



INTEGO
GROUP

DH09: May Your Data Pool Become Super Cool



Amsterdam 2019



ISS – Integrated Summary of Safety

- summary of the clinical studies
- treatment exposure information
- demographics and background characteristics
- adverse events summary
- clinical laboratory assessments details

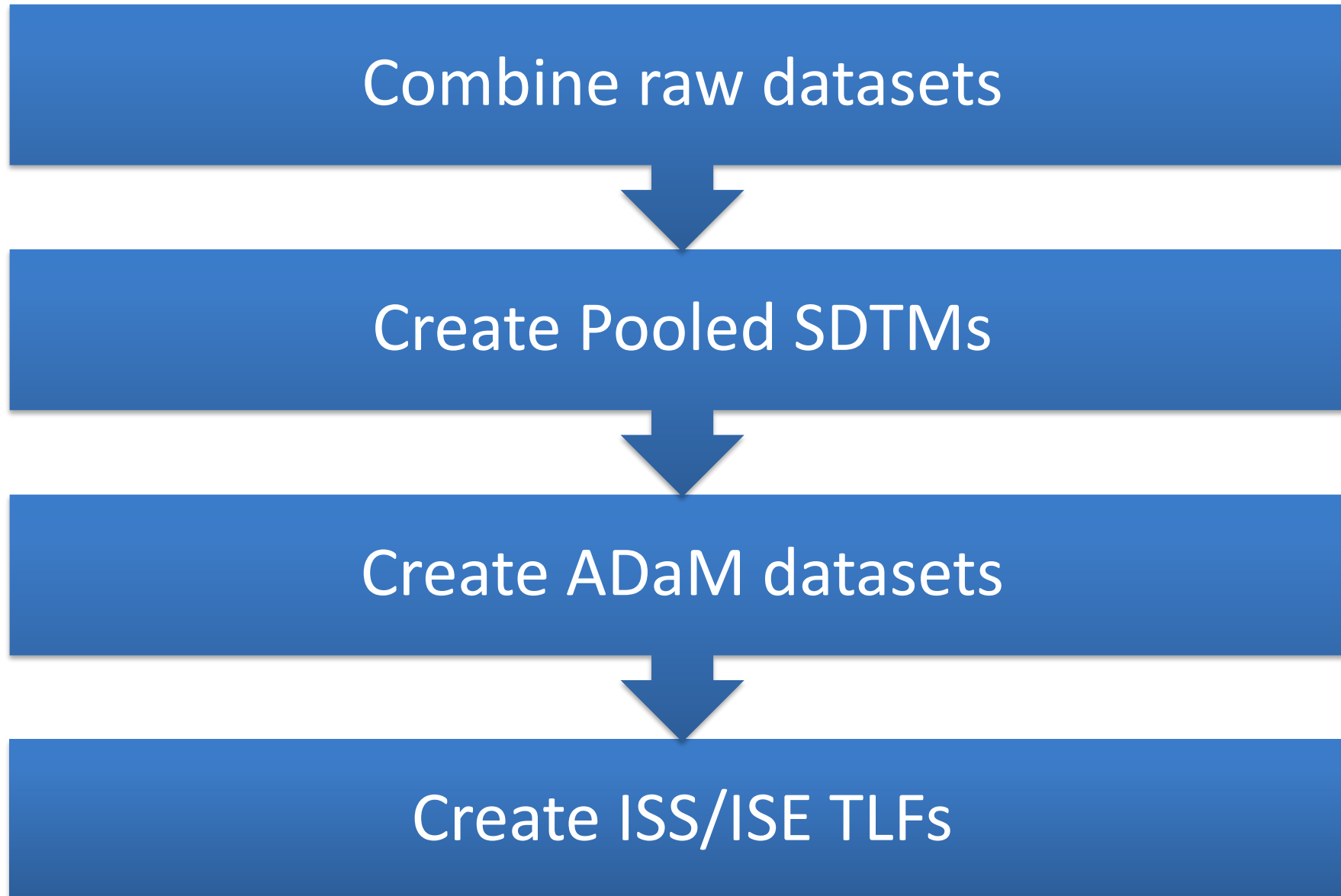
ISE – Integrated Summary of Effectiveness

- Primary endpoints



POOLING DATA **RAW**





Points to remember

- structure of raw datasets could vary from study to study
- different type of variables for the same type of information
- no global standards



POOLING DATA **SDTM**



Combine SDTM datasets



Create pooled ADaM datasets



Create ISS/ISE TLFs

Points to remember

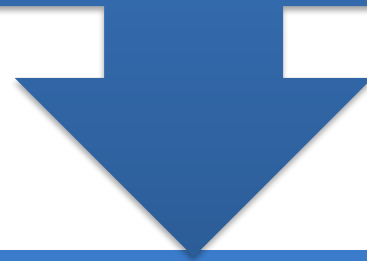
- structure of SDTMs is strictly defined by CDISC
- SDTMs are validated



POOLING DATA **ADAM**



Create pooled ADaM
datasets



Create ISS/ISE TLFs

Points to remember

- same name of variable has different content
- derivations and algorithms could be different from study to study
- only information required for the analysis may be included to the ADaMs



MEDDRA, CTCAE, WHODRUG



Points to know:

- use the most current versions of MedDRA and WhoDRUG
- check if coding specialist is available
- when pooling ADaMs, be careful with the toxicity grades
- no need to up-version the dictionaries for legacy studies



LEGACY STUDIES



If legacy studies are part of ISS the following point should be considered:

- studies may have been conducted many years ago and/or may span across several years. This means that there is a big chance that data will not be standardized using CDISC.
- studies may have been developed by multiple companies and as a result there may be a lack of consistency across studies, the different method of data collection may be used, etc.
- such studies may not reflect current regulatory requirements, because these studies are pre-ICH.



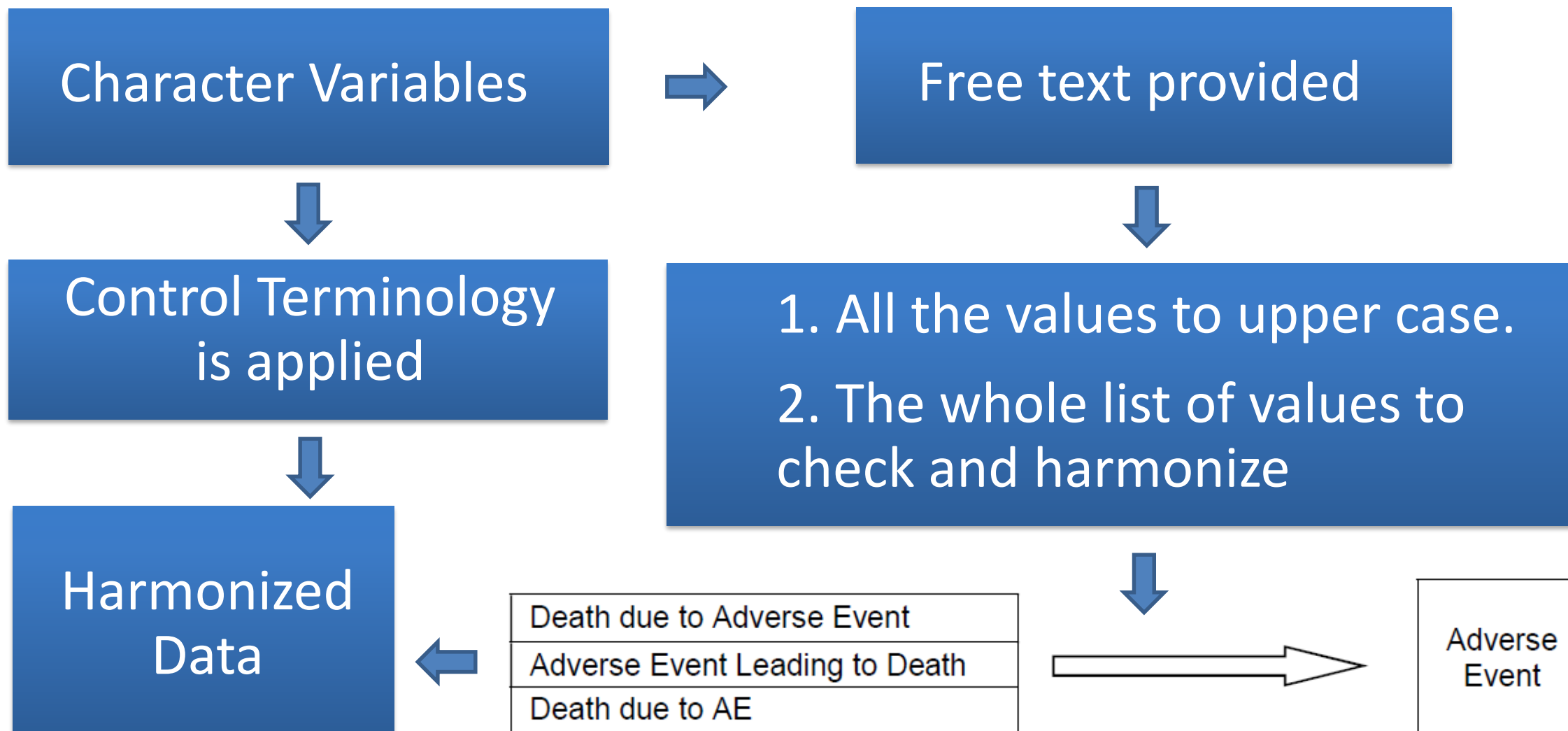
DATA HARMONIZATION



Potential tricky questions. Variable attributes.

- Pinnacle 21 (if pooled on SDTM level)
- “must have” check – all character variables to \$200 before set

Potential tricky questions. Variables Values.



Potential tricky questions. Variables Values.

- **Laboratory, Vitals, ECG:** Normal Ranges to be standardized. Standard unit to be used per each particular test
- **Test names to be checked + QNAM/QVAL from SUPP domains**
- **Partial dates imputation.** Special attention to AE start/end date, since influences TEAE flag
- **Site ID, subject ID unification among the studies**

Variables Harmonization. Summary.

For variables that are common across the pooled studies:

- Determine the most common definition of the variable
- If any imputation is assumed, define the single way of how to proceed
- Make variables in all studies consistent with the “majority”
- Map code lists of categorical or ordinal variables into a common standard if necessary



VISITS AND ANALYSIS WINDOWING



Different visits schedule:

Visit name	Visit Number		
	Study A	Study B	Pooling
Screening	-1	-1	-1
Day 1	1	1	1
Week 3	2	NA	2
Week 4	NA	2	3
Week 6	3	NA	4
Week 8	NA	3	5
Week 9	4	NA	6
Week 12	5	4	7

Different spelling:

‘Follow-up’, ‘Follow-Up’,
‘Follow-up Visit 10’, ‘Follow-up, Day 28’

Baseline definition:

- before or before/on treatment start
- study day 1 results without time to be considered?
- baseline type: last / worst / average

The same visit name, but different window:

	Week 12 Analysis window	Week 12 Target Day
Study A (visiting schedule is once per week)	[78 ; 84]	81
Study B (visiting schedule is once per 3 weeks)	[74 ; 95]	84

Visits Harmonization

- **All the unique values** of study ID/ visit name and number to be selected
- Check all the **individual study protocols** in terms of schedule of assessments
- Create the whole list of visits in the pool database. **Unify the names** of the visits. **Define** a logical **order** for them
- Use **non-integer values to distinguish** between visit numbers:
“Week 3” → 3; “Week 3 Day 2” → 3.02
- **Unscheduled visits to be numbered logically**
- Study **period/phase** is included **in the name** of the visit:
“Double Blind Period, Week 2” → 1000.2;
“Open Label Extension, Week 2” → 2000.2

**Work on the
required visits
only!**



Treatment Groups Assignment



The points to be checked for treatment groups:

- Regimen, dose frequency and value
- Studies with multiple periods of treatment as well as for the cross-over designed studies (some of the periods to be excluded?)
- Studies with up/down titrations – typical for Phase I

The points to be checked for treatment groups:

- Add-on therapies to be considered, since can influence the treatment effect
- Lots of different treatments to be combined: placebo-controlled studies vs. uncontrolled or active-controlled studies
- Treatment groups mapping to be well documented. Output headers are not necessary very informative → include footnotes with the explanation
- All the answers should be well discussed in advance with the team



Pooling Strategy. Programming point of view



- Develop the pooling strategy as early as possible, define the purpose, objectives and requirements of the pooled analysis
- Define which studies and which data should be pooled to add value to the integrated analysis
- Data standards to be used for each individual study to be on the safe side
- All the codes for individual studies optimal and well organized, not depending on the number of patients / studies / treatment groups and so on
- Ensure that the pooled data represent the data from individual studies consistently. If there are any discrepancies due to known algorithm changes they have to be documented

Pooling = huge number of data, that's why try to:

- **Avoid sorting** datasets or re-setting if this is not needed
- Make **use of as few data steps as possible**
- Make use of **DROP or KEEP statement** to remove **un-needed variables** even in temporary datasets
- Make **use of WHERE statement** while setting instead of making record selection in the data steps
- **Drop temporary** datasets when they are not needed anymore
- Carefully **reduce LENGTH** of the variables
- Use **indexes or hash-objects** where possible



Conclusion



To sum up:

- ISS/ISE **planning** process should be **started as soon as the decision** to submit integrated data is reached
- **Pooled datasets** can be created **from the existing raw or derived datasets**.
Which way to use depends on the availability of standard and derived data, common endpoints, etc.

- It is important to **understand the rationale** of pooling data and **the purpose** it is created for. The default is **not to pool everything** but to do so when it adds value
- **Even** with a well-defined data **standard in place** a certain amount of **variability** is to be **expected**
- **Detailed documentation** is important to be prepared as well as **the validation** that may be based **on reports or by study checks**



ANY QUESTIONS?





THANK YOU

Daniil Shliakhov

email: Daniil.Shlyakhov@intego-group.com

Anastasiia Tiurdo

email: Anastasiia.Tiurdo@intego-group.com

Intego Group, LLC

555 Winderley Place, Ste. 129,

Maitland, FL 32751

Phone: +1 (407) 512-1006

www.intego-group.com

