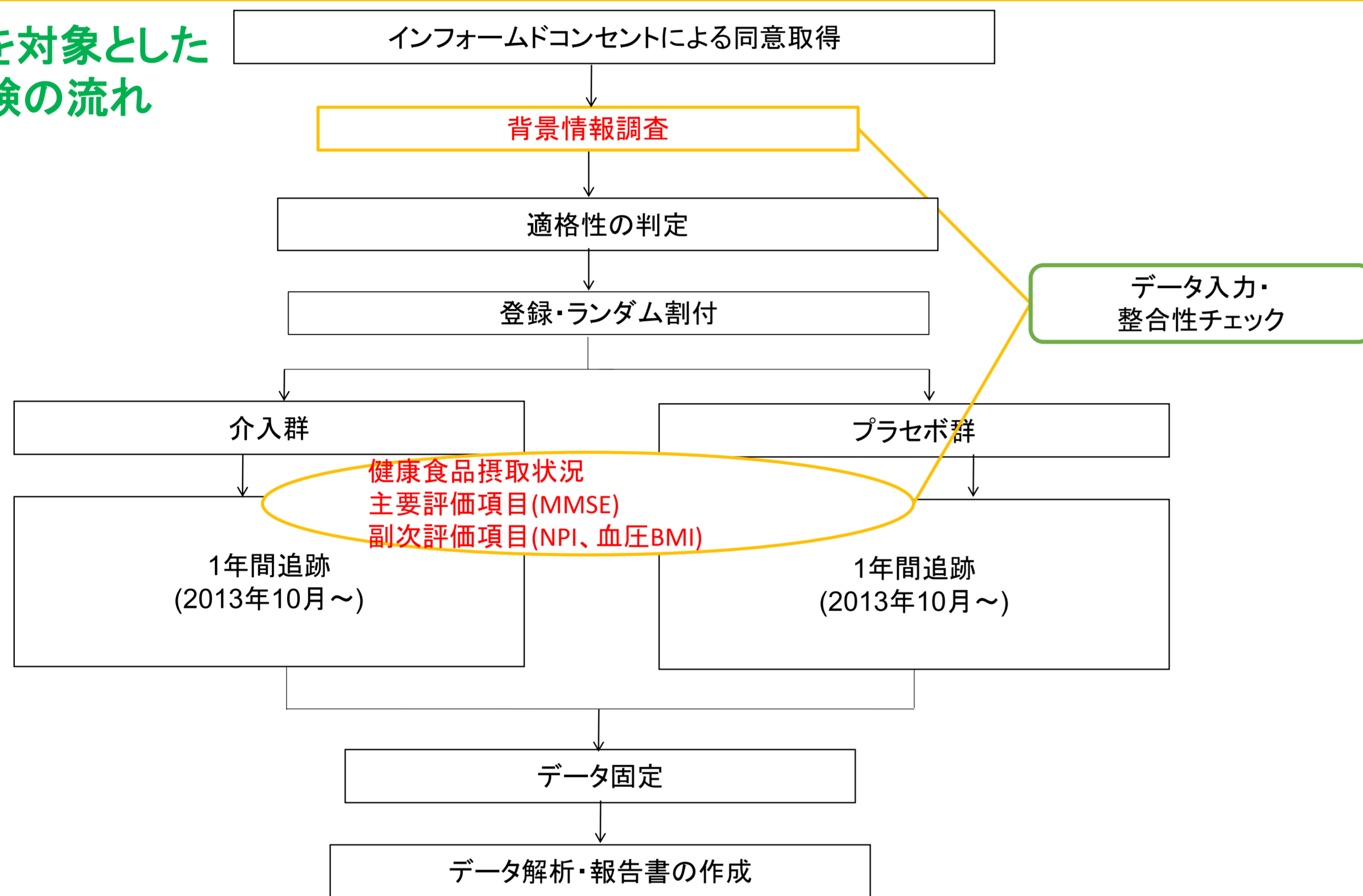
当研究室では大学院生・学部生の研究テーマとして臨床研究が取り入れられていることから、原資料からCase Report Form (CRF)、統計解析のためのデータシートへのデータ入力が欠かせない。入力データの整合性チェックは、データの信頼性確保に極めて重要であるが、研究者主導の臨床研究の場合、そのエラーの発生状況は必ずしも明らかではない¹⁾。

【目的】

自ら実施した研究者主導の臨床試験の経験から、その入力データにおけるエラー割合を調査し、データマネジメントの質の向上について検討する。

【方法】

認知症を対象とした臨床試験の流れ



データシートへの入力方法

- ・2013年10月から開始した認知症機能低下に対する食品(茶飲量)の有効性に関するランダム化比較試験を対象として、調査を行なった。
- ・背景情報、健康食品摂取状況、主要評価項目(Mini-Mental State Examination:MMSE)、副次評価項目(Neuropsychiatric Inventory:NPI)、血圧、Body Math Index(BMI)について、2名の入力者が独立してデータシートに入力した。
- ・その後、ExcelにおけるIF関数により整合性をダブルチェックし、それぞれのエラー数、エラー割合(エラー数/該当データの総数)を算出した。
- ・エラー数は、入力者の単純なミスによるエラーと、原資料の不鮮明によるエラーの2つに分類した。なお、原資料の欠損は原資料不鮮明に含め、ローカルデータマネジメントで解消した。
- ・次いで、一致しなかったデータを原資料(診療録等)、CRFに戻り不一致を解消し、併せて原因を精査した。

背景情報の入力例

No	4	5
生年月	1941年5月	1923年10月
性別	女	女
基礎疾患 ¹⁾	VD	VD
合併症 ²⁾	HT 脳出血 虚血性心疾患	HT DL 多発性脳梗塞
服用薬名	バイアスピリン	バイアスピリン アルダクトン リビートル アーテスト
日常生活動作	一部介助	自立
介護度 ³⁾	2	3
頭部画像	CT	CT
脳萎縮 ⁴⁾	0	1
喫煙習慣	0	0

- 1)基礎疾患は、AD:アルツハイマー病
VD:脳血管性認知症
DLB:レビー小体型認知症の略語を使用
- 2)合併症は、HT:高血圧
DL:脂質異常症
DM:糖尿病の略語を使用
- 3)介護度は、要介護(1・2・3・4・5)の5段階で評価
- 4)脳萎縮は、医師の判断による。0:萎縮無し
1:萎縮有り

摂取状況の入力例

ID 5	1日	2日	3日	4日	5日
朝	1	1	1	1	1
昼	1	1	1	1	1
夕	1	1	1	1	1
介入飲料以外の茶飲用量(mL)	600	600	600	600	600

記入方法は、1:全量飲んだ、1/2:半量飲んだ、0:飲まなかったの3通り。

MMSEの入力例

登録番号	評価項目(満点)	時に関する見当識(5)	場所に関する見当識(5)	記銘(3)	注意と計算(5)	再生(3)	呼称(2)	復唱(1)	理解(3)	読字(1)	書字(1)	描画(1)	合計点(30)
3		4	4	3	5	0	2	1	3	1	1	1	25
4		2	0	0	0	0	2	0	0	0	0	0	4
5		1	5	3	4	0	2	1	3	1	1	1	22

NPIの入力例

ID	妄想	幻覚	興奮	うつ	不安	多幸	無関心	脱抑制	易怒性	異常行動	夜間行動	食行動	合計
7	0	0	0	0	0	0	0	0	1	0	0	0	1
8	0	0	0	0	0	0	0	0	0	1	0	0	1

※各項目重症度:0-3(4段階)負担度:0-5(6段階)で記入

血圧,BMIの入力例

ID	SBP	DBP	BMI
15	100	48	19.3
16	104	70	21.7
17	128	70	22.9
18	120	76	20.2
19	132	80	23.8
20	122	58	17.7
21	154	64	21.9

SBP:systolic blood pressure(収縮期血圧)
DBP:diastolic blood pressure(拡張期血圧)

【結果】

データ収集期間:2013年10月～2014年4月
対象被験者数:30人

	総計
データ数	24,336
エラー数	90
エラー割合(%)	0.370

	背景情報	摂取状況	MMSE	NPI	血圧,BMI
データ数	1,025	20,179	972	1,920	240
エラー数	26	54	4	3	3
エラー割合(%)	2.54	0.268	0.412	0.156	1.25

最も高い

背景情報における入力エラー割合

背景情報	計	(テキスト抜き)
データ数	1,025	784
エラー数	26	8
エラー割合(%)	2.54	1.02

*テキストデータは単語1つをデータ数1としてカウントした。

摂取状況における入力エラー割合

摂取状況	2013年10月	11月	12月	2014年1月	2月	3月	4月	計
データ数	1,302	3,177	3,348	3,295	2,912	3,179	2,966	20,179
エラー数	0	5	26	4	4	7	8	54
エラー割合(%)	0	0.157	0.777	0.121	0.137	0.220	0.270	0.268

MMSEにおける入力エラー割合

MMSE	開始前(2013年10月)	3ヶ月後(2014年1月)	6ヶ月後(2014年4月)	計
データ数	360	312	300	972
エラー数	0	4	0	4
エラー割合(%)	0	1.28	0	0.412

血圧,BMIにおける入力エラー割合

血圧,BMI	開始前(2013年10月)	3ヶ月後(2014年1月)	6ヶ月後(2014年4月)	計
データ数	84	81	75	240
エラー数	0	1	2	3
エラー割合(%)	0	1.23	2.67	1.25

【考察】

- ・調査結果より、全体のエラー割合は0.4%以下であった。この結果は、患者満足度調査の報告におけるデータ入力作業においてヒトが間違える確率(最大7.8%)²⁾と比較し、それほど高くはなかったと言える。
- ・項目別では、背景情報がエラー割合2.5%と高い。背景情報には、数字データのみならずテキストデータも含まれることから、入力の煩雑や入力者の判断の違いなどがエラーの一因であると考えられる。
- ・入力エラーが頻出したのは、摂取状況で2ヶ月後、MMSEで3ヶ月後、NPIで3ヶ月後であり、試験開始後3ヶ月までの、比較的早期のデータに入力エラーが多い傾向が示された。

【結論】

- ・データの精度が重要視される統計解析において、データ入力時の不一致は解析結果の信頼性に大きな影響を与えることから、エラーの発生を0に近づけること及び試験早期からのデータの整合性チェックの必要性が示された。

【参考文献】

1. 村山季実子、他:研究者主導臨床試験におけるローカルデータマネジメントの質とCRC有無・登録症例数の関連:薬理と治療 2013年 41巻Suppl 2号 p. 103-110:ライフサイエンス出版
2. 多田敏彦、他:ヒトが間違える確率を応用した組織改善:山形県立病院医学雑誌 (0389-0848)47巻1号 Page61-66(2013.01)

本演題発表に関連して、開示すべきCOI関係にある企業等はありません。