Analysis Datasets for Baseline Characteristics and Exposure Time for Multi-stage Clinical Trials

Kim Umans, Katherine Riester, Manjit McNeill, Wei Liu, and Lukasz Kniola

PhUSE 2016, CD06, Barcelona

Introduction and Disclaimer

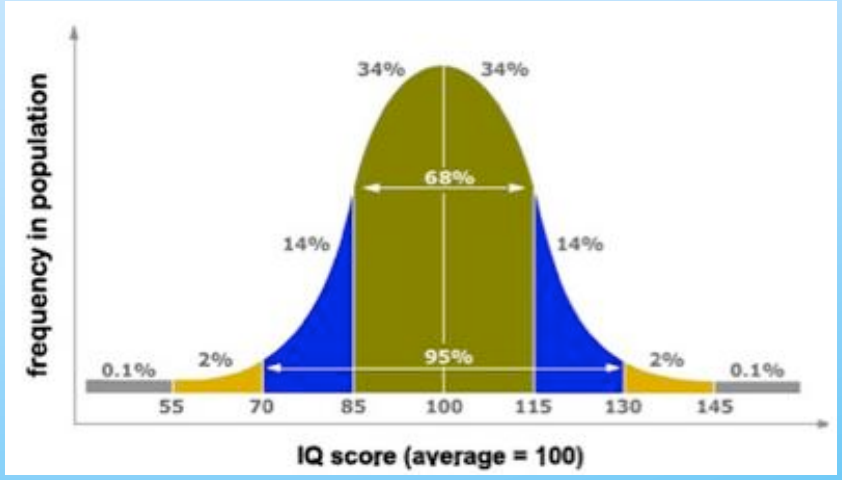
* Thank you to the organizers for their work on the CD conference stream and for including our paper
* All authors are employees of Biogen, Inc.
* This presentation and accompanying paper represent the opinions of the authors only and do not represent the position of their employer
* The interpretations of CDISC standards presented are not necessarily correct and do not represent the CDISC consortium

In the beginning…

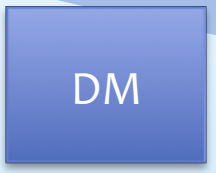
In almost all clinical trials, **baseline** information is of important interest

* Summarizing for its own sake
    * Who are the patients in the trial?
    * Are treatment arms comparable?
    * Will the results be generalizable to the expected real world patient population?

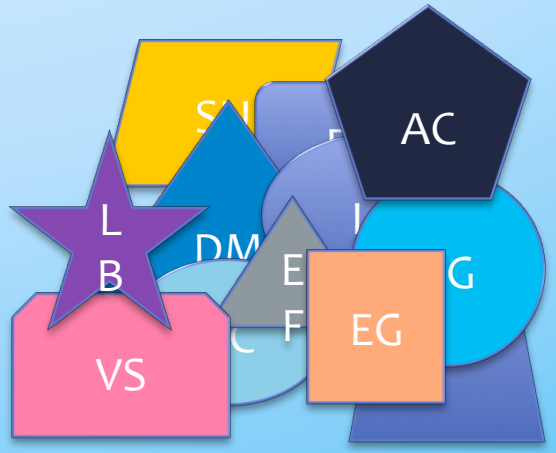
* Using baseline characteristics as covariates (i.e., explanatory variables or prognostic factors) in subgroup analyses or inferential statistical models



Baseline data comes in all shapes and sizes!



- Demographic: age, sex, race, country
- Medical history: for indication and other, including details for indication
- Other history: smoking
- Previous treatment: for indication and other
- Current disease severity
- Baseline vital signs: weight, height
- Miscellaneous: dominant hand



CONSEQUENCE: lots of SDTM domains get involved

Challenge: need an ADaM dataset to hold this diverse baseline data

* "Traditional" solution – put all baseline information in ADSL
* Problem #1 (minor) – too many baseline variables makes ADSL unwieldy
* Problem #2 (major) – modern trials are often complex with multiple, competing versions of "baseline"

Many meanings for “baseline”

* E.g., blinded phase + open label phase, in lieu of an extension study
  * Baseline at start of blinded phase?
  * Baseline at start of open label phase?
  * Baseline at first start of active drug?
* Substudies nested within bigger trials, in lieu of a stand alone study
* Even studies that start simple may be “complexified” later

Proposed solution: ADBASE

* Separate ADaM dataset to hold baseline data
* OCCDS (occurrence) structure, with one record per subject per applicable baseline type
* Critical variables: BASETYPE, BASETYPN, BASERFDT
* BASETYPE examples in a 2-part study:
  * First dose in Part 1
  * First dose in Part 2
  * First dose of active drug

Sample ADBASE rows

Subject	BASE-TYPN	BASETYPE	BASERFDT	AAGE	WEIGHTBL	DIAGMOS
01	1	Period 1	01Sep2014	24	70	15
01	2	Period 2	02Jan2015	25	71	19
01	3	Start of Active Drug	02Jan2015	25	71	19
02	1	Period 1	11Oct2014	50	67	31
03	1	Period 1	01Mar2015	48	81	24
03	3	Start of Active Drug	01Mar2015	48	81	24

Sample ADBASE specification

Variable	Label	Derivation / Codelist
BASETYPE	Baseline Type	Create one record for each applicable baseline type for the subject with Codelist: “Period 1”, “Period 2”, “Start of Active Drug”, “Substudy 1”
BASETYPN	Baseline Type (N)	1, 2, 3, 4 In correspondence with BASETYPE
BASERFDT	Baseline Reference Date	AP01SDT if BASETYPN=1, AP02SDT if BASETYPN=2, AP01SDT if BASETYPN=3 and TRT01PN=1, AP02SDT if BASETYPN=3 and TRT01PN=0, SS01SDT if BASETYPN=4
AAGE	Analysis Age	Age in years on BASERFDT
WEIGHTBL	Baseline Weight	VSSTRESN for VSTESTCD=‘WEIGHT’ at last record on or before BASERFDT

Note: all dates in derivation of BASERFDT expected to be in ADSL.

Exposure time measures

* Similar challenge: as trial complexity grows, so do the number of measurements of exposure time that are needed for analyses
* Similar solution: move the exposure time variables out of ADSL and into their own ADaM dataset

ADTEXP

* BDS structure, with one record per subject per parameter
* Critical variables: PARAM, ASTDT, AENDT, AVAL
  * AVAL calculated as duration: AENDT – ASTDT + 1 day, converted to appropriate units for the parameter
* Critical metadata: In addition to the usual variable specification, ADTEXP requires a parameter lookup table

Sample ADTEXP lookup table

PARAMN	PARAM	units	ASTDT	AENDT
1	Time on Study Period 1 (days)	days	AP01SDT	AP01EDT
2	Time on Study Period 1 (weeks)	weeks	AP01SDT	AP01EDT
3	Time on Study Period 2 (days)	days	AP02SDT	AP02EDT
4	Time on Treatment Period 1 (days)	days	TR01SDT	TR01EDT
5	Time Exposed to Treatment Period 1 (days)	days	TR01SDT	TR01EDT + 25 days, or AP01EDT whichever first
6	Time Exposed to Active Drug (days)	days	TR01SDT if TRT01PN=1, TR02SDT if TRT02PN=0	TRTEDT + 25 days, or STDEDT whichever first

Note: all start and end dates expected to be in ADSL.

Sample ADTEXP rows

SUBJID	PARAM	ASTDT	AENDT	AVAL
004	Time on Study Period 1 (days)	01Sep2014	02Jan2015	124
004	Time on Study Period 1 (weeks)	01Sep2014	02Jan2015	17.714
004	Time on Study Period 2 (days)	02Jan2015	01May2015	120
004	Time on Study (days)	01Sep2014	01May2015	243
004	Time on Treatment Period 1 (days)	02Sep2014	27Dec2014	117
004	Time Exposed to Treatment Period 1 (days)	02Sep2014	02Jan2015	123
005	Time on Study Period 1 (days)	11Oct2014	11Nov2014	32

Advantages

* ADBASE and ADTEXP datasets move a potentially large, cumbersome subset of variables out of ADSL
  * Late changes to these datasets can be made without affecting ADSL, causing fewer ripples downstream
* ADBASE allows for ADSL to remain in the traditional one-record-per-subject structure, even for integrated analyses
  * Subsetting to the appropriate BASETYPE and merging gives a baseline type specific version of ADSL
* Allows for more detailed metadata

Moltes gràcies

Thank you!!!

danke schön

¡muchas gracias!