

Minimising 'Data Loss' During Mapping of Raw Data to SDTM

John Monaghan, Manager of Statistical Programming, IQVIA



Agenda

- + Brief introduction to SDTM mapping process (for those not familiar with this)
- + General pain point for SDTM mapping – Non Standard Data Adds Complexity
- + Specific pain point for SDTM mapping – Data Complexity Increases Risk of Data Loss
- + Solution for pain point – A tool for quickly/easily identifying “at risk” data
- + Tool validation and testing
- + Discussion of results
- + Future directions

What is Standard Data Tabulation Model (SDTM)

- The main output from a clinical trial is **data**.
- **Data** must be analysed and submitted to regulatory authorities (e.g., FDA) to get approval for a therapeutic.
- Raw **data** is collected from a variety of sources
 - Forms filled out by doctors describing effect of the therapeutic on a patient (Case report form (CRF)).
 - Forms filled out by patients (Patient reported outcomes (ePRO)).
 - Other sources (laboratory tests, wearable devices etc.).
- Raw **data** must be “standardized/converted/mapped” to produce SDTM data.
- SDTM **data** is then analysed to produce ADaM **data**, tables, listings and figures.
- SDTM **data**, ADaM **data**, tables, figures and listings are submitted to regulatory authorities. Raw data is not submitted.





**Pain point with SDTM mapping
process – Non Standard Data Adds
Complexity!**

General Pain Point – Non Standard Data Adds Complexity!

Some sources of non-standard data:

- Free text fields

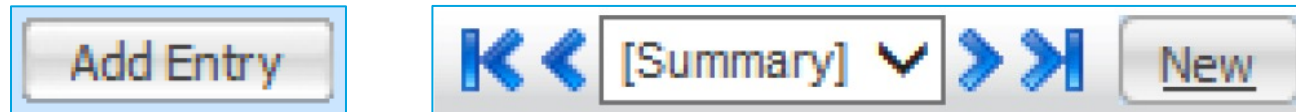


[ICF Version]

☐ [EOS Other unit]
Other, specify:

[Accession Number]
Accession Number:

- Dynamic page generation



Add Entry [Summary] ▼ New

- Complex vendor data (e.g., eCOA, central labs, biomarkers, wearables)
- Trial design (multi-phase, complex efficacy, dose escalation etc.)

Risks to Data Integrity Arising from SDTM Mapping Process

- Misinterpretation of SDTM rules ([Microsoft Word - PharmaSUG2020 Paper#329 OvercomingPitfallsOfDS.docx \(lexjansen.com\)](#))
- Incorrect implementation of SDTM rules ([SDTM Validation Tool \(lexjansen.com\)](#))
- Raw data that does not easily "fit" SDTM (<https://www.lexjansen.com/phuse/2010/cd/CD02.pdf>)
- Failure to map “all collected” data.



**Pain point with SDTM mapping
process – Data Complexity
Increases Risk of Data Loss!**

Pain Point – Data Complexity Increases Risk of Data Loss!

- Risk of data loss is partially addressed by aCRF process ([FDA - Study Data Technical Conformance Guide](#)):

4.1.4.6 Annotated Case Report Form (aCRF) for SDTM

An Annotated Case Report Form (aCRF) is a PDF document that maps the clinical data collection fields used to capture subject data (electronic or paper) to the corresponding variables or discrete variable values contained within the SDTM datasets. Regardless of whether the clinical database is in a format supported by the Catalog, an aCRF should be submitted preferably at the time a protocol is submitted. The aCRF should be provided as a PDF with the file name “acrf.pdf.”³²

The aCRF should include treatment assignment forms, when applicable, and should map each variable on the CRF to the corresponding variables in the datasets (or database). The aCRF should include the variable names and coding for each CRF item.

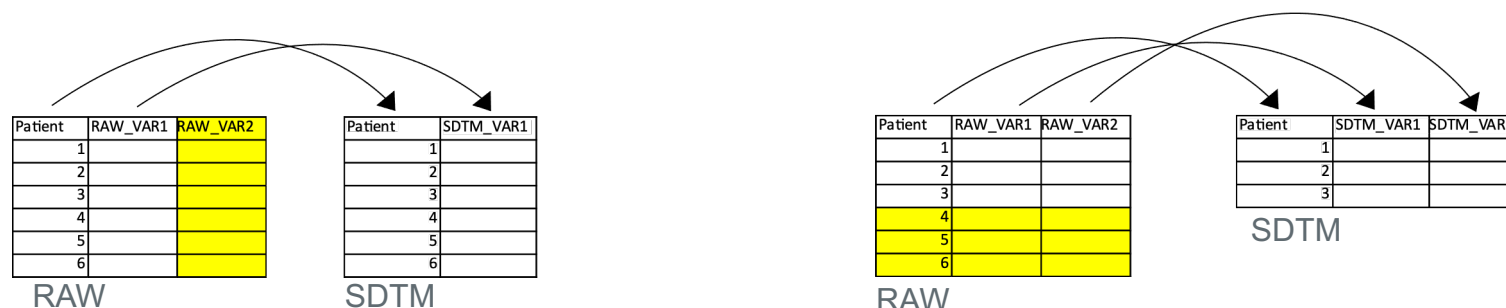
When data are recorded on the CRF but are not submitted, the CRF should be annotated with the text "NOT SUBMITTED." There should be an explanation in the SDRG stating why data have not been submitted.

Visit date	Fixed Unit: yyyy mmm dd
	SVSTDTC
Current follow up status of the subject	On site follow up ☐
	NOT SUBMITTED Telephone follow up ☐
	Survival follow up ☐
Will RECIST Assessment/Scan Continue	No ☐
	NOT SUBMITTED Yes ☐

- At IQVIA we have custom Data Lineage and Export Comparison (DLEC) tool for automatic identification of changes in datasets (new variables, datasets, changes to variable data types etc.) which complements the aCRF tool.

Pain Point – Data Complexity Increases Risk of Data Loss!

The aCRF process mitigates risk of **MISSING COLUMNS** of raw data but what about **MISSING ROWS/VALUES**!



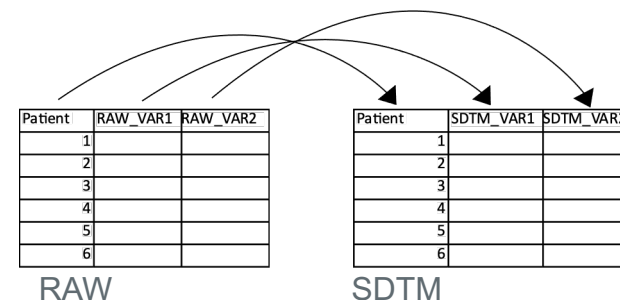
Question: How could rows/values be lost?

Answers:

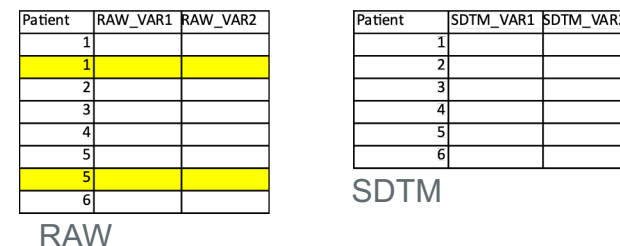
- SDTM mapping is not always simple!
- Dataset complexity grows with time!
- Data complexities can be rare – associated with a small number of patients/visits.
- Rare data complexities can be easily overlooked.

Extra data rows for some patients

DATA CUT 1

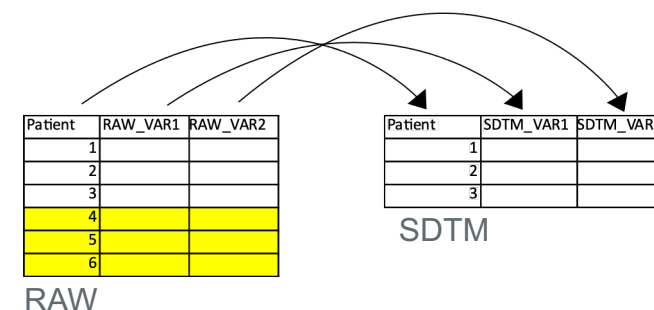
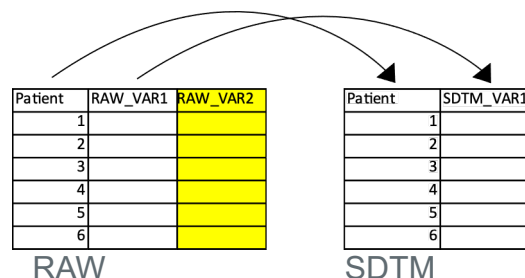


DATA CUT 2



Pain Point – Data Complexity Increases Risk of Data Loss!

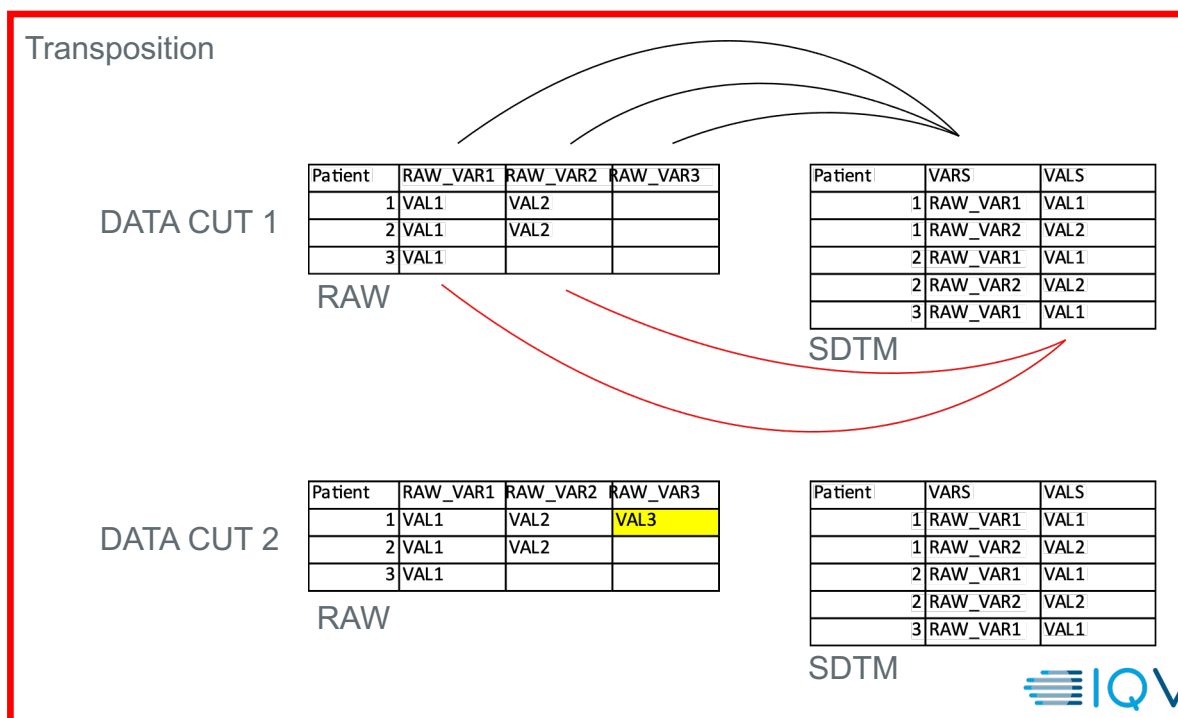
The aCRF process mitigates risk of **MISSING COLUMNS** of raw data but what about **MISSING ROWS/VALUES**!



Question: How could rows/values be lost?

Answers:

- SDTM mapping is not always simple!
- Dataset complexity grows with time!
- Data complexities can be rare – associated with a small number of patients/visits.
- Rare data complexities can be easily overlooked.



Pain Point – Data Complexity Increases Risk of Data Loss!

Questions:

- **How common is loss of data rows/values during SDTM mapping?**
 - Not much information in literature.
 - Anecdotally – does happen, is picked up during the QC process and leads to rework.
- **How is this risk mitigated?**
 - Pinnacle 21? – No
 - Risk based monitoring software (e.g. Cluepoints)? – No
 - Carefull planning and creation of mapping programs using dual programming with comprehensive review by senior employees – Yes!
- **What do programmers/reviewers need to prevent loss of rows/values?**
 - Expertise (Experience to know what type of raw data is at risk)
 - Time (Required to investigate raw data and create programmatic solutions)



**“We should all invest time in improving
and reassessing our quality processes”**

– Leo Valmas

Solution for Reducing Risks Associated with Complex Data

Data Investigator Macro

- **Reduces time and expertise** required to identify “complexities” in data
- Allows user to quickly count rows and list values associated with patients, visits, tests, dates etc.

```
/*Data Investigator*/
```

```
%let    RAW_DATASET  
%let    VARIABLES  
%let    VALUES  
%let    INFORMATION_FORMAT  
%let    ORDERED_BY_COUNT  
%let    DISTINCT
```

```
= dataset;  
= var1 var2 var3;  
= "val1" "val2";  
= "count";  
= "Y";  
= "N";
```

```
/*Macro*/
```

```
%include '\\quintiles.net\Enterprise\Sites\IEDUB\Users\q1064628\Useful stuff\Phuse macro\Phuse_include_file.sas';
```

Your raw dataset

A list of variables for subcategorising your quantification/description of the data

A list of values you can

Choose to count (quantify)
— or all values in data

Choose to quantify/describe unique
values or all values of data

The macro packaged in
an include statement

Validation of the Data Investigator Macro

Can the tool be used to help a reviewer find rare complexities in the data that might be missed during the SDTM mapping process?

- 3 studies at development stage:
 - First draft SDTM data
 - Fully dual programmed

Validation of the Data Investigator Macro

Can the tool be used to help a reviewer find rare complexities in the data that might be missed during the SDTM mapping process?

- 3 studies at development stage:
 - First draft SDTM data
 - Fully dual programmed
- Concentrated on efficacy and ePRO data because:
 - Data structures are complex (for example tumor progression datasets and dynamic questionnaires)
 - Data collection process can be novel (ePRO using dynamic questionnaires on tablet computers)
 - Data is unfamiliar (efficacy and ePRO - can be very specific to the study)

Study 1 – Oncology – Bone scan data

```
9  /*Data Investigator*/
10
11  %let    RAW_DATASET      = t_bone;
12  %let    VARIABLES       = trial_name patientID;
13  %let    VALUES         = "";
14  %let    INFORMATION_FORMAT = "count";
15  %let    ORDERED_BY_COUNT = "Y";
16  %let    DISTINCT        = "N";
17
18  /*Macro*/
19  %include '\\quintiles.net\Enterprise\Sites\IEDUB\Users\q1064628\Useful
```

Log Output Data (2)

< Where Query Builder Tasks ▾

NUM_CHAR_DAT	TRIAL_NAME	PATIENTID
RESULTS	1 redacted	24

Parameters chosen to count the number of rows in the dataset (i.e., non-distinct patient IDs)

24 rows in the dataset

Study 1 – Oncology – Bone scan data

```
9  /*Data Investigator*/
10
11  %let    RAW_DATASET          = t_bone;
12  %let    VARIABLES           = trial_name patientID;
13  %let    VALUES              = "";
14  %let    INFORMATION_FORMAT   = "count";
15  %let    ORDERED_BY_COUNT     = "Y";
16  %let    DISTINCT             = "Y";
17
18  /*Macro*/
19  %include '\\quintiles.net\Enterprise\Sites\IEDUB\Users\q1064628\Us
```

Log Output Data (2)

< Where Query Builder Tasks ▾

NUM_CHAR_DAT	TRIAL_NAME	PATIENTID
RESULTS	1 redacted	14

Change the parameter to Y to count the number of patients in the dataset (i.e., distinct patient IDs)

14 patients in the dataset

Study 1 – Oncology – Bone scan data

```
9  /*Data Investigator*/
10
11  %let    RAW_DATASET      = t_bone;
12  %let    VARIABLES       = patientnumber trial_name;
13  %let    VALUES         = "";
14  %let    INFORMATION_FORMAT = "count";
15  %let    ORDERED_BY_COUNT = "Y";
16  %let    DISTINCT        = "N";
17
18  /*Macro*/
19  %include '\\quintiles.net\Enterprise\Sites\IEDUB\Users\q10646
```

Change the parameters slightly to count the number of rows per patient

Log Output Data (2)

< Where Query Builder Tasks

NUM_CHAR_DAT

RESULTS

	PATIENTNUMBER	TRIAL_NAME
1	S009-03	5
2	S007-04	4
3	S006-07	3
4	S006-09	2
5	S002-01	1
6	S002-02	1
7	S002-03	1
8	S004-02	1
9	S008-05	1

Patient S009-03 has the most rows of data (5 rows).

Most patients have 1 row of data. 4 patients have more than this –
COMPLEX DATA – REVIEW THESE ROWS CLOSELY!

Study 1 – Oncology – Bone scan data

Examination of data belonging to subject S006-07 revealed that the programmer had not accounted for bone scan done on first day of treatment (baseline bonescan):

RAW

?

	⚠ PATIENTNUMBER	⚠ VISITREFNAME	⚠ BONEDAT_DT	⚠ BONEDAT1_DT	⚠ BONEDAT2_DT
1	S006-07	DISVS	03/12/2020	05/22/2020	
2	S006-07	DISVS			07/09/2020
3	S006-07	DISVS			09/15/2020

SDTM

	⚠ USUBJID	⚠ TRMETHOD	⚠ EPOCH	⚠ TRDTC
1	redacted -S006-07	BONE SCAN	SCREENING	2020-03-12
2	redacted -S006-07	BONE SCAN	TREATMENT	2020-07-09
3	redacted -S006-07	BONE SCAN	TREATMENT	2020-09-15

Study 2 – Respiratory – eDiary questionnaires

```
9  /*Data Investigator*/
10
11  %let    RAW_DATASET      = ert_ed diary_qs;
12  %let    VARIABLES       = studyid usubjid;
13  %let    VALUES         = "" ; /*List of v
14  %let    INFORMATION_FORMAT = "count";
15  %let    ORDERED_BY_COUNT = "Y";
16  %let    DISTINCT        = "N";
17
18  /*Macro*/
19  %include '\\quintiles.net\Enterprise\Sites\IEDUB\Us
```

Log Output Data (2)

< Where Query Builder Tasks ▾

NUM_CHAR_DAT	STUDYID	USUBJID
RESULTS	1 redacted	1048454

1048454 rows in the dataset

Study 2 – Respiratory – eDiary questionnaires

```
9  /*Data Investigator*/
10
11  %let    RAW_DATASET      = ert_ed diary_qs;
12  %let    VARIABLES       = studyid usubjid;
13  %let    VALUES         = "" ;
14  %let    INFORMATION_FORMAT = "count";
15  %let    ORDERED_BY_COUNT = "Y";
16  %let    DISTINCT        = "Y";
17
18  /*Macro*/
19  %include '\\quintiles.net\Enterprise\Sites\IEDUB\Users
```

Log Output Data (2)

< Where Query Builder Tasks ▾

NUM_CHAR_DAT	STUDYID	USUBJID
RESULTS	1 redacted	679

679 subjects in the dataset

Study 2 – Respiratory – eDiary questionnaires

```
9  /*Data Investigator*/
10
11  %let    RAW_DATASET      = ert_ed diary_qs;
12  %let    VARIABLES       = usubjid studyid ;
13  %let    VALUES         = " " ;
14  %let    INFORMATION_FORMAT = "count";
15  %let    ORDERED_BY_COUNT = "Y";
16  %let    DISTINCT        = "N";
17
18  /*Macro*/
19  %include '\\quintiles.net\Enterprise\Sites\IEDUB\Users
```

Log Output Data (2)

< Where Query Builder Tasks ▾

NUM_CHAR_DAT

RESULTS

	USUBJID	STUDYID
1	redacted 1734-001	4288
2	redacted 1796-001	4208
3	redacted 1752-015	4148
4	redacted 1752-007	4128
5	redacted 1760-001	4102
6	redacted 1761-001	4088
7	redacted 1761-002	4088
8	redacted 1761-003	4088
9	redacted 1761-004	4088
10	redacted 1761-005	4088
11	redacted 1761-006	4088
12	redacted 1761-007	4088
13	redacted 1761-008	4088
14	redacted 1761-009	4088
15	redacted 1761-010	4088
16	redacted 1761-011	4088
17	redacted 1761-012	4088
18	redacted 1761-013	4088
19	redacted 1761-014	4088
20	redacted 1761-015	4088
21	redacted 1761-016	4088
22	redacted 1761-017	4088
23	redacted 1761-018	4088
24	redacted 1761-019	4088
25	redacted 1761-020	4088
26	redacted 1761-021	4088
27	redacted 1761-022	4088
28	redacted 1761-023	4088
29	redacted 1761-024	4088
30	redacted 1761-025	4088
31	redacted 1761-026	4088
32	redacted 1761-027	4088
33	redacted 1761-028	4088
34	redacted 1761-029	4088
35	redacted 1761-030	4088
36	redacted 1761-031	4088
37	redacted 1761-032	4088
38	redacted 1761-033	4088
39	redacted 1761-034	4088
40	redacted 1761-035	4088
41	redacted 1761-036	4088
42	redacted 1761-037	4088
43	redacted 1761-038	4088
44	redacted 1761-039	4088
45	redacted 1761-040	4088
46	redacted 1761-041	4088
47	redacted 1761-042	4088
48	redacted 1761-043	4088
49	redacted 1761-044	4088
50	redacted 1761-045	4088
51	redacted 1761-046	4088
52	redacted 1761-047	4088
53	redacted 1761-048	4088
54	redacted 1761-049	4088
55	redacted 1761-050	4088
56	redacted 1761-051	4088
57	redacted 1761-052	4088
58	redacted 1761-053	4088
59	redacted 1761-054	4088
60	redacted 1761-055	4088
61	redacted 1761-056	4088
62	redacted 1761-057	4088
63	redacted 1761-058	4088
64	redacted 1761-059	4088
65	redacted 1761-060	4088
66	redacted 1761-061	4088
67	redacted 1761-062	4088
68	redacted 1761-063	4088
69	redacted 1761-064	4088
70	redacted 1761-065	4088
71	redacted 1761-066	4088
72	redacted 1761-067	4088
73	redacted 1761-068	4088
74	redacted 1761-069	4088
75	redacted 1761-070	4088
76	redacted 1761-071	4088
77	redacted 1761-072	4088
78	redacted 1761-073	4088
79	redacted 1761-074	4088
80	redacted 1761-075	4088
81	redacted 1761-076	4088
82	redacted 1761-077	4088
83	redacted 1761-078	4088
84	redacted 1761-079	4088
85	redacted 1761-080	4088
86	redacted 1761-081	4088
87	redacted 1761-082	4088
88	redacted 1761-083	4088
89	redacted 1761-084	4088
90	redacted 1761-085	4088
91	redacted 1761-086	4088
92	redacted 1761-087	4088
93	redacted 1761-088	4088
94	redacted 1761-089	4088
95	redacted 1761-090	4088
96	redacted 1761-091	4088
97	redacted 1761-092	4088
98	redacted 1761-093	4088
99	redacted 1761-094	4088
100	redacted 1761-095	4088
101	redacted 1761-096	4088
102	redacted 1761-097	4088
103	redacted 1761-098	4088
104	redacted 1761-099	4088
105	redacted 1761-100	4088
106	redacted 1761-101	4088
107	redacted 1761-102	4088
108	redacted 1761-103	4088
109	redacted 1761-104	4088
110	redacted 1761-105	4088
111	redacted 1761-106	4088
112	redacted 1761-107	4088
113	redacted 1761-108	4088
114	redacted 1761-109	4088
115	redacted 1761-110	4088
116	redacted 1761-111	4088
117	redacted 1761-112	4088
118	redacted 1761-113	4088
119	redacted 1761-114	4088
120	redacted 1761-115	4088
121	redacted 1761-116	4088
122	redacted 1761-117	4088
123	redacted 1761-118	4088
124	redacted 1761-119	4088
125	redacted 1761-120	4088
126	redacted 1761-121	4088
127	redacted 1761-122	4088
128	redacted 1761-123	4088
129	redacted 1761-124	4088
130	redacted 1761-125	4088
131	redacted 1761-126	4088
132	redacted 1761-127	4088
133	redacted 1761-128	4088
134	redacted 1761-129	4088
135	redacted 1761-130	4088
136	redacted 1761-131	4088
137	redacted 1761-132	4088
138	redacted 1761-133	4088
139	redacted 1761-134	4088
140	redacted 1761-135	4088
141	redacted 1761-136	4088
142	redacted 1761-137	4088
143	redacted 1761-138	4088
144	redacted 1761-139	4088
145	redacted 1761-140	4088
146	redacted 1761-141	4088
147	redacted 1761-142	4088
148	redacted 1761-143	4088
149	redacted 1761-144	4088
150	redacted 1761-145	4088
151	redacted 1761-146	4088
152	redacted 1761-147	4088
153	redacted 1761-148	4088
154	redacted 1761-149	4088
155	redacted 1761-150	4088
156	redacted 1761-151	4088
157	redacted 1761-152	4088
158	redacted 1761-153	4088
159	redacted 1761-154	4088
160	redacted 1761-155	4088
161	redacted 1761-156	4088
162	redacted 1761-157	4088
163	redacted 1761-158	4088
164	redacted 1761-159	4088
165	redacted 1761-160	4088
166	redacted 1761-161	4088
167	redacted 1761-162	4088
168	redacted 1761-163	4088
169	redacted 1761-164	4088
170	redacted 1761-165	4088
171	redacted 1761-166	4088
172	redacted 1761-167	4088
173	redacted 1761-168	4088
174	redacted 1761-169	4088
175	redacted 1761-170	4088
176	redacted 1761-171	4088
177	redacted 1761-172	4088
178	redacted 1761-173	4088
179	redacted 1761-174	4088
180	redacted 1761-175	4088
181	redacted 1761-176	4088
182	redacted 1761-177	4088
183	redacted 1761-178	4088
184	redacted 1761-179	4088
185	redacted 1761-180	4088
186	redacted 1761-181	4088
187	redacted 1761-182	4088
188	redacted 1761-183	4088
189	redacted 1761-184	4088
190	redacted 1761-185	4088
191	redacted 1761-186	4088
192	redacted 1761-187	4088
193	redacted 1761-188	4088
194	redacted 1761-189	4088
195	redacted 1761-190	4088
196	redacted 1761-191	4088
197	redacted 1761-192	4088
198	redacted 1761-193	4088
199	redacted 1761-194	4088
200	redacted 1761-195	4088
201	redacted 1761-196	4088
202	redacted 1761-197	4088
203	redacted 1761-198	4088
204	redacted 1761-199	4088
205	redacted 1761-200	4088
206	redacted 1761-201	4088
207	redacted 1761-202	4088
208	redacted 1761-203	4088
209	redacted 1761-204	4088
210	redacted 1761-205	4088
211	redacted 1761-206	4088
212	redacted 1761-207	4088
213	redacted 1761-208	4088
214	redacted 1761-209	4088
215	redacted 1761-210	4088
216	redacted 1761-211	4088
217	redacted 1761-212	4088
218	redacted 1761-213	4088
219	redacted 1761-214	4088
220	redacted 1761-215	4088
221	redacted 1761-216	4088
222	redacted 1761-217	4088
223	redacted 1761-218	4088
224	redacted 1761-219	4088
225	redacted 1761-220	4088
226	redacted 1761-221	4088
227	redacted 1761-222	4088
228	redacted 1761-223	4088
229	redacted 1761-224	4088
230	redacted 1761-225	4088
231	redacted 1761-226	4088
232	redacted 1761-227	4088
233	redacted 1761-228	4088
234	redacted 1761-229	4088
235	redacted 1761-230	4088
236	redacted 1761-231	4088
237	redacted 1761-232	4088
238	redacted 1761-233	4088
239	redacted 1761-234	4088
240	redacted 1761-235	4088
241	redacted 1761-236	4088
242	redacted 1761-237	4088
243	redacted 1761-238	4088
244	redacted 1761-239	4088
245	redacted 1761-240	4088
246	redacted 1761-241	4088
247	redacted 1761-242	4088
248	redacted 1761-243	4088
249	redacted 1761-244	4088
250	redacted 1761-245	4088
251	redacted 1761-246	4088
252	redacted 1761-247	4088
253	redacted 1761-248	4088
254	redacted 1761-249	4088
255	redacted 1761-250	4088
256	redacted 1761-251	4088
257	redacted 1761-252	4088
258	redacted 1761-253	4088
259	redacted 1761-254	4088
260	redacted 1761-255	4088
261	redacted 1761-256	4088
262	redacted 1761-257	4088
263	redacted 1761-258	4088
264	redacted 1761-259	4088
265	redacted 1761-260	4088
266	redacted 1761-261	4088
267	redacted 1761-262	4088
268	redacted 1761-263	4088
269	redacted 1761-264	4088
270	redacted 1761-265	4088
271	redacted 1761-266	4088
272	redacted 1761-267	4088
273	redacted 1761-268	4088
274	redacted 1761-269	4088
275	redacted 1761-270	4088
276	redacted 1761-271	4088
277	redacted 1761-272	4088
278	redacted 1761-273	4088
279	redacted 1761-274	4088
280	redacted 1761-275	4088
281	redacted 1761-276	4088
282	redacted 1761-277	4088
283	redacted 1761-278	4088
284	redacted 1761-279	4088
285	redacted 1761-280	4088
286	redacted 1761-281	4088
287	redacted 1761-282	4088
288	redacted 1761-283	4088
289	redacted 1761-284	4088
290	redacted 1761-285	4088
291	redacted 1761-286	4088
292	redacted 1761-287	4088
293	redacted 1761-288	4088
294	redacted 1761-289	4088
295	redacted 1761-290	4088
296	redacted 1761-291	4088
297	redacted 1761-292	4088
298	redacted 1761-293	4088
299	redacted 1761-294	4088
300	redacted 1761-295	4088
301	redacted 1761-296	4088
302	redacted 1761-297	4088
303	redacted 1761-298	4088
304	redacted 1761-299	4088
305	redacted 1761-300	4088
306	redacted 1761-301	4088
307	redacted 1761-302	4088
308	redacted 1761-303	4088
309	redacted 1761-304	4088
310	redacted 1761-305	4088
311	redacted 1761-306	4088
312	redacted 1761-307	4088
313	redacted 1761-308	4088
314	redacted 1761-309	4088
315	redacted 1761-310	4088
316	redacted 1761-311	4088
317	redacted 1761-312	4088
318	redacted 1761-313	4088
319	redacted 1761-314	4088
320	redacted 1761-315	4088
321	redacted 1761-316	4088
322	redacted 1761-317	4088
323	redacted 1761-318	4088
324	redacted 1761-319	4088
325	redacted 1761-320	4088
326	redacted 1761-321	4088
327	redacted 1761-322	4088
328	redacted 1761-323	4088
329	redacted 1761-324	4088
330	redacted 1761-325</	

Study 2 – Respiratory – eDiary questionnaires

```
9  /*Data Investigator*/
10
11  %let    RAW_DATASET      = ert_ed diary_qs;
12  %let    VARIABLES       = usubjid visit QSDTC studyid ;
13  %let    VALUES         = "" ;
14  %let    INFORMATION_FORMAT = "count";
15  %let    ORDERED_BY_COUNT = "Y";
16  %let    DISTINCT        = "N";
17
18  /*Macro*/
19  %include '\\quintiles.net\Enterprise\Sites\IEDUB\Users\q1064628
```

Log Output Data (2)

< Where Query Builder Tasks ▾

NUM_CHAR_DAT

RESULTS

		USUBJID	VISIT	QSDTC	STUDYID
1	redacted	-0100-001	VIS1	2021-02-08T08:32:23	51
2	redacted	-0202-001	VIS1	2021-04-19T07:45:00	51
3	redacted	-0202-001	VIS2	2021-05-31T07:50:51	51
4	redacted	-0202-001	VIS3	2021-07-08T07:55:57	51
5	redacted	-0202-002	VIS1	2021-04-16T08:32:30	51

Subjects visits have between 3 and 51 rows

52130	redacted	2-1819-005	VIS1	2021-05-13T08:44:22	3
52131	redacted	2-1819-005	VIS2	2021-06-24T08:33:28	3
52132	redacted	2-1819-006	VIS1	2021-07-12T09:10:26	3
52133	redacted	2-1819-007	VIS1	2021-07-14T09:40:01	3

Study 2 – Respiratory – eDiary questionnaires

```
9  /*Data Investigator*/
10
11  %let    RAW_DATASET          = ert_ed diary_qs;
12  %let    VARIABLES           = QSCAT studyid ;
13  %let    VALUES              = "" ;
14  %let    INFORMATION_FORMAT   = "count";
15  %let    ORDERED_BY_COUNT     = "Y";
16  %let    DISTINCT             = "N";
17
18  /*Macro*/
19  %include '\\quintiles.net\Enterprise\Sites\IEDUB\
```

Log Output Data (2)

< Where Query Builder Tasks

NUM_CHAR_DAT	QSCAT	STUDYID
RESULTS		
1	DailyDiary	987860
2	SGRQ	53499
3	EQ-5D-5L	4044
4	TDI	1671
5	BDI	1380

5 different questionnaires are listed in the data. DailyDiary has the most rows.

Study 2 – Respiratory – eDiary questionnaires

```
9  /*Data Investigator*/
10
11  %let    RAW_DATASET      = ert_ed diary_qs;
12  %let    VARIABLES       = QSCAT QSTEST ;
13  %let    VALUES         = " " ;
14  %let    INFORMATION_FORMAT = "count";
15  %let    ORDERED_BY_COUNT = "Y";
16  %let    DISTINCT        = "Y";
17
18  /*Macro*/
19  %include '\\quintiles.net\Enterprise\Sites\IEDUB\U
```

Log Output Data (2)

< Where Query Builder Tasks ▾

NUM_CHAR_DAT

RESULTS

	QSCAT	QSTEST
1	SGRQ	51
2	DailyDiary	20
3	EQ-5D-5L	6
4	BDI	3
5	TDI	3

The questionnaires can produce between 3 and 51 rows of data.

Study 2 – Respiratory – eDiary questionnaires

```
9  /*Data Investigator*/
10
11  %let    RAW_DATASET      = ert_ed diary_qs;
12  %let    VARIABLES       = QSCAT VISIT studyid ;
13  %let    VALUES         = "" ;
14  %let    INFORMATION_FORMAT = "count";
15  %let    ORDERED_BY_COUNT = "Y";
16  %let    DISTINCT        = "N";
17
18  /*Macro*/
19  %include '\\quintiles.net\Enterprise\Sites\IEDUB\Users\q
```

Log Output Data (2)

< Where Query Builder Tasks

NUM_CHAR_DAT

RESULTS

	QSCAT	VISIT	STUDYID
1	DailyDiary		987860
2	SGRQ	VIS1	23766
3	SGRQ	VIS2	15402
4	SGRQ	VIS3	10455
5	SGRQ	VIS5	3621
6	EQ-5D-5L	VIS1	2796
7	BDI	VIS1	1374
8	EQ-5D-5L	VIS3	1224
9	TDI	VIS2	873
10	TDI	VIS3	579
11	SGRQ	UNS1	255
12	TDI	VIS5	210
13	EQ-5D-5L	UNS1	24
14	TDI	UNS1	9
15	BDI	UNS1	6

This table shows what visits each questionnaire is collected at.

There are a small number of UNSCHEDULED visits listed –
COMPLEX DATA – REVIEW THESE ROWS CLOSELY!

Study 2 – Respiratory – eDiary questionnaires

Examination of data associated with UNSCHEDULED TDI questionnaires reveals mishandling of unscheduled visits:

RAW

	USUBJID	QSTEST	QSCAT	VISIT	QSDTC	
1	redacted	-0215-002	Change in Functional Impairment	TDI	UNS1	2021-06-28T08:21:49
2	redacted	-0215-002	Change in Magnitude of Task	TDI	UNS1	2021-06-28T08:21:49
3	redacted	-0215-002	Change in Magnitude of Effort	TDI	UNS1	2021-06-28T08:21:49
4	redacted	-1757-008	Change in Functional Impairment	TDI	UNS1	2021-06-01T07:12:20
5	redacted	-1757-008	Change in Magnitude of Task	TDI	UNS1	2021-06-01T07:12:20
6	redacted	-1757-008	Change in Magnitude of Effort	TDI	UNS1	2021-06-01T07:12:20
7	redacted	-1819-003	Change in Functional Impairment	TDI	UNS1	2021-06-23T07:48:45
8	redacted	-1819-003	Change in Magnitude of Task	TDI	UNS1	2021-06-23T07:48:45
9	redacted	-1819-003	Change in Magnitude of Effort	TDI	UNS1	2021-06-23T07:48:45

SDTM

	USUBJID	QSTEST	QSCAT	VISIT	QSDTC	
1	redacted	-0215-002	Change in Functional Impairment	TDI		2021-06-28T08:21:49
2	redacted	-0215-002	Change in Magnitude of Task	TDI		2021-06-28T08:21:49
3	redacted	-0215-002	Change in Magnitude of Effort	TDI		2021-06-28T08:21:49
4	redacted	1757-008	Change in Functional Impairment	TDI	Unscheduled 4.01	2021-06-01T07:12:20
5	redacted	1757-008	Change in Magnitude of Task	TDI	Unscheduled 4.01	2021-06-01T07:12:20
6	redacted	1757-008	Change in Magnitude of Effort	TDI	Unscheduled 4.01	2021-06-01T07:12:20
7	redacted	-1819-003	Change in Functional Impairment	TDI		2021-06-23T07:48:45
8	redacted	1819-003	Change in Magnitude of Task	TDI		2021-06-23T07:48:45
9	redacted	-1819-003	Change in Magnitude of Effort	TDI		2021-06-23T07:48:45

Study 3 – Oncology – eDiary questionnaire

```
9  /*Data Investigator*/
10
11  %let    RAW_DATASET      = EORTCLC;
12  %let    VARIABLES       = study subject;
13  %let    VALUES         = " " ;
14  %let    INFORMATION_FORMAT = "count";
15  %let    ORDERED_BY_COUNT = "Y";
16  %let    DISTINCT        = "N";
17
18  /*Macro*/
19  %include '\\quintiles.net\Enterprise\Sites\IEDUB\
```

Log

Output Data (2)

< Where | Query Builder | Tasks ▾

NUM_CHAR_DAT

STUDY

SUBJECT

RESULTS

1

redacted

338

Dataset has 338 rows

Study 3 – Oncology – eDiary questionnaires

```
9  /*Data Investigator*/
10
11  %let    RAW_DATASET      = EORTCLC;
12  %let    VARIABLES       = study visit;
13  %let    VALUES         = "" ;
14  %let    INFORMATION_FORMAT = "list";
15  %let    ORDERED_BY_COUNT = "Y";
16  %let    DISTINCT        = "Y";
17
18  /*Macro*/
19  %include '\\quintiles.net\Enterprise\Sites\IEDUB
```

Log Output Data (2)

< Where Query Builder Tasks

NUM_CHAR_DAT	STUDY	VISIT
RESULTS	1 redacted	700.0
	2 redacted	510.0
	3 redacted	310.0
	4 redacted	300.0
	5 redacted	290.0
	6 redacted	90.00
	7 redacted	80.00
	8 redacted	70.00
	9 redacted	60.00
	10 redacted	50.00
	11 redacted	40.00
	12 redacted	30.00
	13 redacted	20.00
	14 redacted	10.00

Visit numbering has unusual progression from 10, 20, 30, 40 initially to 310, 510, 700 at the end?

Study 3 – Oncology – eDiary questionnaires

```
9  /*Data Investigator*/
10
11  %let    RAW_DATASET      = EORTCLC;
12  %let    VARIABLES       = visit study;
13  %let    VALUES         = "" ;
14  %let    INFORMATION_FORMAT = "count";
15  %let    ORDERED_BY_COUNT = "Y";
16  %let    DISTINCT        = "N";
17
18  /*Macro*/
19  %include '\\quintiles.net\Enterprise\Sites\IEDUB\U
```

Log Output Data (2)

< Where Query Builder Tasks ▾

NUM_CHAR_DAT ⊕ VISIT ⊕ STUDY

RESULTS

1	10.00	102
2	20.00	85
3	30.00	65
4	40.00	41
5	50.00	26
6	60.00	8
7	70.00	4
8	80.00	1
9	90.00	1
10	290.0	1
11	300.0	1
12	310.0	1
13	510.0	1
14	700.0	1

Each of the strange visits has only one row in the dataset – **COMPLEX DATA – REVIEW THESE ROWS CLOSELY!**

Study 3 – Oncology – eDiary questionnaires

Examination of data associated with unusual visit numbers reveals possible mishandling of these visits – visit 50.00 and 290.00 are both assigned to Cycle 7 – Day 1 despite occurring nearly 1 month apart:

SDTM

	USUBJID	QSCAT	VISIT	QSDTC	
1	redacted	003001	EORTC QLQ-LC13	Cycle 1 - Day 1	2019-01-10T09:29
2	redacted	003001	EORTC QLQ-LC13	Cycle 2 - Day 1	2019-02-07T09:03
3	redacted	003001	EORTC QLQ-LC13	Cycle 3 - Day 1	2019-03-07T09:10
4	redacted	003001	EORTC QLQ-LC13	Cycle 5 - Day 1	2019-05-02T13:10
5	redacted	003001	EORTC QLQ-LC13	Cycle 7 - Day 1	2019-05-30T08:41
6	redacted	003001	EORTC QLQ-LC13	Cycle 7 - Day 1	2019-06-27T09:55
7	redacted	003001	EORTC QLQ-LC13	Cycle 9 - Day 1	2019-08-22T09:56
8	redacted	003001	EORTC QLQ-LC13	Cycle 11 - Day 1	2019-10-17T09:26

RAW

	SUBJECT	VISIT	VIS_DAT	QSCAT
1	E6003001	10.00	2019-01-10	EORTC QLQ-LC13
2	E6003001	20.00	2019-02-07	EORTC QLQ-LC13
3	E6003001	30.00	2019-03-07	EORTC QLQ-LC13
4	E6003001	40.00	2019-05-02	EORTC QLQ-LC13
5	E6003001	50.00	2019-05-30	EORTC QLQ-LC13
6	E6003001	290.0	2019-06-27	EORTC QLQ-LC13
7	E6003001	300.0	2019-08-22	EORTC QLQ-LC13
8	E6003001	310.0	2019-10-17	EORTC QLQ-LC13

30

Discussion

- Data Investigator Macro helped identify unexpected complex data in the datasets reviewed.
- Review of complex data revealed the potential for SDTM mapping issues in the datasets reviewed and facilitated its correction.
- Amount of “at risk” data identified was very low (< 1 % of total data). One of the core tenets of Good Clinical Practice requires us to ensure **integrity of the data and validity of analyses**.
- The tool **empowers reviewers** to increase the depth and quality of the review process without increasing the time required to do so.

Future directions

- SDTM mapping processes:
 - » Integrate Data Investigator Macro at all steps:
 - SDTM specification writing
 - Programming
 - Review.
- Risk Based Validation of SDTM mapping:
 - » Alternative approach to validation.
 - » Allows for more appropriate validation methods depending on the risk profile/complexity.
 - » Data Investigator Macro could form part of a suite of tools.
- Future development:
 - » Create a web based GUI to match other IQVIA tools.
 - » Extend functionality of the tool to facilitate programmatic comparison of datasets (Raw and SDTM) – DATA COMPARATOOL

Any Questions?

- If so please contact:
 - John Monaghan (John.Monaghan@IQVIA.com)

Appendix slide 1 – Programming approach

- Programming approach
 - Include statement used to call macro instead of traditional macro call approach due to technical limitations of Proc IML submit statement (can't be wrapped in a macro).
 - Proc IML used to create lists from parameters VARIABLES and VALUES.
 - Proc IML and SAS datasets used to dynamically construct the following macrovariables:

&SQL_distinct

&list_of_variables

&from_table

&where_statement

&var_list_del_by_str;

&ordered_by_count

- Proc SQL used to pull macrovariables into a single dynamic query which produces the output

```
194 proc sql;  
195     create table results as select &SQL_distinct  
196     &list_of_variables  
197     from &from_table  
198     &where_statement  
199     group by &var_list_del_by_str;  
200     run;  
201  
202     %if &ordered_by_count = "Y" %then  
203         %do;  
204  
205     proc sort data=results out=results;  
206         by descending &var_last;  
207     run;  
208     %end;
```