NIDA CTN Data Share: A Case Study

Paul VanVeldhuisen, Ph.D.
The EMMES Corporation

May 20, 2013

The NIDA CTN

Mission: To improve the quality of drug abuse treatment throughout the country using science as the vehicle

Trial Characteristics within CTN

Similarities

- Randomized
- Multisite
- Some data collection instruments

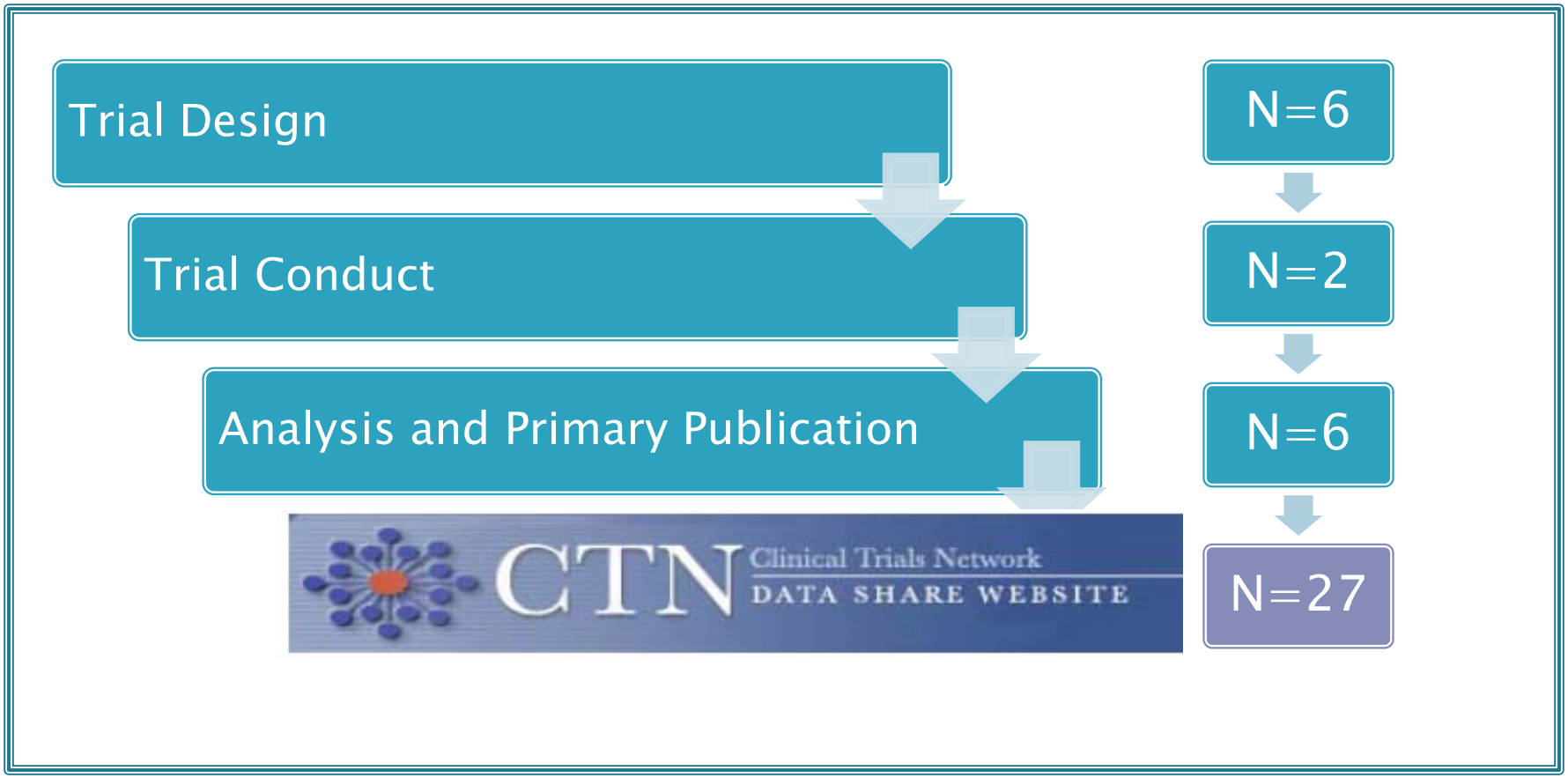
Differences

- Intervention types
  - Pharmacotherapies
  - Behavioral
  - Both
- Substance of use/abuse
- Populations
- Study assessments
- Data collection platforms

CTN Data Share Goals

- Provide participant-level data to research community
- Databases
  - easily accessible
  - released in timely manner
  - free
- Network setting -> promote cross-study analysis
  - standardize data as much as possible
- Optimize investment of original trials

Current CTN Status



NIDA CTN Protocol Example

NIDA-CTN-0010

Study Title: Buprenorphine/Naloxone-Facilitated Rehabilitation for Opioid Dependent Adolescents/Young Adults

Short Description: To compare two 3-month treatments for adolescents/young adults who are addicted to heroin. Both groups get 3 months of psychosocial treatment (individual and/or group drug counseling). The experimental group receives 3 months of buprenorphine stabilization, while the control group receives buprenorphine over a 7 - 14 day period.

30 raw datasets

Data Dictionary

NIDA-CTN-0010 Demographics Form 87

Field Name	Field Label	Field Description	Data Type	Length	Key Field	Null	Valid Responses	Valid Responses Codes	Logical/ Cross Reference Checks
PROTNUM	Protocol Number	Protocol Number (NIDA-CTN-0010)	Text	13	TRUE	Not Null			
NODE	Node Number	Participating Node Number	Integer	2	TRUE	Not Null			
SITE	Site Number	Site ID within a Node (Pri-Sub Site)	Integer	4	TRUE	Not Null			
ID	Subject ID	Subject ID	Integer	4	TRUE	Not Null			
VISDATE	Visit Date	Date of Assessment	Date	10	FALSE	Not Null			
PHASE	Visit Phase	Phase	Integer	1	FALSE	Not Null	0;	Screening;	
VISWEEK	Visit Week	Visit Week	Integer	2	FALSE	Not Null	-8;	N/A	VISWEEK eq sysmis. (Sysmis recode on backend to -8 for IMC purposes)
SEQNUM	Sequence Number	Sequence Number	Integer	2	FALSE	Null			Only used on AE and SAE
DEM001	Sex	Sex	Integer	1	FALSE	Not Null	1; 2; -9;	Male; Female; (X) Subject refused to answer;	
DEM002	Date of Birth	Date of Birth	Date	10	FALSE	Not Null			
DEM00301	White	Ethnicity/Race - White	Integer	1	FALSE	Not Null	0; 1;	No; Yes;	
DEM00302	Black	Ethnicity/Race - Black, African American	Integer	1	FALSE	Not Null	0; 1;	No; Yes;	

Case Report Forms

Study STUDYID DOMAIN: DM, SC ID USUBJID

Date __ __/__ __/_____ DMDTC/SC

Demographics Form

1. Sex

1 = Male DM.SEX

2 = Female

DM.BRTHDTC

2. Date of Birth __ __/__ __/_ _ _ _

3. Ethnicity/Race

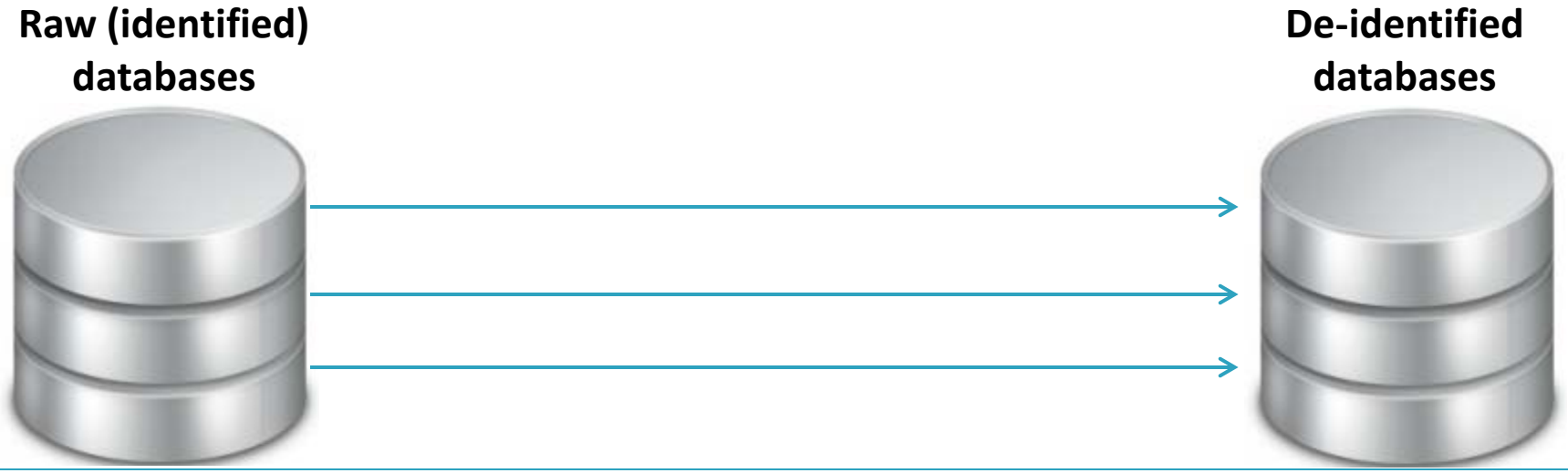
For each of the following, choose **"Yes"** to all that apply to you and **"No"** to those that do not.

DM.RACE (if multiple, then 'MULTIPLE')

SCORRES Yes	No	SCTEST
1	0	White
1	0	Black , African-American
1	0	American Indian or Alaskan Native
1	0	Spanish, Hispanic, or Latino (mark all that apply)

Case Report Forms

De-identification Process



- Original raw database structure maintained
- Identifiers according to HIPAA (18 variables) **emptied** (not typically collected)
- Long text fields (e.g. Other-specify), comment boxes, SAE narratives **emptied**
- Dates
  - Convert to "**days on study**" where Day 0 is date of randomization
  - **Age at randomization** replaces date of birth
- Participant ID (which includes site) **recoded** by unique random identifying number
- Site identifying number **dropped** (not recoded)

SDTM vs CRF-level

- Case Report Form (CRF) format
  - Preserves data structure exactly as originally collected in protocol
  - Easier for secondary analysis within single protocol
- Study Data Tabulation Model (SDTM) format developed by Clinical Data Interchange Standards Consortium (CDISC):
  - CDISC standards are widely accepted by pharmaceutical industry, but not familiar to academicians
  - FDA requires sponsors to submit clinical safety data in SDTM format alongside analysis data format
  - Labor intensive: need certified professionals to do mapping to SDTM: **Retrospective Standardization**
  - Data structures move away from raw data structures
  - Facilitates data merge for cross-study analysis

CRF-Level vs SDTM Example

CRF Level

Study 0001 Database: dem001	
ID	DEM01
01_037378	1
01_002844	2

Study 0019 Database: dem	
SUBJNO	DEM001
19_000117	2
19_000197	2

SDTM

Study 0001 Database: DM	
USUBJID	SEX
01_037378	M
01_002844	F

Study 0019 Database: DM	
USUBJID	SEX
19_000117	F
19_000197	F

What De-Identified Data to Share?

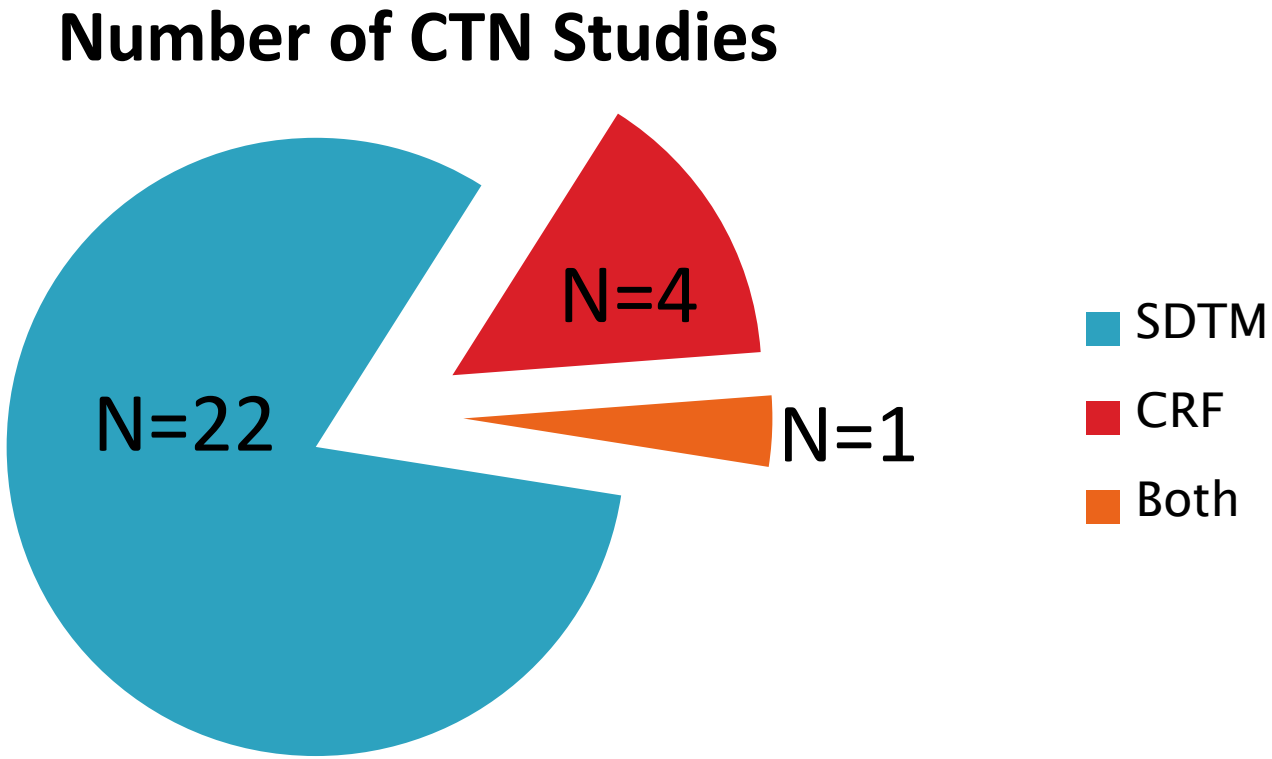
?

CRF-level databases

?

CDISC/SDTM databases

CRF-Level vs SDTM in CTN Data Share



User Expertise Expectations

- Data management
  - Ability to merge database
- Statistical
  - Access to statistical software
    - SAS XPORT files (can import into SPSS)
    - Software that can import ASCII data
    - May require data conversion software for use in other types of statistical software

Requires an investment of time and effort to get meaningful results

Assessments Link on Website

- Detailed information on a trial level
- Assessments may be searched either by:
  - Assessment category (e.g., mental health) and subcategory (e.g. depression)
  - Protocol number (e.g. CTN 0001)

Assessments Example

Home » Assessments Main Page » Assessments Main Page

Assessments Main Page

Home | By Assessment | By Protocol

Tobacco:

- ☐ CTP Smoking Survey
- ☑ Fagerstrom Test for Nicotine Dependence
- ☐ Reasons for Quitting
- ☐ Smoking Cessation Medication
- ☐ Tobacco Use Assessment
- ☐ Withdrawal Scale for Tobacco

Protocol	FTND
NIDA-CTN-0009	✔
NIDA-CTN-0029	✔
NIDA-CTN-0030	✔
NIDA-CTN-0031	✔
NIDA-CTN-0027	✔

Helping End Users

- "Contact Us" link on Website
- Technical questions
  - Handled by Data Coordinating Center
  - Examples: trouble accessing, downloading, data formats
- Protocol-specific research questions
  - Reach out to lead investigators
  - Example: how missing data were handled related to scoring instruments
- General research questions
  - Handled by Data Coordinating Center
  - Example: Do you know where we can get data on assessment X?

Limitations

- Minimal control over requests and usage
  - Encourage public use vs lack of control over how data used
- CRF-level vs SDTM
- Impact of de-identification
  - Information loss
  - Inability to replicate primary findings
- Statistical and database manipulation expertise
  - Post “analysis” datasets?

Future Directions

- Prospective standardization of instruments and CRFs
  - Improve ability to do cross-study analysis
- Extend Data Share website to include NIDA trials outside the CTN


Thank you!