



統合解析用データセットの自動 構築に向けて

SAS Forumユーザー会 学術大会, 2004年 7月30日

角谷伸一, 北西由武, 長谷川貴大,

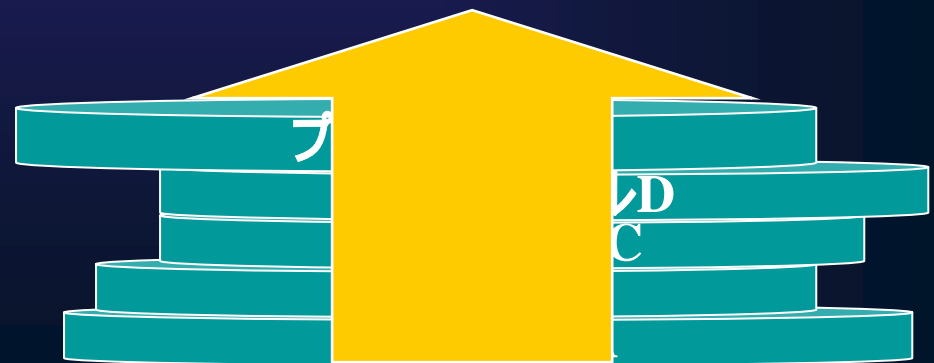
平野勝也, 田崎武信

塩野義製薬(株)



統合解析

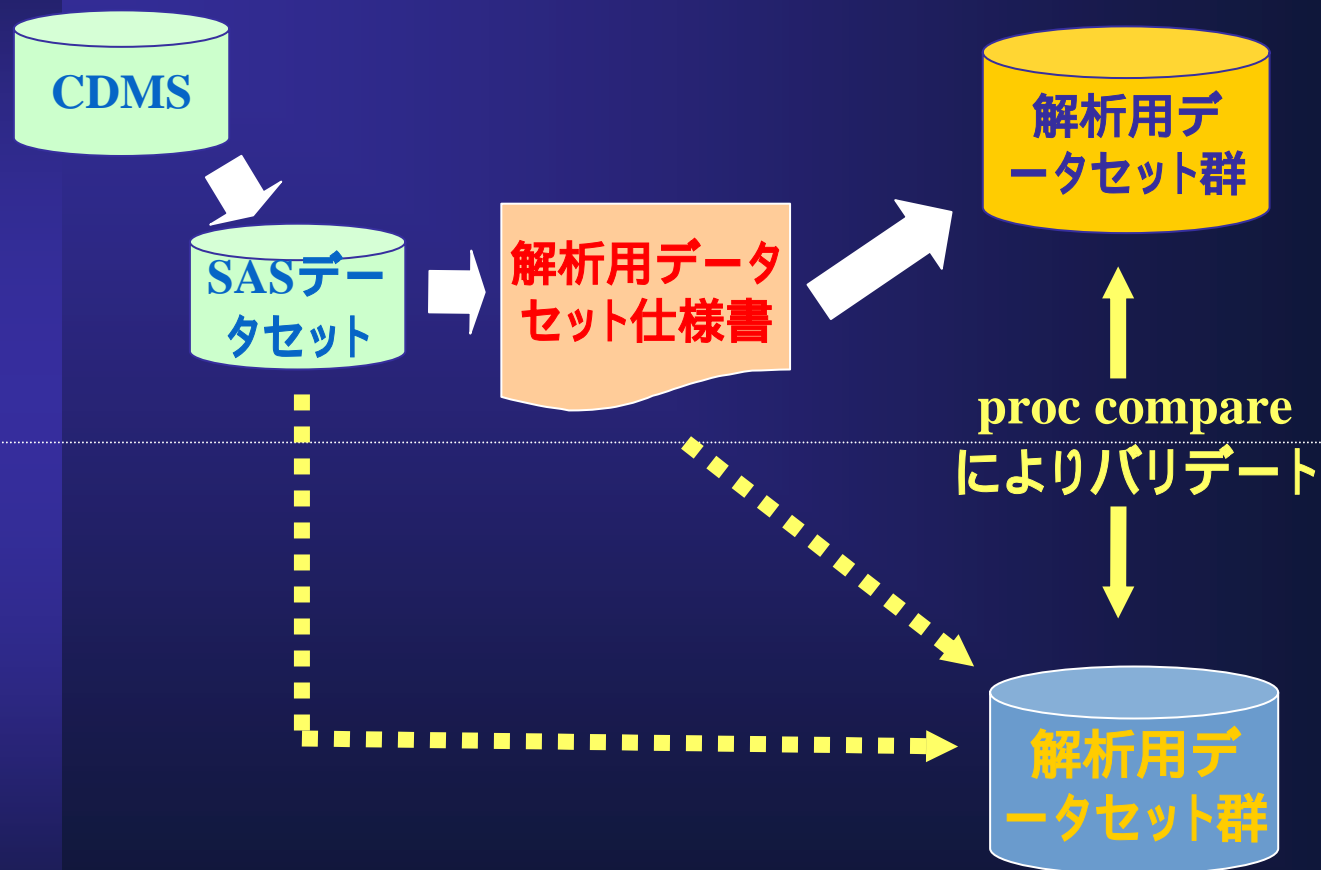
- ◆ 申請資料で個々の試験とは別に、全試験を通して、有効性と安全性の結果を要約することが要求されている。
- ◆ その資料を作成するために、統合解析用データセットの作成が不可欠。
- ◆ 個々の試験の解析用データセットを標準化することを出発点として考えた。



解析業務主担当者

バリデーション担当者

これまでの手順





現状の手順の問題点

- ◆ 個々の試験ごとに解析業務主担当者が存在する
プロトコル間(担当者間, 担当者内)のばらつきの発生
解析用データセットの名称, 変数の属性(名称, ラベル,
タイプ, レンクス)が**プロトコル間で異なることがある**.
- ◆ 統合解析時に, 個々の試験の解析用データセット
仕様書, 解析用データセットを見直すことが必要





問題への対処

- ◆ 個々の試験ごとに作成する解析用データセットの標準化を検討

そのために、

- 変数辞書(VAR_DIC)の導入
解析用データセットに必要な変数情報を一元管理
- 変数辞書を効率よく利用するための、解析用データセット仕様書.xlsの開発



変数辞書 (VAR_DIC)

- ◆ 解析用データセットに必要な変数情報を登録, 保存, 管理
- ◆ 新規変数の管理
- ◆ 項目名が同じSAS変数はひとつとして管理
重複がないように管理



変数辞書の内容

形式:SASデータセット
データセット名:
VAR_DIC

変数名	変数名に対するラベル	内容
memname	データセット名	解析用データセット名
cdms	CDMS	CDMS のどのデータセットから取得したものがわかるように, データセット名を入力
name	変数名	SAS 変数名を入力
type	変数タイプ	N: 数値変数, C: 文字変数
length	変数の長さ	変数の長さ
format	フォーマット情報	SAS 変数にフォーマットがある場合, フォーマット名を入力
label_J	変数のラベル(日本語)	SAS 変数の日本語のラベルを入力
label_E	変数のラベル(英語)	SAS 変数の英語のラベルを入力
product	品目	品目名を入力
belong	所属	CDMS 由来のものか, CDMS にはない新規変数なのかを識別.




変数辞書導入の利点

解析用データ
セットに必要な
変数情報を登
録



変数辞書



変数辞書に登録
されている情報
を参照



解析用デ
ータセット群

変数辞書を活用することで、
プロトコル間(担当者間, 担当者内)のばらつきがなくなる。
標準的な解析用データセットを作成できる



変数辞書を効率よく利用するためのExcelの開発
解析用データセット仕様書.xls



解析用データセット仕様書.xls

- ◆ 変数辞書を有効活用し、標準的な解析用データセットを作成するためのツール
 - ・解析用データセット仕様書作成機能
 - ・変数辞書に登録されていない新規変数を辞書へ登録させる機能
 - ・空データセット作成機能
 - ・辞書情報を一覧表示させる機能

裏で、SASプログラムを実行させる。

3つのシート(MAIN, REGESTER, DICTIONARY)で実施。



解析用データセット仕様書.xls

Microsoft Excel - 解析用データセット仕様書(検索機能追加)

ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) データ(D) ウィンドウ(W) ヘルプ(H)

MSゴシック 11 B I U

	A	B	C	D	E	F	G	H	I	J	K
1	開発品目:		KEY情報:								
2	プロトコル:		実行結果								
3	データ情報取得(辞書情報更新)										
4	適合性チェック										
5	空データセット作成										
6	データセット名	CDMS	変数名	タイプ	長さ	フォーマット	ラベル(日本語)	ラベル(英語)	品目	所属	
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											

CDMS 上のデータベースからの情報がここに出力される。
解析に必要な追加変数がここに入力される。

Microsoft Excel - 解析用データセット仕様書(検索機能追加)

ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) データ(D) ウィンドウ(W) ヘルプ(H)

MSゴシック 11 B I U

	A	B	C	D	E	F	G	H	I	J
1	辞書登録									
2	実行結果									
3	データセット名									
4	データセット名	CDMS	変数名	タイプ	長さ	フォーマット	ラベル(日本語)	ラベル(英語)	品目	所属
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										
29										

辞書に登録されていない新規変数の SAS 変数の一覧がここに
表示される。

Microsoft Excel - 解析用データセット仕様書(検索機能追加)

ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) データ(D) ウィンドウ(W) ヘルプ(H)

MSゴシック 11 B I U

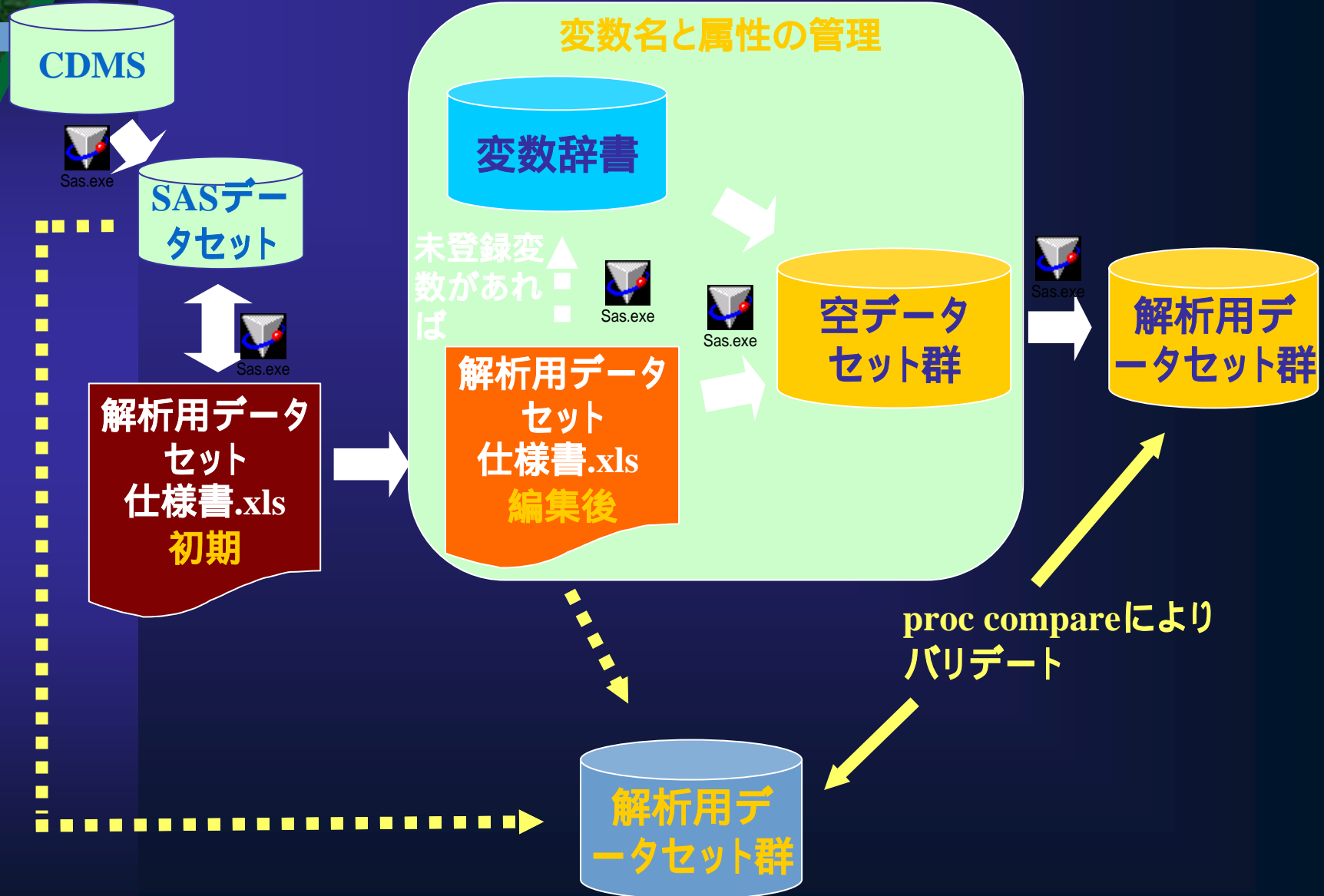
	A	B	C	D	E	F	G	H	I	J
1	CDMS	変数名	タイプ	長さ	フォーマット	ラベル(日本語)	ラベル(英語)	品目	所属	
2										
3										
4										
5										
6										
7										
8										
9	CDMS	変数名	タイプ	長さ	フォーマット	ラベル(日本語)	ラベル(英語)	品目	所属	
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										
29										

オートフィルタの表示 キーワード検索の実行 クリアボタン

新しい手順

解析業務主担当者

バリデーション担当者





解析用データセット仕様書.xls の機能

STEP1: 解析用データセット仕様書の作成

Microsoft Excel - 解析用データセット仕様書.xls

ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) データ(D) ウィンドウ(W) ヘルプ(H)

MS ゴシック 11 B I U

D9 =

	A	B	C	D	E	F	G	H	I
1	開発品目:	PQ_TEST	KEY情報:						
2	プロトコル:	PRG_TEST		実行結果					
3				メタ情報取得(辞書情報更新)					
4	整合性チェック		空データセット作成						
5	データセット名	CDMS	変数名	タイプ	長さ	フォーマット	ラベル(日本語)	品目	所属
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									

MAIN / REGISTER / DICTIONARY /

STEP1: 解析用データセット仕様書の作成

Microsoft Excel - 解析用データセット仕様書.xls

ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) データ(D) ウィンドウ(W) ヘルプ(H)

MS ゴシック 11 B I U

C43 =

	A	B	C	D	E	F	G	H	I
1	開発品目:	PQ_TEST	KEY情報:						
2	プロトコル:	PRG_TEST	実行結果						
3			×タ情報取得(辞書情報更新)				正常終了		
4		整合性チェック	空データセット作成						
5	データセット名	CDMS	変数名	タイプ	長さ	フォーマット	ラベル(日本語)	品目	所属
6	DEMO	DEMOS	SUBJID	C	6		症例番号	COMMON	CDMS
7	DEMO	DEMOS	DEPDSM	C	5	\$DSMFOR	DSM-IVによる病型分類	PQ_TEST	CDMS
8	DEMO	DEMOS	DEPDSDYN	C	5	\$YN	気分的一致・不一致	PQ_TEST	CDMS
9	DEMO	DEMOS	DEPDUR	C	5	\$DEPDUR	今回発病時期	PQ_TEST	CDMS
10	DEMO	DEMOS	DEPFIAGE	N	8		初発年齢	PQ_TEST	CDMS
11	DEMO	DEMOS	DEPTIMES	C	5	\$DEPTIM	病相回数	PQ_TEST	CDMS
12	DEMO	DEMO_A	SUBJID	C	6		症例番号	PQ_TEST	CDMS
13	DEMO	DEMO_A	AGE	N	8		年齢	PQ_TEST	CDMS
14	DEMO	DEMO_A	BIRTHDT	N	8	YYMMDD	生年月日	PQ_TEST	CDMS
15	DEMO	DEMO_A	IO	C	5	\$INOUT	入院・外来	PQ_TEST	CDMS
16	DEMO	DEMO_A	NURSEYN	C	5	\$UM	授乳の有無	PQ_TEST	CDMS
17	DEMO	DEMO_A	PRGTESYN	C	5	\$UM	妊娠検査の有無	PQ_TEST	CDMS
18	DEMO	DEMO_A	PRGYN	C	5	\$UM	妊娠の有無	PQ_TEST	CDMS
19	DEMO	DEMO_A	SEX	C	5	\$SEX	性別	PQ_TEST	CDMS
20	DEPHAMD	DEPHAMD	SUBJID	C	6		症例番号	PQ_TEST	CDMS
21	DEPHAMD	DEPHAMD	DEP01HAM	N	8		抑うつ気分	PQ_TEST	CDMS
22	DEPHAMD	DEPHAMD	DEP02HAM	N	8		仕事と活動	PQ_TEST	CDMS
23	DEPHAMD	DEPHAMD	DEP03HAM	N	8		生殖器症状	PQ_TEST	CDMS
24	DEPHAMD	DEPHAMD	DEP04HAM	N	8		身体症状、消化器系	PQ_TEST	CDMS

ここで、仕様書を作成する

MAIN REGISTER DICTIONARY

STEP2: 新規変数の辞書への登録

Microsoft Excel - 解析用データセット仕様書.xls

ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) データ(D) ウィンドウ(W) ヘルプ(H)

MS ゴシック 11 B I U

A40 = DEPHAMD

	A	B	C	D	E	F	G	H	I
1	開発品目:	PQ_TEST	KEY情報:						
2	プロトコル:	PRG_TEST	実行結果						
3			メタ情報取得(辞書情報更新)				正常終了		
4	整合性チェック		空データセット作成				辞書に含まれていない変数があります		
5	データセット名	CDMS	変数名	タイプ	長さ	フォーマット	ラベル(日本語)	品目	所属
6	DEMO	DEMOS	SUBJID	C	6		症例番号	COMMON	CDMS
7	DEMO	DEMOS	DEPDSM	C	5	\$DSMFOR	DSM-IVによる病型分類	PQ_TEST	CDMS
8	DEMO	DEMOS	DEPDSDMYN	C	5	\$YN	気分一致・不一致	PQ_TEST	CDMS
9	DEMO	DEMOS	DEPDUR	C	5	\$DEPDUR	今回発病時期	PQ_TEST	CDMS
10	DEMO	DEMOS	DEPIAGE	N	8		初発年齢	PQ_TEST	CDMS
11	DEMO	DEMOS	DEPTIMES	C	5	\$DEPTIM	病相回数	PQ_TEST	CDMS
12	DEMO	DEMO_A	SUBJID	C	6		症例番号	PQ_TEST	CDMS
13	DEMO	DEMO_A	AGE	N	8		年齢	PQ_TEST	CDMS
14	DEMO	DEMO_A	BIRTHDT	N	8	YYMMDD	生年月日	PQ_TEST	CDMS
15	DEMO	DEMO_A	IO	C	5	\$INOUT	入院・外来	PQ_TEST	CDMS
16	DEMO	DEMO_A	NURSEYN	C	5	\$UM	授乳の有無	PQ_TEST	CDMS
17	DEMO	DEMO_A	PRGTESYN	C	5	\$UM	妊娠検査の有無	PQ_TEST	CDMS
18	DEMO	DEMO_A	PRGYN	C	5	\$UM	妊娠の有無	PQ_TEST	CDMS
19	DEMO	DEMO_A	SEX	C	5	\$SEX	性別	PQ_TEST	CDMS
20	DEPHAMD	DEPHAMD	SUBJID	C	6		症例番号	PQ_TEST	CDMS
21	DEPHAMD	DEPHAMD	DEP01HAM	N	8		抑うつ気分	PQ_TEST	CDMS
22	DEPHAMD	DEPHAMD	DEP02HAM	N	8		仕事と活動	PQ_TEST	CDMS
23	DEPHAMD	DEPHAMD	DEP03HAM	N	8		生殖器症状	PQ_TEST	CDMS
24	DEPHAMD	DEPHAMD	DEP04HAM	N	8		身体症状、消化器系	PQ_TEST	CDMS

変数辞書に登録されていない、新規変数が赤字で表示される。

MAIN / REGISTER / DICTIONARY

STEP2: 新規変数の辞書への登録

Microsoft Excel - 解析用データセット仕様書.xls

ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) データ(D) ウィンドウ(W) ヘルプ(H)

MS ゴシック 11 B I U

A4 = DEMO

	A	B	C	D	E	F	G	H	I
1							実行結果		
2									
3	データセット名	CDMS	変数名	タイプ	長さ	フォーマット	ラベル(日本語)	品目	所属
4	DEMO	DEMOS	DEPDMS	C	5	\$DSMFOR	DSM-IVによる病型分類	PQ_TEST	CDMS
5	DEMO	DEMOS	DEPDMSYN	C	5	\$YN	気分的一致・不一致	PQ_TEST	CDMS
6	DEMO	DEMOS	DEPDUR	C	5	\$DEPDUR	今回発病時期	PQ_TEST	CDMS
7	DEMO	DEMOS	DEPFIAGE	N	8		初発年齢	PQ_TEST	CDMS
8	DEMO	DEMOS	DEPTIMES	C	5	\$DEPTIM	病相回数	PQ_TEST	CDMS
9	DEMO	DEMO_A	AGE	N	8		年齢	PQ_TEST	CDMS
10	DEMO	DEMO_A	BIRTHDT	N	8	YYMMDD	生年月日	PQ_TEST	CDMS
11	DEMO	DEMO_A	IO	C	5	\$INOUT	入院・外来	PQ_TEST	CDMS
12	DEMO	DEMO_A	NURSEYN	C	5	\$UM	授乳の有無	PQ_TEST	CDMS
13	DEMO	DEMO_A	PRGTESYN	C	5	\$UM	妊娠検査の有無	PQ_TEST	CDMS
14	DEMO	DEMO_A	PRGYN	C	5	\$UM	妊娠の有無	PQ_TEST	CDMS
15	DEMO	DEMO_A	SEX	C	5	\$SEX	性別	PQ_TEST	CDMS
16	DEMO	DEMO_A	SUBJID	C	6		症例番号	PQ_TEST	CDMS
17	DEPHAMD	DEPHAMD	DEP01HAM	N	8		抑うつ気分	PQ_TEST	CDMS
18	DEPHAMD	DEPHAMD	DEP02HAM	N	8		仕事と活動	PQ_TEST	CDMS
19	DEPHAMD	DEPHAMD	DEP03HAM	N	8		生殖器症状	PQ_TEST	CDMS
20	DEPHAMD	DEPHAMD	DEP04HAM	N	8		身体症状、消化器系	PQ_TEST	CDMS
21	DEPHAMD	DEPHAMD	DEP05HAM	N	8		体重減少	PQ_TEST	CDMS
22	DEPHAMD	DEPHAMD	DEP06HAM	N	8		入眠障害	PQ_TEST	CDMS
23	DEPHAMD	DEPHAMD	DEP07HAM	N	8		熟眠障害	PQ_TEST	CDMS
24	DEPHAMD	DEPHAMD	DEP08HAM	N	8		早朝睡眠障害	PQ_TEST	CDMS
25	DEPHAMD	DEPHAMD	DEP09HAM	N	8		身体症状、一般的	PQ_TEST	CDMS
26	DEPHAMD	DEPHAMD	DEP10HAM	N	8		罪業感	PQ_TEST	CDMS
27	DEPHAMD	DEPHAMD	DEP11HAM	N	8		自殺	PQ_TEST	CDMS

辞書登録

新規変数の
名称など問題
がないことを
確認する

MAIN REGISTER DICTIONARY

STEP2: 新規変数の辞書への登録

Microsoft Excel - 解析用データセット仕様書.xls

ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) データ(D) ウィンドウ(W) ヘルプ(H)

MS ゴシック 11 B I U

G18 =

	A	B	C	D	E	F	G	H	I
1							実行結果		
2							正常終了		
3	データセット名	CDMS	変数名	タイプ	長さ	フォーマット	ラベル(日本語)	品目	所属
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
25									
26									
27									

新規変数が辞書へ登録され、この画面から情報がなくなる

MAIN REGISTER DICTIONARY

STEP2: 新規変数の辞書への登録

Microsoft Excel - 解析用データセット仕様書.xls

ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) データ(D) ウィンドウ(W) ヘルプ(H)

MS ゴシック 11 B I U % , +.00 +.00 75%

A5 = DEMO

	A	B	C	D	E	F	G	H	I
1	データセット名	CDMS	変数名	タイプ	長さ	フォーマット	ラベル(日本語)	品目	所属
2									
3	オートフィルタの表示		キーワード検索の実行		クリアボタン				
4	データセット名	CDMS	変数名	タイプ	長さ	フォーマット	ラベル(日本語)	品目	所属
5	DEMO	DEMOS	DEPDSM	C	5	\$DSMFOR	DSM-IVによる病型分類	PQ_TEST	CDMS
6	DEMO	DEMOS	DEPDSMYN	C	5	\$YN	気分一致・不一致	PQ_TEST	CDMS
7	DEMO	DEMOS	DEPDUR	C	5	\$DEPDUR	今回発病時期	PQ_TEST	CDMS
8	DEMO	DEMOS	DEPFIAGE	N	8		初発年齢	PQ_TEST	CDMS
9	DEMO	DEMOS	DEPTIMES	C	5	\$DEPTIM	病相回数	PQ_TEST	CDMS
10	DEMO	DEMOS	SUBJID	C	6		症例番号	COMMON	CDMS
11	DEMO	DEMO_A	AGE	N	8		年齢	PQ_TEST	CDMS
12	DEMO	DEMO_A	BIRTHDT	N	8	YYMMDD	生年月日	PQ_TEST	CDMS
13	DEMO	DEMO_A	IO	C	5	\$INOUT	入院・外来	PQ_TEST	CDMS
14	DEMO	DEMO_A	NURSEYN	C	5	\$UM	授乳の有無	PQ_TEST	CDMS
15	DEMO	DEMO_A	PRGTESYN	C	5	\$UM	妊娠検査の有無	PQ_TEST	CDMS
16	DEMO	DEMO_A	PRGYN	C	5	\$UM	妊娠の有無	PQ_TEST	CDMS
17	DEMO	DEMO_A	SEX	C	5	\$SEX	性別	PQ_TEST	CDMS
18	DEMO	DEMO_A	SUBJID	C	6		症例番号	PQ_TEST	CDMS
19	DEPHAMD	DEPHAMD	DEP01HAM	N	8		抑うつ気分	PQ_TEST	CDMS
20	DEPHAMD	DEPHAMD	DEP02HAM	N	8		仕事と活動	PQ_TEST	CDMS
21	DEPHAMD	DEPHAMD	DEP03HAM	N	8		生殖器症状	PQ_TEST	CDMS
22	DEPHAMD	DEPHAMD	DEP04HAM	N	8		身体症状、消化器系	PQ_TEST	CDMS
23	DEPHAMD	DEPHAMD	DEP05HAM	N	8		体重減少	PQ_TEST	CDMS
24	DEPHAMD	DEPHAMD	DEP06HAM	N	8		入眠障害	PQ_TEST	CDMS
25	DEPHAMD	DEPHAMD	DEP07HAM	N	8		熟眠障害	PQ_TEST	CDMS
26	DEPHAMD	DEPHAMD	DEP08HAM	N	8		早朝睡眠障害	PQ_TEST	CDMS
27	DEPHAMD	DEPHAMD	DEP09HAM	N	8		身体症状	PQ_TEST	CDMS

MAIN REGISTER **DICTIONARY**

STEP3: 空データセットの作成

Microsoft Excel - 解析用データセット仕様書.xls [読み取り専用]

ファイル(F) 編集(E) 表示(V) 挿入(I) 書式(O) ツール(T) データ(D) ウィンドウ(W) ヘルプ(H)

MS ゴシック 11 B I U

M44 =

	A	B	C	D	E	F	G	H	I
1	開発品目:	PQ_TEST	KEY情報:		SUBJID, PROTOCOL				
2	プロトコル:	PRG_TEST					実行結果		
3			メタ情報取得(辞書情報更新)				正常終了		
4	整合性チェック		空データセット作成				正常終了		
5									
6	DEMO	DENOS	DEPDEN	C	5	\$DSMFOR	DSM-IVによる病型分類	PQ_TEST	COMS
7	DEMO	DENOS	DEPDENYN	C	5	\$YN	気分一致・不一致	PQ_TEST	COMS
8	DEMO	DENOS	DEPDUR	C	5	\$DEPDUR	今回発症時期	PQ_TEST	COMS
9	DEMO	DENOS	DEPFIAGE	N	8		初発年齢	PQ_TEST	COMS
10	DEMO	DENOS	DEPTIMES	C	5	\$DEPTIM	発症回数	PQ_TEST	COMS
11	DEMO	DEMO_A	SUBJID	C	6		症例番号	COMMON	COMS
12	DEMO	DEMO_A	AGE	N	8		年齢	PQ_TEST	COMS
13	DEMO	DEMO_A	BIRTHDT	N	8	YYMMDD	生年月日	PQ_TEST	COMS
14	DEMO	DEMO_A	IO	C	5	\$IINOUT	入院・外来	PQ_TEST	COMS
15	DEMO	DEMO_A	NURSEYN	C	5	\$UM	授乳の有無	PQ_TEST	COMS
16	DEMO	DEMO_A	PRGTESTYN	C	5	\$UM	妊娠検査の有無	PQ_TEST	COMS
17	DEMO	DEMO_A	PRGYN	C	5	\$UM	妊娠の有無	PQ_TEST	COMS
18	DEMO	DEMO_A	SEX	C	5	\$SEX	性別	PQ_TEST	COMS
19	DEMO	DEMO_B	MHYN	C	5	\$UM	既往症・合併症の有無	PQ_TEST	COMS
20	DEMO	DEMO_C	POYN	C	5	\$UM	前治療の有無	PQ_TEST	COMS
21	DEMO	DEMO_D	PTYN	C	5	\$UM	前治療法の有無	PQ_TEST	COMS
22	DEMO	DEMO_E	CTYN	C	5	\$UM	併用療法有無	PQ_TEST	COMS
23	DEMO	DEMO_F	COYN	C	5	\$UM	併用薬有無	PQ_TEST	COMS
24	DEMO	DEMO_G	AEYN	C	5	\$UM	有害事象有無	PQ_TEST	COMS

MAIN / REGISTER / DICTIONARY



STEP3: 空データセットの作成

SAS

ファイル(F) 表示(V) ツール(T) ソリューション(S) ウィンドウ(W) ヘルプ(H)

VIEWTABLE: Library.Demo

	有害事象有無	年齢	生年月日	併用薬有無	併用療法有無	DSM-IVによる病型分類
1						

列の属性

一般 | 色 | フォント | 閉じる | 適用 | ヘルプ

名前: AGE

ラベル: 年齢

長さ: 8

出力形式: 8. ...

入力形式: 8. ...

タイプ

☐ Character

☒ Numeric

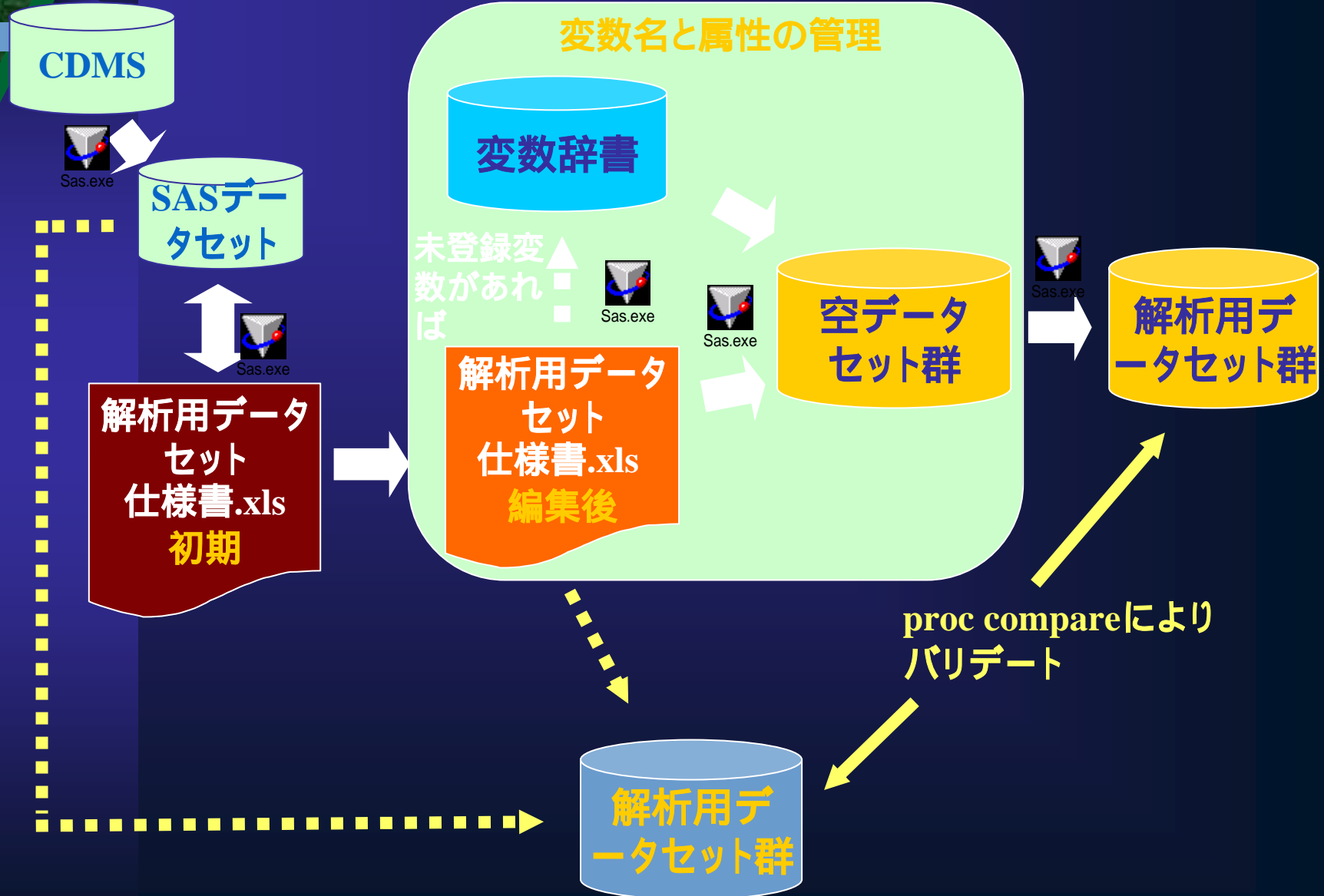
結果 エクス...

C:\Program Files\SAS Ir

新しい手順

解析業務主担当者

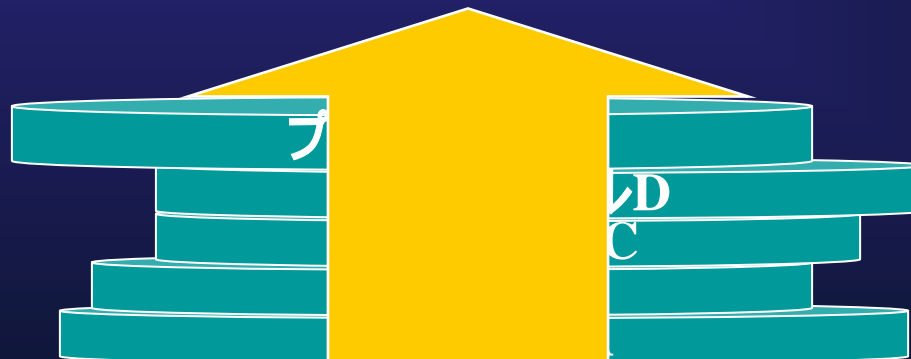
バリデーション担当者





統合解析用データセットの作成

- ◆ 個々の試験で標準化された解析用データセットが完成している。
- ◆ SASのSET文で縦に結合することで統合解析用データセットを作成できる。



SASのSET文

統合解析用データセットの完成



まとめ

- ◆ 統合解析用データセットを作成するために、まず、個々の解析用データセットを標準化することが不可欠である。
- ◆ 変数情報を管理するための変数辞書を導入し、この辞書を有効に利用できるツール、解析用データセット仕様書.xlsを利用することで、標準化された解析用データセットを完成させることが可能となった。
- ◆ 個々の試験の標準化された解析用データセットをSASのSET文で縦に連結すれば、統合解析用データセットが効率よく完成させることが可能となった。

