Developing the data collection sheet

Stephanie Stroever, PhD, MPH

Director, Division of Biostatistics, Clinical Research Institute

Assistant Professor, School of Medicine, Department of Medical Education

Importance of data collection sheet

- Defines and operationalizes your variables
- Provides structure for reliability and accuracy of data collection
- Assembles data in manner usable by biostatistician – saves you time

Must-haves

- Draft data collection tab
- Data dictionary tab

Data collection tab

- One column per variable
  - Do not list more than one data point per cell
  - If you have multiple possibilities, each needs its own column (see example)
- One row per subject
  - If you have repeated measures, keep on same row with a separate column for follow up

Example

id	group	Age	Gender	pulm_exacerbation_baseline	pulm_exacerbation_second	other_medhist_pancreatic_insufficiency	other_medhist_asthma	other_medhist_GERD	diabetes	organ_transplant_yn
1	1	6	1	4	5	1	0	0	0	
2	1	19	1	3	.	1	0	0	0	
3	1	15	0	7	8	1	1	0	0	
4	1	22	0	3	2	1	1	0	0	
5	1	28	0	8	8	1	0	1	0	
6	1	15	1	8	12	1	1	0	1	
7	1	25	0	3	6	1	1	0	0	
8	1	20	1	2	4	1	0	0	0	
9	1	10	1	5	5	1	1	1	0	
10	1	9	1	4	4	1	0	0	0	
11	0	30	0	0	0	0	0	0	0	
12	0	30	0	0	0	0	0	0	0	
13	1	24	1	2	1	1	0	0	0	
14	1	23	1	5	11	1	1	1	0	
15	0	26	0	0	0	0	0	0	0	
16	0	26	0	0	0	0	0	0	0	
17	1	21	0	3	4	0	1	0	1	
18	1	46	1	3	4	1	1	0	1	
19	1	30	0	2	3	1	0	0	1	

Example continued

	V	W	X	Y	Z	AA	
	FEV1_baseline	FEV1_baseline_ percentage	FEV1_second	FEV1_second_ percentage	BPI_IgG_baseline	BPI_IgG_second	Cl r
	0.89	55	1.59	98	3.46	2.97	
	1.30	44	.	.	3.83	.	
	2.62	73	3.70	95	7.37	4.49	
	3.61	88	3.27	71	5.38	2.97	
	3.25	65	3.47	71	3.82	2.93	
	1.31	46	1.82	61	4.74	2.76	
	3.95	89	3.76	85	4.70	2.79	
	3.12	89	3.12	90	2.16	4.48	
	2.04	97	2.26	97	4.39	2.71	
	1.33	85	1.55	84	2.16	4.48	
	.	.	.	.	4.39	.	
	.	.	.	.	3.68	.	
	1.85	41	3.67	81	72.07	17.38	
	2.44	77	2.56	81	32.48	20.93	
	.	.	.	.	2.51	.	
	.	.	.	.	5.39	.	
	2.15	70	2.48	78	2.74	3.63	
	0.82	29	1.25	45	37.75	44.54	
	3.14	77	2.93	72	5.36	4.55	

Example continued

	N	O	P	Q	R	S	T	U
	antibiotic_visit1	antibiotic_visit1_name1	antibiotic_visit1_name2	antibiotic_visit1_name3	antibiotic_visit2	antibiotic_visit2_name1	antibiotic_visit2_name2	antibiotic_visit2_name3 F
	1	1	14	0	1	1	14	0
	1	2	0	0	0	.	.	.
	0	.	.	.	1	7	8	0
	1	4	0	0	0	.	.	.
	1	5	0	0	1	7	0	0
	1	14	0	0	1	15	0	0
	1	2	0	0	1	2	0	0
	1	14	0	0	1	14	0	0
	1	6	14	0	1	14	0	0
	1	14	0	0	1	2	0	0
	0	.	.	.	0	.	.	.
	0	.	.	.	0	.	.	.
	1	8	14	0	1	8	0	0
	0	.	.	.	1	7	6	.
	0	.	.	.	0	.	.	.
	0	.	.	.	0	.	.	.
	1	14	0	0	1	16	0	0
	1	10	0	0	1	8	0	0
	1	2	8	0	1	2	0	0
	1	10	0	0	1	14	7	0
	1	9	0	0	1	8	7	0

Data dictionary tab

- This provides specific instructions for operationalizing your variables
  - Provides guidance to data collectors for precision and accuracy
- Gives the biostatistician information about coding of each variable
- Provides all investigators information for writing manuscript

Example

A	B	C	D
Variable Name	Variable Description	Type	Expected data
id	Assigned ID	Nominal	Value 1-XX
group	Group for the patient - either CF patient or healthy control	Dichotomous	0 = healthy control, 1 = CF patient
Age	age at the time of enrollment	Count	Value 1-XX
Gender	Gender identified in the medical record	Dichotomous	0 = male, 1 = female
pulm_exacerbation_baseline	# of pulmonary exacerbations (# oral or IV antiotics treatments per year)	Count	Value 1-XX
pulm_exacerbation_second	# of pulmonary exacerbations since enrollment in the study	Count	Value 1-XX
other_medhist_pancreatic_insufficiency	Patient has been diagnosed with pancreatic insufficiency *at any time* in their lives (ever/never)	Dichotomous	0 = no, 1 = yes

Example

antibiotic_visit1_name1	The type of antibiotic the patient is taking at the time of the first sample collection - #1	Categorical	0 = none, 1 = cefdinir, 2 = trimethoprim-sulfanomide, 3 = Minocycline, 4 = doxycycline, 5 = moxyfloxacin, 6 = aztreonam, 7 = tobramycin, 8 = azithromycin, 9 = vancomycin, 10 = ciprofloxacin, 11 - colistimethate, 12 = meropenem, 13 = ceftazidime, 14 = levofloxacin, 15 = Ceftaroline, 16 = erythromycin, 17 = cefepime, 18 = clindamycin
antibiotic_visit1_name2	The type of antibiotic the patient is taking at the time of the first sample collection - #2	Categorical	0 = none, 1 = cefdinir, 2 = trimethoprim-sulfanomide, 3 = Minocycline, 4 = doxycycline, 5 = moxyfloxacin, 6 = aztreonam, 7 = tobramycin, 8 = azithromycin, 9 = vancomycin, 10 = ciprofloxacin, 11 - colistimethate, 12 = meropenem, 13 = ceftazidime, 14 = levofloxacin, 15 = Ceftaroline, 16 = erythromycin, 17 = cefepime, 18 = clindamycin
antibiotic_visit1_name3	The type of antibiotic the patient is taking at the time of the first sample collection - #3	Categorical	0 = none, 1 = cefdinir, 2 = trimethoprim-sulfanomide, 3 = bactrim, 4 = doxycycline, 5 = moxyfloxacin, 6 = aztreonam, 7 = tobramycin, 8 = azithromycin, 9 = vancomycin, 10 = ciprofloxacin, 11 -

Common errors

○ Are you defining gender or sex? They are different!

○ Race and ethnicity

 ○ Race and ethnicity are not interchangeable

○ Continuous vs. categorical

 ○ Start continuous if possible – we can always categorize later after we assess the data

Ethnicity according to CMS

Ethnicity Data Standard *Are you Hispanic, Latino/a, or Spanish Origin? (One or more categories may be selected)*

Categories	Notes
a. _____ No, not of Hispanic, Latino/a, or Spanish origin	
b. _____ Yes, Mexican, Mexican American, Chicano/a	
c. _____ Yes, Puerto Rican	These categories roll-up to the Hispanic or Latino category of the OMB standard
d. _____ Yes, Cuban	
e. _____ Yes, Another Hispanic, Latino/a or Spanish origin	

Race according to CMS

Race Data Standard *What is your race?(One or more categories may be selected)*

Categories	Notes
a. _____ White	These categories are part of the current OMB standard
b. _____ Black or African American	
c. _____ American Indian or Alaska Native	
d. _____ Asian Indian	These categories roll-up to the Asian category of the OMB standard
e. _____ Chinese	
f. _____ Filipino	
g. _____ Japanese	
h. _____ Korean	
i. _____ Vietnamese	
j. _____ Other Asian	
k. _____ Native Hawaiian	These categories roll-up to the Native Hawaiian or Other Pacific Islander category of the OMB standard
l. _____ Guamanian or Chamorro	
m. _____ Samoan	

Questions?

Stephanie.Stroever@ttuhsc.edu