

我们为什么需要CDASH和 SDTM 两个标准！

John Wang@dMed, C3C

原作者：Shannon Labout, VP Education, CDISC

Kit Howard, Director Education, CDISC



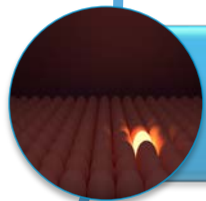
今天您将学到



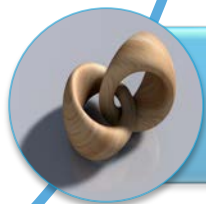
SDTM和CDASH是什么



指出SDTM和CDASH相同点

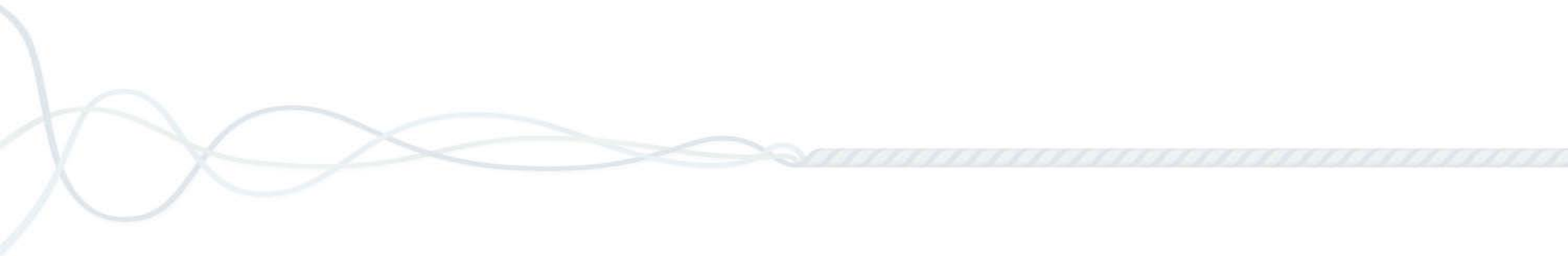


它们如何不同，为什么



为何最好两者一起使用

第1节: SDTM和CDASH是什么



SDTM/IG

定义了采集数据的一致展现方式

- 按照数据间的逻辑关系，汇总到不同的主题中
- 命名规则, 数据类型, 受控术语
- 所有清理干净的数据，外加少许衍生变量

可支持

- 用于递交的合并分析, 数据仓库, 注册
- 监管机构审评, IRB汇报, 安全监督
- 假设检验分析

CDASH/IG

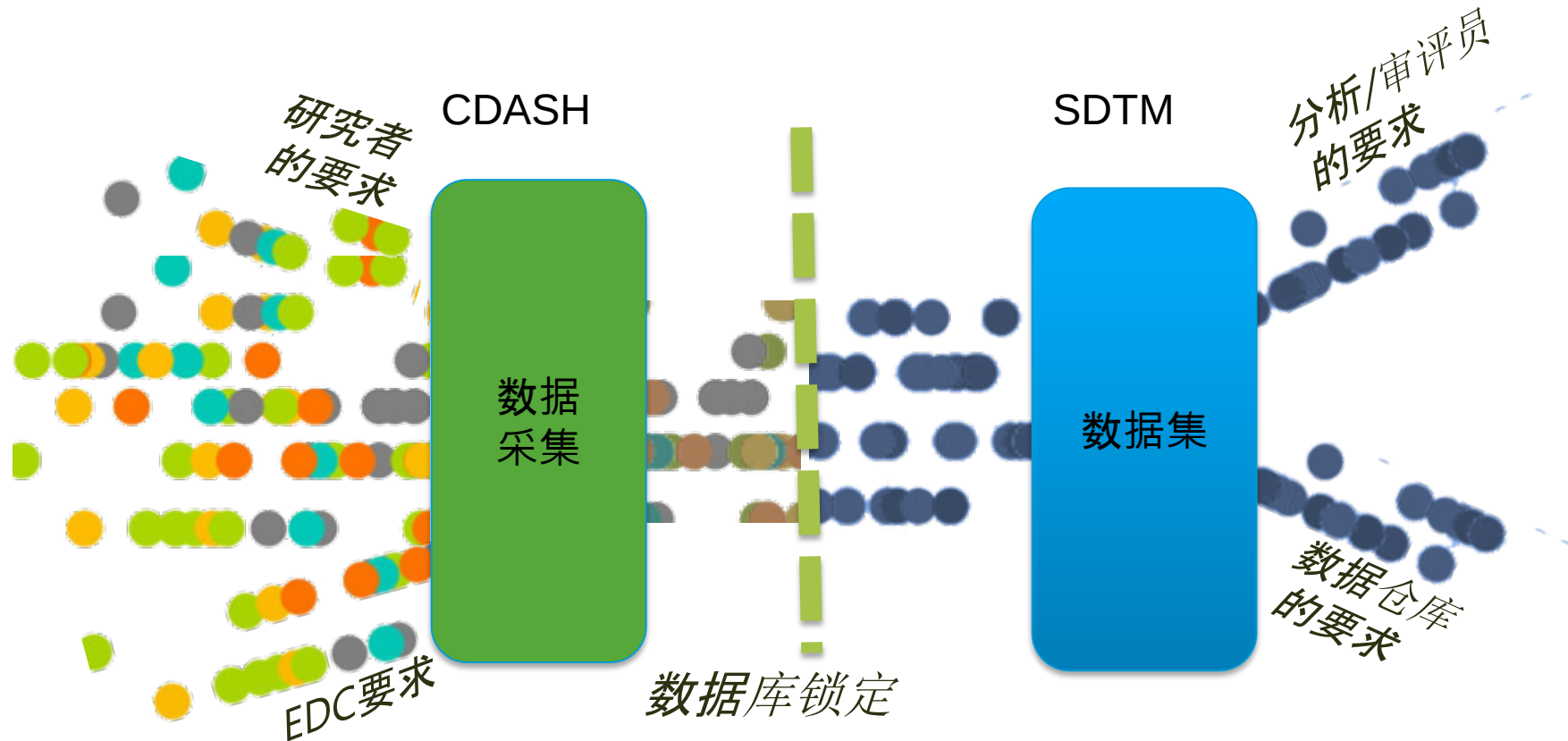
数据采集和管理

- CRF设计一致 – 问题，答案，假设
- 便于研究中心使用
- 数据质量检查规则，可满足SDTM中数据质量的前提假设

可支持

- 以终为始
- 数据从采集到递交的过程平顺
- 数据真实性和可溯源到原始数据

为什么理解SDTM非常关键

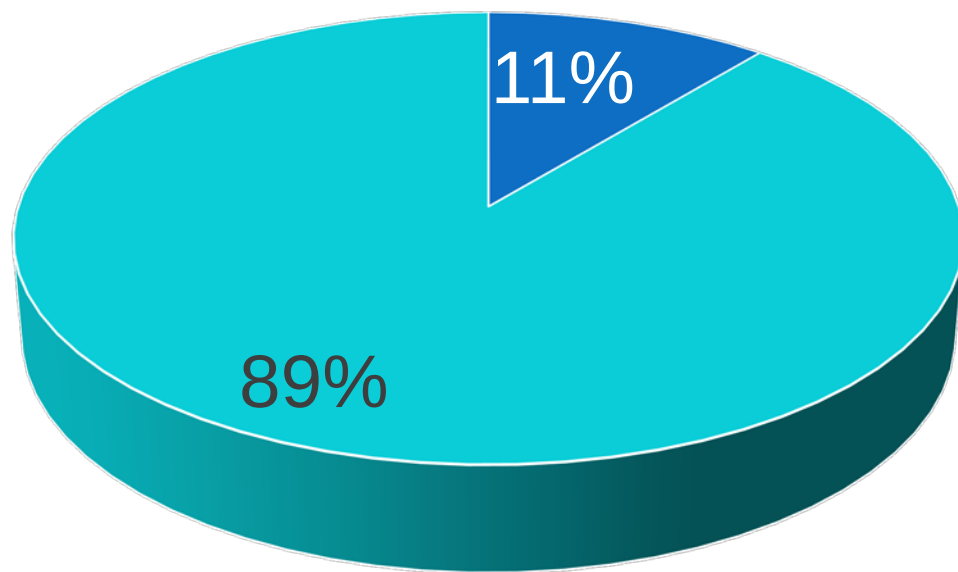


要放入数据仓库
将会是噩梦

设计时考虑了合并分析，数据仓库需求

第2节：相似点

变量重叠: 完全相同 + 标准化映射 (mappings)



■ 不同

■ 完全相同+标准化映射

更多相似点



相同的受控术语
相同的数据假设
相同的将数据按照域的组织方式

- 仅仅包含未符合IE标准的记录
- DA用于追踪药物的发放，而不用于治疗的依从性

第3节：不同点

设计理念驱动

Subject ID: 1|0|2|3 SUBJID AE

Were Any Adverse Events Experienced?

☒ No AEYN Not Submitted

☐ Yes - (please complete the form)

Adverse Event - (please use precise medical terminology)

AETERM

Start Date and Time

Day

Month

Year

AESTDAT
AESTDTC

缺失数据

SDTM

Show me the data, not lack of data

CDASH

Absence of evidence is not evidence of absence: must check that missing data is missing

SDTM

USUBJID	AETERM	AESTDTC	AEENDTC
2006-34-1022	HEADACHE	2006-10-14	2006-10-20
2006-34-1022	RASH	2006-10-14	2006-10-21
2006-34-1024	TIRED	2006-09-24	2006-09
2006-34-1024	HEADACHE	2006-09-24	2006-09-24

原理

SDTM assumes that if there is no record then nothing happened. This only works if the data was checked in data capture, which requires a question and record (e.g., Were there any AEs?)

人类可读vs机器可读的数据

SDTM

USUBJID	HOTERM	HOSTDTC	HODUR
ABC1201	Hospitalization	2011-08-06	P6H

CDASH

SUBJID	HOTERM	HOSTDAT	HOCDUR	HOCDURU
1201	Hospitalization	06-AUG-2011	6	HOURS

HOSPITALIZATION

HOTERM

HO

Was the subject hospitalized during this study?

☐ No

HOYN

Not Submitted

☒ Yes If "Yes," please complete the following:

Admission Date:

06

AUG

2011

HOSTDAT

HOSTDTC

HOCDUR

HOCDURU

Duration of Hospitalization:

6

Number

☒ Hours

☐ Days

HODUR

SDTM

机器可读:
Dates/Times: ISO
8601, 1 variable,
YYYY-MM-
DDThh:mm:ss
Duration: P1M3D

CDASH

人类可读:
Dates/Times: 2 or
more variables,
DD- MM-YYYY,
HH:MM:SS
Duration: 1 month,
3 days

原理

SDTM中一些变量使用了机器可读的格式，如日期类变量，对于数据的重用很方便，但是用于数据收集确不用户友好。人们在用不熟悉的格式记录数据时更容易犯错误。

数据的组织

SDTM

CDASH

数据必须按照域进行组织到数据集中

每个CRF页上的数据，是由一起收集什么来主导，而不一定非得按照域进行组织。

原理

Each CRF should have the data that makes sense to collect together. Per CDASH, this can mix domains if standard variable names are used, and in the end the data appear in the right domains.

In SDTM, data MUST appear only in the correct domain.

Demographics

Birth Date

1	3	D	E	C	1	9	4	8
Day		Month			Year			

BTHDTC

BTHDAT

Domain = DM

Sex

☐ Male

☐ Female

SEX

Race

☐ White

☐ Black

☐ Other, (please specify) _____

RACE

RACE1

RACE2

RACEOTH

Reproductive Status (if female)

☒ Premenarchal

☐ Postmenarchal

☐ Postmenopausal

Domain = RP

RPORRES when
RPTESTCD = MENOSTAT
or (CDASH)
MENOSTAT

DM CDASH

RP CDASH

DM SDTM

USUBJID	SUBJID	BTHDTC	SEX	RACE
ABC1201	01	1948-12-13	M	WHITE
ABC1202	02	1955-03-22	M	BLACK

RP SDTM

USUBJID	RPTESTCD	RPORRES	RPSTRESC
ABC1201	MENOSTAT	Premenarchal	Premenarchal
ABC1201	NUMLIV	2	2

EDC CRF

Hamilton Depression Scale 21 Items (HAM-D21)

1. DEPRESSED MOOD (<i>Sadness, hopeless, help-</i>) ☐ Absent ☐ These feeling states indicated only on questioning ☐ These feeling states spontaneously reported verbally	4. INSOMNIA EARLY ☐ No difficulty falling asleep ☐ Complains of occasionally difficulty falling asleep ☐ Complains of nightly difficulty falling asleep
3	5. INSOMNIA MIDDLE ☐ No difficulty ☐ Complains of being restless and disturbed ☐ Waking during night - any getting out of bed rates 2

QSORRES when QSTESTCD = HAMD101 or (CDASH) HAMD101

HAMD101

HAMD104

HAMD105

CDASH

SUBJID	HAMD101	HAMD104	HAMD105
1	Absent	No difficulty falling asleep	No difficulty

SDTM

USUBJ ID	QSTESTCD	QSTEST	QSORRES	QSSTR ESC
2324-P0001	HAMD101	HAMD1-Depressed Mood	Absent	0
2324-P0001	HAMD104	HAMD1-Insomnia Early - Early Night	No difficulty falling asleep	0

水平结构vs竖直结构

SDTM

发现类的数据必须是规范化的格式或者竖直结构；答案是已经知晓。

CDASH

发现类的数据可以是水平横置的，以便于每个测试有自己的值列表；SDTM CT可用于变量的命名和CRF的提示

原理

In normalized data, each test is on a different record and may need different controlled terms, called “value-level metadata,” which cannot be modeled in most data capture systems; some EDC systems can’t handle normalized data at all

元数据内容

CDASH Metadata

Data Collection Scenario	CDASHIG Variable	Question Text	Prompt Type	CRF Completion Instructions
--------------------------	------------------	---------------	-------------	-----------------------------

SDTM Metadata

Variable Name	Variable Label	Type	Controlled Terms Codelist or Format	Role	CDISC Notes
---------------	----------------	------	-------------------------------------	------	-------------

SDTM

元数据包含变量标签及角色

CDASH

元数据包含问题的描述或提示，CRF填写指南

原理

SDTM关注于将数据以表格形式展现。CDASH主要解决数据收集的需求，设计时要求确保问题清晰明确，能得到一致的答案。他们的目的不同，因而他们的元数据不同。

无法使用的变量

Exposure as Collected EC CRF

Exposure

EC

Line #	Dose	Dose Frequency	Dosing Time Period
1	ECDSTXT Capsules ECDOSU	☐ BID ☐ TID ECDOSFRQ	Start Date ECSTDAT Day Month Year End Date ECENDAT Day Month Year

SDTM

某些变量需等到研究结束后才可以衍生出来，
例如：USUBJID，
STRESC/STRESN，
EXDOSE

CDASH

某些变量仅在研究的情境下才有意义，例
如：SUBJID，
ECDOSE

原理

USUBJID在一次递交中必须是唯一的；
STRESC和STRESN在一次递交中将发现类的结果进行了标化；在双盲研究里，直到最后揭盲时，才能知道治疗组的剂量信息。
SDTM主要关注将数据合并汇总并用于递交；CDASH主要关注在单个研究里数据收集。

USUBJID	EXTRT	EXDOSE	EXDOSU	EXDOSFRQ	EXSTDTC	EXENDTC
ABC1201	BestDrug	350	mg	BID	2011-08-06	2011-09-01

两者一起使用的好处

确保相同的问题，相同的答案，和分析时所用一致

优化研究中心的数据收集及其结构，便于传输以及分析

支持数据溯源到采集点

最小化编程和验证所需资源，增加从数据采集转换到展现时的质量

针对前面所述的问题，皆可解决

可将数据存入数据仓库以备未来之需

寻求疾病疗法 *Unlocking Cures*

