Data Management

Responsible Conduct of Research (RCR)

Chalinee Monsereenusorn, M.D.

Assistant Professor

Pediatric Oncology Attending

Division of Hematology-Oncology, Department of Pediatrics

Phramongkutklao Hospital, Bangkok, THAILAND

ORDPMK
Office of Research Development
Phramongkutklao Hospital & Phramongkutklao College of Medicine

Objectives

- Describe step of data management during the study
- Learn what's the "Data Quality"
- Types of database
- How to develop data sheet or case record form to answer your research question to obtain analyzable data

Sections in Proposal

- Introduction
- Rationale and background
- Objectives or outcomes
- Benefit of the research
- Study design
- Population
- Recruitment : Inclusion and Exclusion criteria
- Methodology
- Statistical analysis
- Literature reviews
- Appendix: Case record form (CRF), inform and consent

Data Management Steps

Finalize Protocol to Identify Data Variables

For Each Variable, Ask…

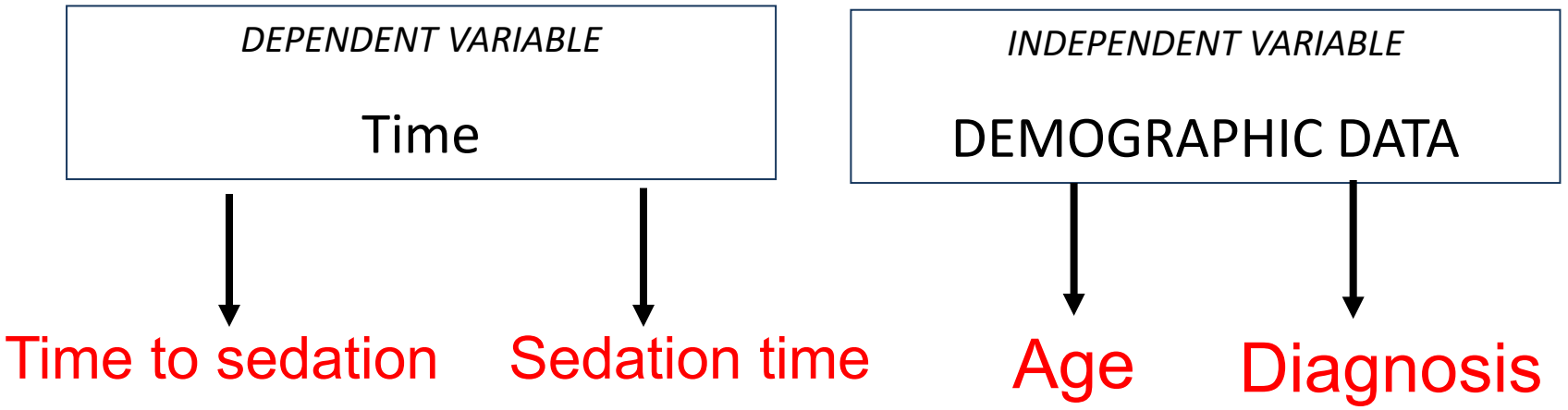
- Where will be data be collected from (source)?
- How often will data be collected?
- When will measurement be conducted (timing)?
- Who will complete the study measurements?

Example

In a randomized control trial evaluating The Efficacy of Midazolam with Fentanyl versus Midazolam with Ketamine for Bedside Invasive Procedural Sedation in Pediatric Oncology Patients, as the primary outcome measure.

What variables are needed to measure the efficacy of sedation drugs??

ESSENTIAL VARIABLE: EFFICACY



Midazolam with Fentanyl

vs.

Midazolam with Ketamine

Data Variables

Data set_IM Demo.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
4	Family	Numeric	8	0		{0, no}...	None	8	Right	Nominal
5	Familyposi	String	750	0	Detail of Famil	None	None	12	Left	Nominal
6	Type	Numeric	8	0		{0, solitary}...	None	8	Right	Nominal
7	Non_viscer	Numeric	8	0		{1, Head and	None	8	Right	Nominal
8	Boneinvol	Numeric	8	0		{0, no}...	None	8	Right	Nominal
9	Visceralinv	Numeric	8	0		{0, no}...	None	8	Right	Nominal
10	Visceralinv	Numeric	8	0		{1, GI tract}...	None	8	Right	Nominal
11	AgeDx	Numeric	8	2		None	None	8	Right	Scale
12	Followtime	Numeric	8	2		None	None	8	Right	Scale
13	Biopsysite	Numeric	8	0	Biopsy site	{1, Head and	None	8	Right	Nominal
14	Typebiopsy	Numeric	8	0	Type of biopsy	{0, incisional}.	None	8	Right	Nominal
15	Margin	Numeric	8	0	Margin	{0, negative}...	None	8	Right	Nominal
16	PATH	Numeric	8	0	PATH result	{0, inconclusiv	None	8	Right	Nominal
17	Immuno	Numeric	8	0	Immunohistoc	{0, no}...	None	8	Right	Nominal
18	ActinMSA	Numeric	8	0	Actin/MSA	{0, negative}...	None	8	Right	Nominal
19	Actin_HHF	Numeric	8	0	Actin-HHF35	{0, negative}...	None	8	Right	Nominal
20	ActinA4	Numeric	8	0	Actin A4	{0, negative}...	None	8	Right	Nominal
21	SMA	Numeric	8	0	SMA	{0, negative}...	None	8	Right	Nominal
22	Desmin	Numeric	8	0	Desmin	{0, negative}...	None	8	Right	Nominal
23	S_100	Numeric	8	0		{0, negative}...	None	8	Right	Nominal
24	vimentin	Numeric	8	0		{0, negative}...	None	8	Right	Nominal
25	FactorVIII	Numeric	8	0		{0, negative}...	None	8	Right	Nominal
26	CD31	Numeric	8	0		{0, negative}...	None	8	Right	Nominal
27	CD34	Numeric	8	0		{0, negative}...	None	8	Right	Nominal
28	NSE	Numeric	8	0		{0, negative}...	None	8	Right	Nominal
29	GFAP	Numeric	8	0		{0, negative}...	None	8	Right	Nominal
30	EMA	Numeric	8	0		{0, negative}...	None	8	Right	Nominal
31	Cam5.2	Numeric	8	0		{0, negative}...	None	8	Right	Nominal
32	CD68	Numeric	8	0		{0, negative}...	None	8	Right	Nominal
33	CD163	Numeric	8	0		{0, negative}...	None	8	Right	Nominal
34	CK	Numeric	8	0		{0, negative}...	None	8	Right	Nominal
35	Myogenin	Numeric	8	0		{0, negative}...	None	8	Right	Nominal
36	ALK_1	Numeric	8	0		{0, negative}...	None	8	Right	Nominal
37	Betacateni	Numeric	8	0		{0, negative}...	None	8	Right	Nominal
38	Cytogeneti	Numeric	8	0		{0, not done}..	None	8	Right	Nominal
39	Cytogeneti	Numeric	8	0		{0, normal}...	None	8	Right	Nominal
40	Diagnosticl	Numeric	8	0	Diagnostic Im	{1, Plain film}...	None	8	Right	Nominal
41	Surveillanc	Numeric	8	0	Surveillance Im	{1, Plain film}..	None	8	Right	Nominal
42	LNinvolve	Numeric	8	0	LN involvemen	{0, no}...	None	8	Right	Nominal
43	PATH_LN	Numeric	8	0	If LN +, PATH	{0, negative}...	None	8	Right	Nominal
44	findmass	Numeric	8	0	Imaging findin	{0, cystic}...	None	8	Right	Nominal
45	Descriptim	String	750	0	Descript imagi	None	None	8	Left	Nominal
46	regressioni	Numeric	8	0	Documented r	None	None	8	Right	Nominal
47										
48										
49										
50										
51										

Data View | Variable View

1. Numeric
   - *Categorical*
     0=Male
     1=Female
   - *Continuous* 1.875
2. Sting
   Text

Sources of Data

Where the data is first recorded is called a **source document**

- Medical record: paper, electronic Medical Record
- Lab report
- Survey/questionnaires

Additional Sources of Data

- Patients (e.g., Diary)
- External Data
  - Data collected from specimens or testing done and recorded via a separate database (e.g., pathology reports).
- Other

Data Management Steps

Finalize Protocol to Identify Data Variables

Create Case Report Forms *and Develop Database*

Case Record Form (CRF)

- A printed or electronic document designed to capture all study variables specified in the IRB approved protocol
  - The vehicle by which your study variables move from idea to reality, i.e. from the written protocol to an analyzable data set

Good CRF
Bullet points
Concise

Subject initials _____

V 3.0 30 มี.ค. 55

Subject ID# on each page

Case number [_][_][_]

Assessment date [_][_][_][_][_][_]

Visit / event date & data collector ID on each page

Visit 1

Visit date [_][_][_][_][_][_] ddmmyy

Weight _____kg Height _____cmBSA _____m^2

Type of Procedure ☐ Lumbar puncture ☐ Bone marrow examination

Type and dose of intrathecal chemotherapy

☐ Methotrexate [_][_] mg

☐ Hydrocortisone [_][_] mg

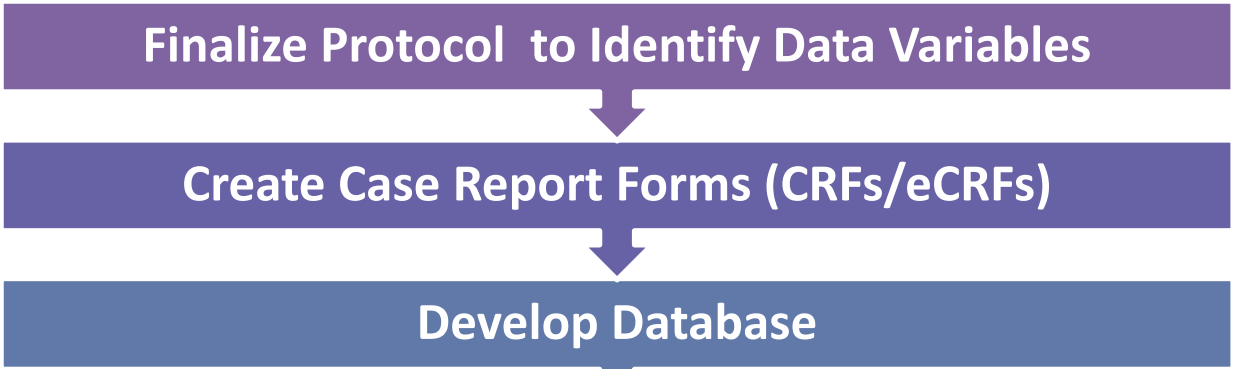
☐ Ara-C [_][_] mg

Avoid opened-handwriting

Study drug administered number _____

Randomization

Data Management Steps



Case Report Form to Database

CRF

SECTION A: GENERAL INFORMATION

A1. Subject ID number: ___ ___ - ___ ___ ___

A2. Subject initials: ___ ___ ___

A3. Date form completed: (MM/DD/YYYY) ___ ___ / ___ ___ / ___ ___ ___ ___

A4. Initials of person completing form: ___ ___ ___

Instructions: Information on this form is to be collected by interview with the mother in person or over the phone and from the infant's medical chart. Assign study number once patient is deemed eligible.

SECTION B: INFANT CHART REVIEW:

B1. Was the infant born less than or equal to 32 6/7 weeks of gestation?

Yes ..... O

No ..... O

B2. Was the infant less than or equal to 1500grams at birth?

Yes ..... O

No ..... O **(IF "YES" to B1, continue, if "NO" to B1 and B2 NOT ELIGIBLE)**

B3. Has the infant had a positive CMV test (PCR, culture or antigenemia)?

Yes..... O **(GO TO B3a)**

No..... O

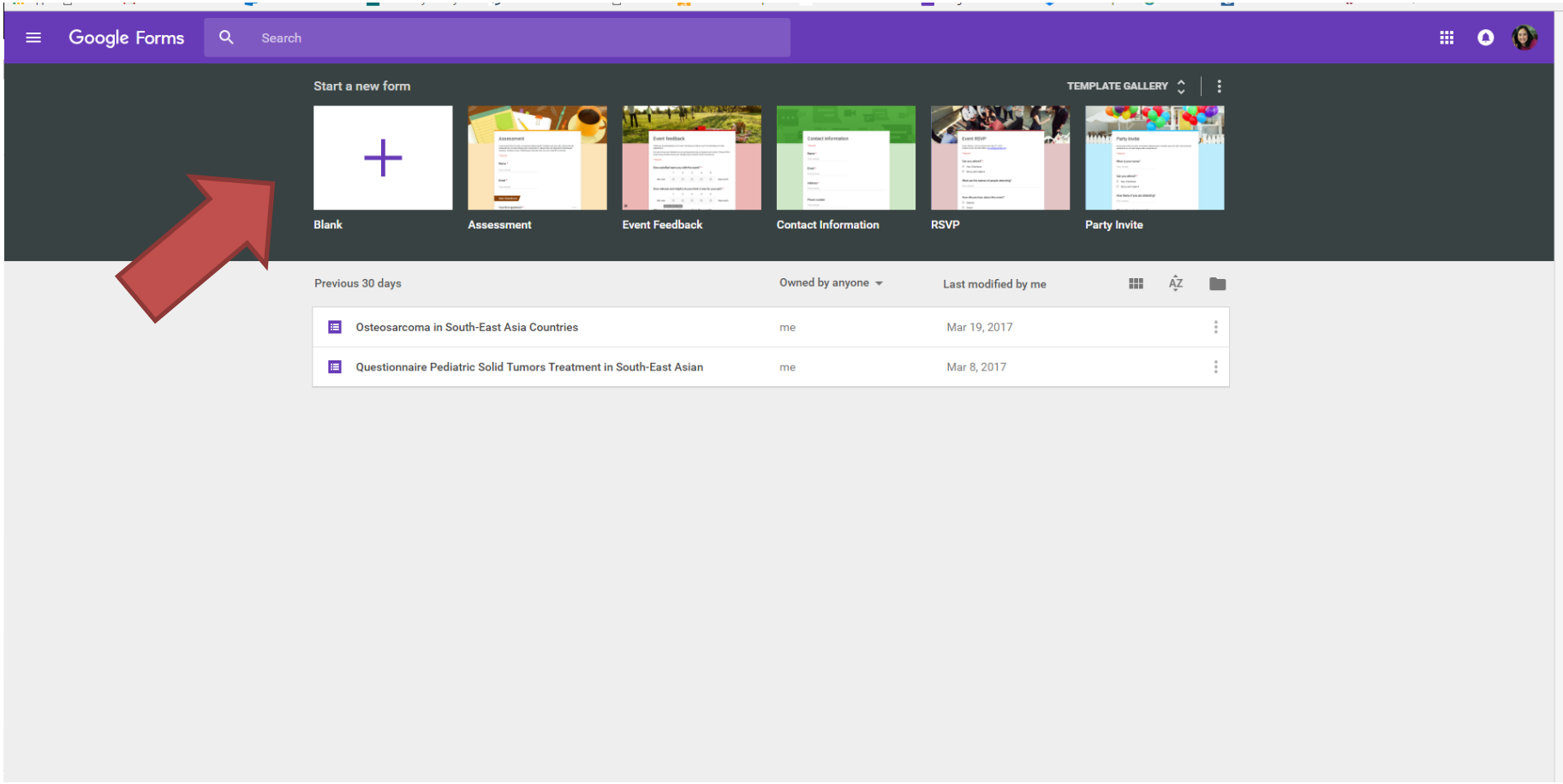
Database Excel

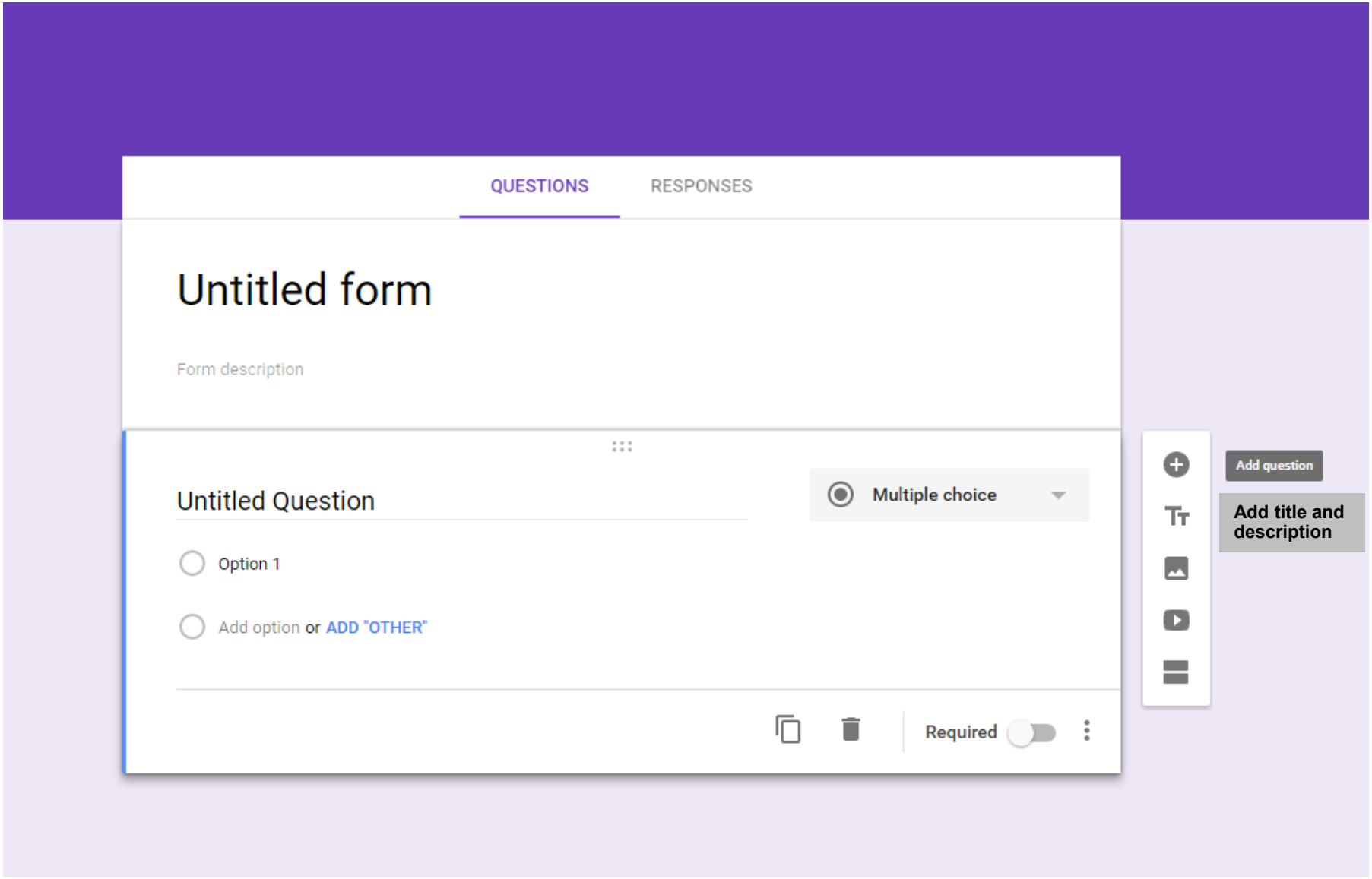
Event Name: **Screening/Baseline Labs (Arm 1: CMV Negative Mothers)**

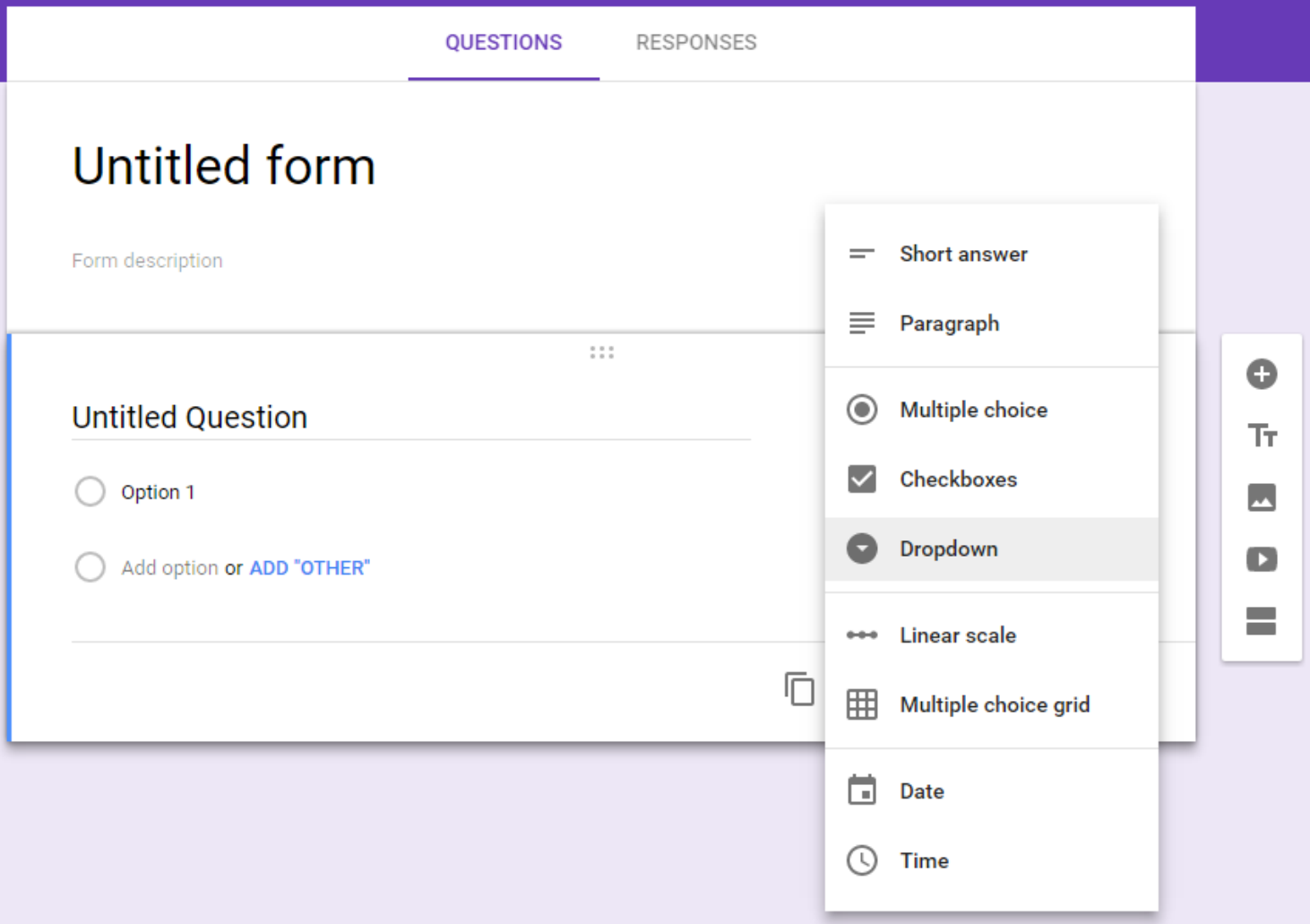
A1. Study ID number	01025 (To rename this record, modify the value immediately below)
A1. Study ID number	01025
SECTION A: General Information	
A2. Subject Initials	MK
A3. Date of screening	2009-10-15 MM-DD-YYYY
A4. Intials of person completeing form	AKC
SECTION B: Infant Chart Review	
B1. Infant born < 32 6/7 weeks?	(•) Yes ( ) No reset value
B2. Infant less than 1500g?	(•) Yes ( ) No reset value
B3. Has infant had a positive CMV test?	( ) Yes (•) No reset value

Recommend: similar in layout and appearance to help avoid error….

Google forms


Google Forms
Search
Start a new form
TEMPLATE GALLERY
Blank
Assessment
Event Feedback
Contact Information
RSVP
Party Invite
Previous 30 days
Owned by anyone
Last modified by me
Osteosarcoma in South-East Asia Countries
me
Mar 19, 2017
Questionnaire Pediatric Solid Tumors Treatment in South-East Asian
me
Mar 8, 2017


QUESTIONS
RESPONSES
Untitled form
Form description
Untitled Question
Multiple choice
Option 1
Add option or ADD "OTHER"
Required
Add question
Add title and description

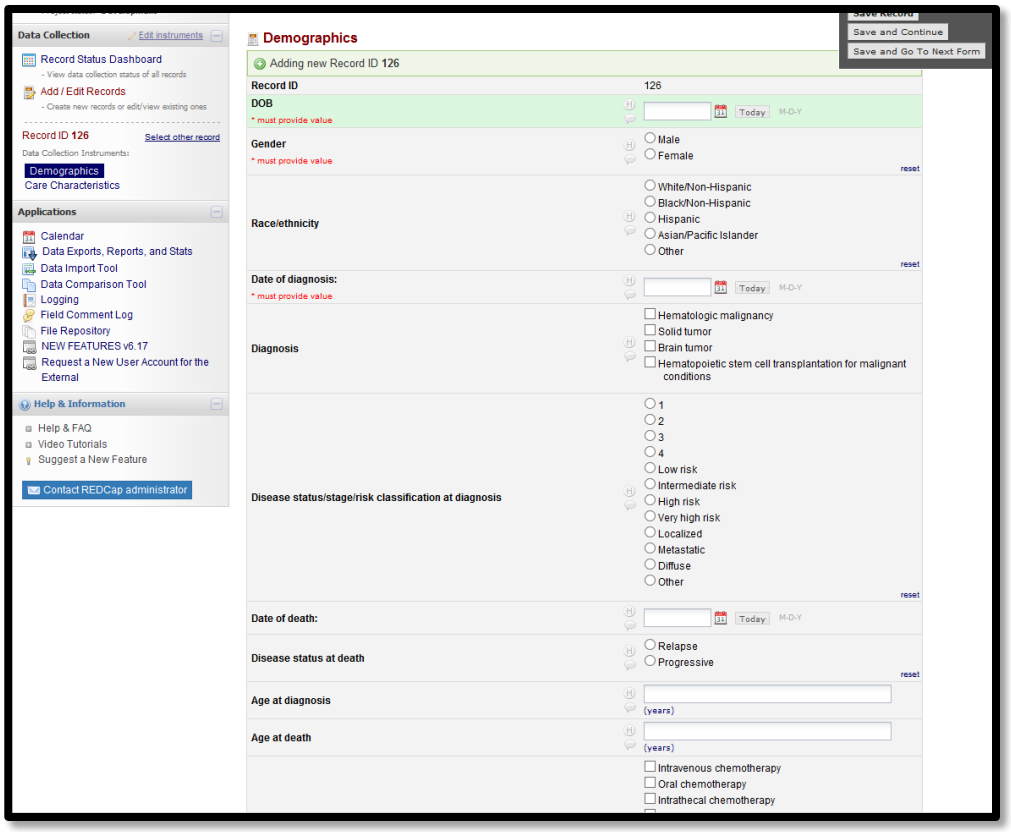

QUESTIONS
RESPONSES
Untitled form
Form description
Untitled Question
Option 1
Add option or ADD "OTHER"
Short answer
Paragraph
Multiple choice
Checkboxes
Dropdown
Linear scale
Multiple choice grid
Date
Time

CRF and database in the same place!!!

What's REDCap

- REDCap (Research Electronic Data Capture)
- Electronic databases
- User-friendly electronic data capture (EDC) tools for research studies
- Supporting data capture for clinical and translational research studies
- Free, secure, web-based application designed
- Intranet or internet access
- Online vs. offline

Database to Case Record Form



Demographics

Record ID _____

DOB _____

Gender
- ○ Male
- ○ Female

Race/ethnicity
- ○ White/Non-Hispanic
- ○ Black/Non-Hispanic
- ○ Hispanic
- ○ Asian/Pacific Islander
- ○ Other

If other, please specify: _____

Date of diagnosis: _____

Diagnosis
- ☐ Hematologic malignancy
- ☐ Solid tumor
- ☐ Brain tumor
- ☐ Hematopoietic stem cell transplantation for malignant conditions

Hematologic malignancy
- ☐ Acute lymphoblastic leukemia (ALL)
- ☐ Acute myeloid leukemia (AML)
- ☐ Chronic myeloid leukemia (CML)
- ☐ Juvenile myelomonocytic leukemia (JMML)
- ☐ Hodgkin Lymphoma
- ☐ Acute lymphoblastic lymphoma
- ☐ Other non-Hodgkin lymphoma
- ☐ Other

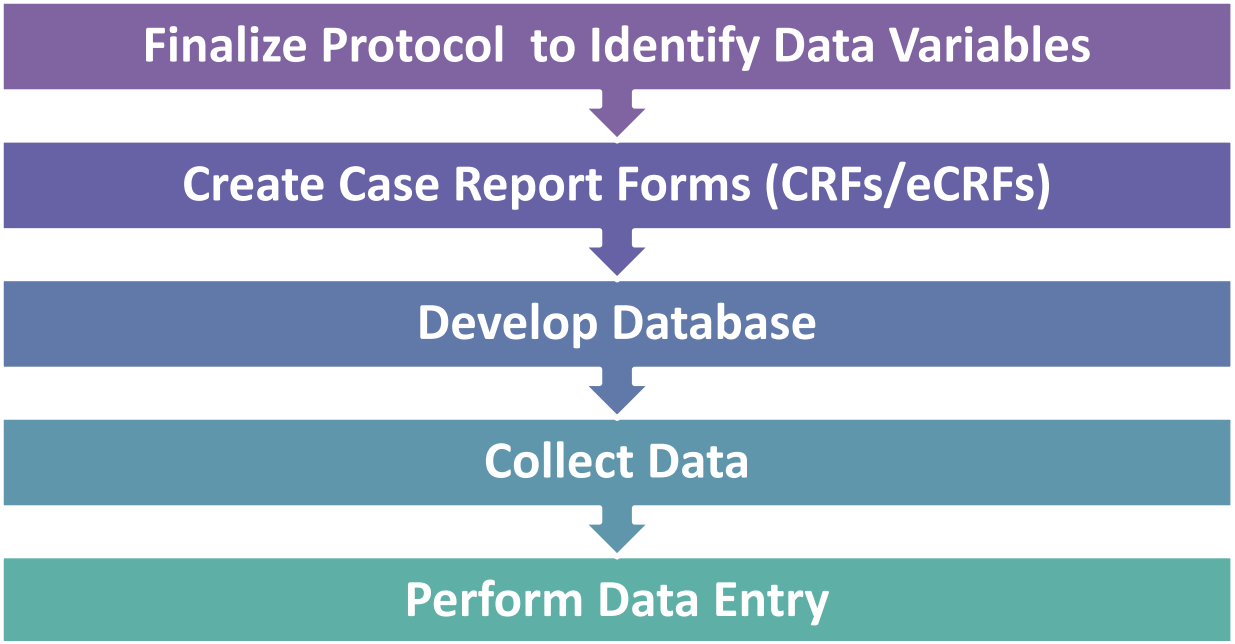
If Other heme malignancy, please specify: _____

Solid tumor
- ☐ Osteosarcoma
- ☐ Ewing sarcoma
- ☐ Hepatoblastoma
- ☐ Hepatocellular carcinoma
- ☐ Germ cell tumor
- ☐ Neuroblastoma
- ☐ Rhabdomyosarcoma
- ☐ Other

If other solid tumor, please specify: _____

Brain tumor
- ☐ Diffuse intrinsic pontine glioma (DIPG)
- ☐ PNET
- ☐ Glioblastoma multiforme (GBM)
- ☐ High grade glioma
- ☐ Medulloblastoma
- ☐ CNS germ cell tumor

Data Management Steps



Data Entry Best Practices

- Only enter what you see on CRF
- Fill out all “blank”
- Detailed important notification
- Do not enter anything illegible/unclear/open to interpretation
  - Unless defined as ‘Self Evident Correction’
- Flag using ‘Sticky Notes’ or ‘Remark’
- Many CRFs to fill -> summarize how to fill them out

ตารางแสดงจำนวนแบบสอบถามในแต่ละชนิดของผู้ป่วยมะเร็งตามช่วงอายุต่างๆ

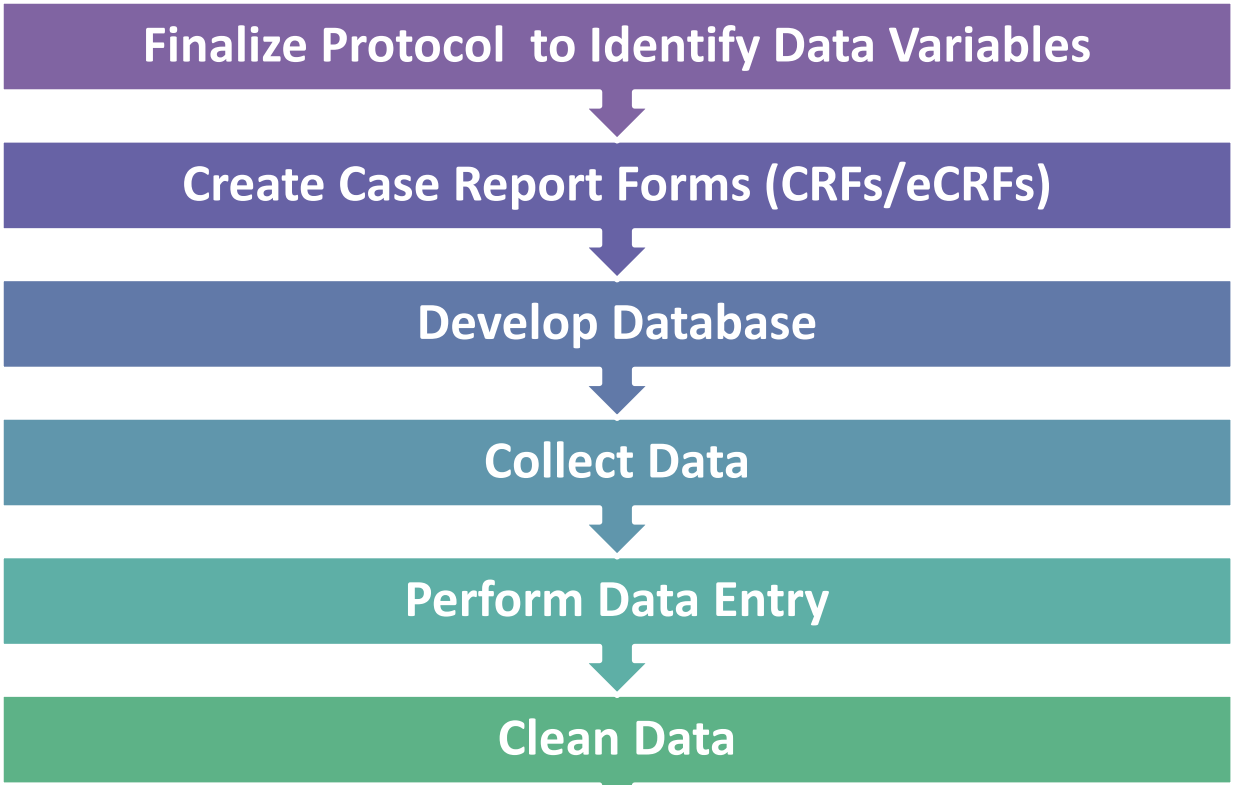
ชนิดมะเร็ง	กลุ่มอายุ (ปี)	CRF	Generic		Cancer module				Brain module	
					ผู้ป่วย (ตัวเด็ก)		บิดา /มารดา			
			ผู้ป่วย (ตัวเด็ก)	บิดา / มารดา	7 day (acute)	1 month	7 day (acute)	1 month	ผู้ป่วย (ตัวเด็ก)	บิดา/มารดา
มะเร็งธรรมดา (Non-brain tumor)	2-4	/		/			/	/		
	5-7	/	/	/	/	/	/	/		
	8-12	/	/	/	/	/	/	/		
	13-18	/	/	/	/	/	/	/		
	18-25	/			/	/	/	/		
มะเร็งสมอง (Brain tumor)	2-4	/		/						/
	5-7	/	/	/					/	/
	8-12	/	/	/					/	/
	13-25*	/	/	/					/	/

*หมายเหตุ กลุ่มผู้ป่วยมะเร็งสมอง (Brain tumor) อายุ 18-25 ปี ให้ใช้แบบสอบถามฉบับเดียวกันกับผู้ป่วยอายุ 13-18 ปี

Data Entry Best Practices

- Initial and date CRF to indicate data entered
- For complex or multi-center studies, universalized CRF or used uniformed database. i.e., REDCap or locked Excel sheets
- Need organizational system
  - CRFs Ready for Data Entry
  - CRFs Flagged (Incomplete, Incorrect, Need Resolution)
  - CRFs Entered (already input data into database)

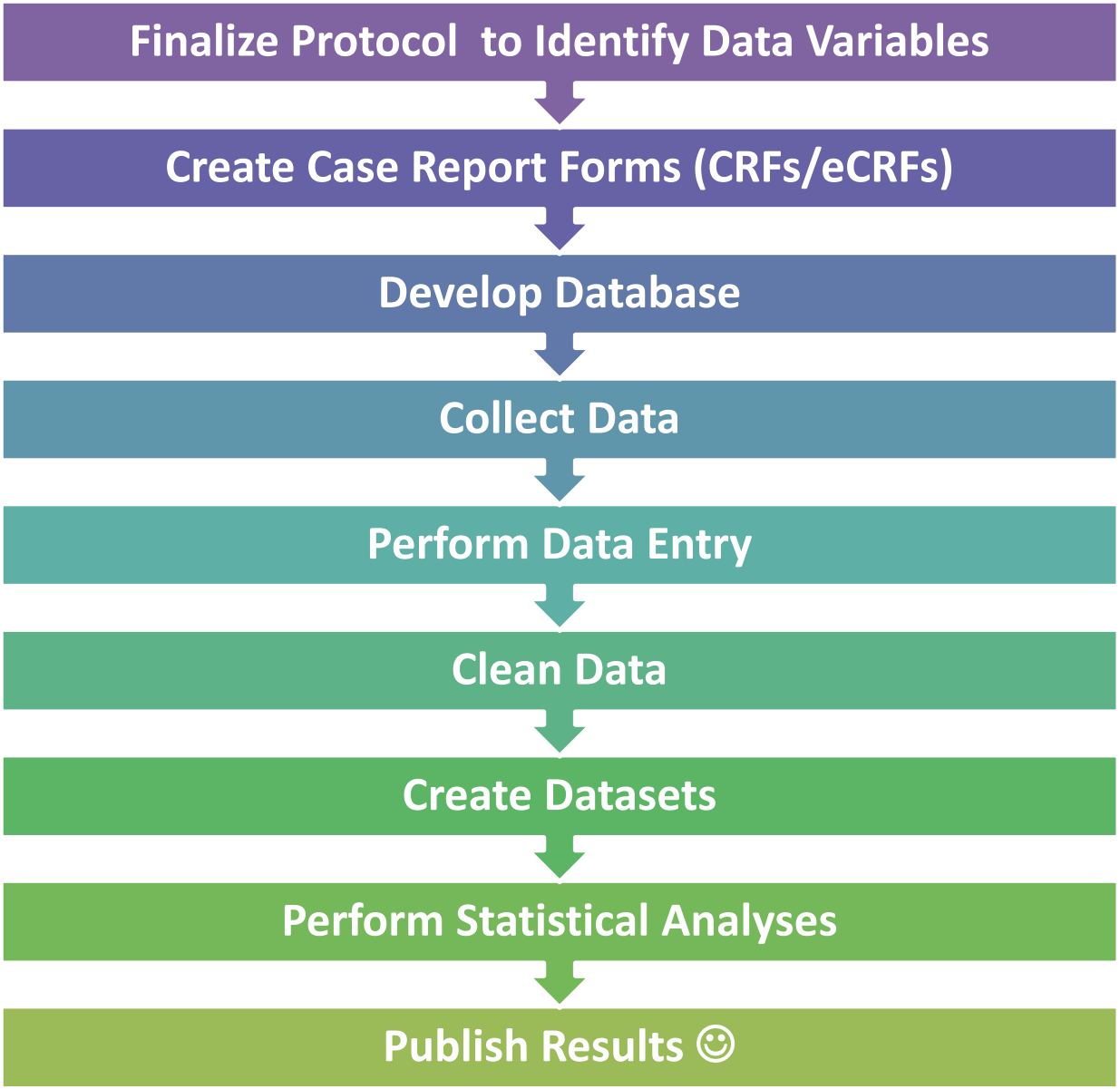
Data Management Steps



Typical Data Cleaning Tasks

- Dates – most common error type!
- Free text – avoid, transfer to numerical variables
- Confirm data is logical-date of death vs. date of birth
- Missing values where data is required
- Verify accurate data entry – compare to source
- Numeric data – check allowable ranges
- Identify duplicate entries
  - Check uniqueness of data values (study ID)
  - Ensure each patient is assigned a unique study number!

Data Management Steps



Common Mistakes

- CRF MUST be done before submit the proposal to IRB
- CRF is be able to edit rely on IRB recommendation
- Once IRB approval, any edited CRF MUST be processed as amendment

Conclusion:
Plan Ahead Before Implementing Your Study

- Develop definitive list of essential study variables
  - But collect only the data that are needed!
- Develop Case Report Forms (CRF)
- Develop CRF Completion Guides (e.g., how to collect the data)
- Develop (Program) a database to match your study's needs
  - Paper, REDCap, Google form
- Plan to perform targeted Data Cleaning
- Involve a Statistician – draft analytic plan

Key Concepts for Data Management

- Your research question will drive your study design.
- Choose your variables and database carefully. Involve a Statistician!
- Quality data doesn't come easy. Careful steps are required to succeed.
- Collect only the data needed and approved (in the IRB protocol).
- Inaccurate / incomplete data = inaccurate conclusions ☹

THANK YOU

chalinee_monsereenusorn@pedpmk.org