

Research Professionals Network Workshop Series

DEVELOPING EFFECTIVE DATA COLLECTION TOOLS

Alana Ewen, MPH
Data Analyst, Graduate Medical Education
Alana.Ewen@bmc.org

 Clinical & Translational Science Institute

 BOSTON
MEDICAL
CENTER
EXCEPTIONAL CARE. WITHOUT EXCEPTION.

 The University
of Vermont
WARREN COLLEGE OF MEDICINE

 THE
University of Vermont
MEDICAL CENTER

Objectives

Following the session, learners should be able to:

- 1) Define “data collection”
- 2) Outline the importance of having and maintaining a good data collection form
- 3) Determine what type of data to collect
- 4) Apply discussed concepts to develop one’s own data collection form

Data Collection - Definition

- The process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes.

- U.S. Department of Health & Human Services

Data Collection

- All facets of the data collection process should be carefully reviewed in order to ensure reliable and valid data
 - Make sure your data are collected in the same fashion from beginning to end of your research study

Key questions to ask yourself prior to data collection:

- 1) What information are you going to collect?
- 2) How are you going to collect your data?
 - a) Interviews/focus groups
 - b) Phone records
 - c) Surveys/questionnaires
 - d) Medical records (or clinic and office charts)
 - e) Recorded data from EEGs, ECGs, etc.
- 3) How will data be analyzed?

Protocol → Data Collection Process

WonderDrug Protocol 95-06

Site Number: 126
 Subject Screening Number: 126-001
 Initials: IT

Demographics
 Date of birth (Day/Month/Year): _____
 Gender: Male Female
 Race: White Black Hispanic/Latino Asian Other _____

Eligibility
 Does the subject meet all of the inclusion eligibility criteria? Yes No
 If no, specify the unmet inclusion criterion (by number): _____
 Does the subject meet any of the exclusion criteria? Yes No
 If yes, specify the met exclusion criterion (by number): _____
If the subject does not meet all of the eligibility criteria, do not proceed with randomization.

Randomization
 Was the subject randomized? Yes No
 If not, why not? Serious adverse event Screening lab test abnormality
 Withdrew consent Other reason

Evaluation of Primary Infection
 Date of enrollment: _____
 Duration of signs and symptoms prior to randomization (in days) ____
 Type of skin infection
 Decubitus ulcer Diabetic foot
 Major abscess Cellulitis (check only if complicated)
 Postoperative wound infection
 Wound classification stage if a pressure ulcer: ____

Cultures
 Were specimens sent for culture? Yes No
 Source of specimen
 Blood Pus Biopsy Curettage
 Aspirate Other
 Was this a deep culture? Yes No
If no, specimen is not acceptable; obtain another sample.
 Was a Gram stain done? Yes No
 Did the Gram stain show wbc*? Yes No
 Did the Gram stain show Gram positive cocci in clumps? Yes No
If no, patient must be dropped from the study.

Culture Results

Accession number	Date	Source	Site	Result (use organism code)

Sensitivity Results

	Organism 1	Organism 2, etc.
Bacteria isolated	MRSA	E. coli
Report sensitivities to the following antibiotics	MIC (ug/ml)	MIC (ug/ml)
Cephakillital		
Wondercillin		
Vancomycin		
Bactrim		
Rifampin		

* White blood cells, i.e., signs of infection

Importance of Collecting “Good”/Valid Data

- Precision of data is important to maintain the integrity of your research
- Written record (i.e. paper surveys) and electronic records should be properly maintained in the event that aspects of your research protocol need to be modified

Data Collection Approaches in Clinical Research

Advantages & Disadvantages

Questionnaire Survey Data

Advantages	Disadvantages
<ul style="list-style-type: none">▪ Can collect personal and/or risk factor data not typically contained in hospital/ambulatory care records	<ul style="list-style-type: none">• Validating individual survey responses can be difficult, burdensome, costly, and of questionable utility
<ul style="list-style-type: none">• Can elicit information in an analytically desirable and standardized manner	<ul style="list-style-type: none">• If response rates are less than desirable, one may question the representativeness of the study sample and its generalizability
<ul style="list-style-type: none">• Can maintain high survey response rates through various financial or other incentives	<ul style="list-style-type: none">• Responses might differ if questions are asked in-person vs. by phone vs. by mail/internet

Data Collection Approaches in Clinical Research

Advantages & Disadvantages

Hospital/Ambulatory Care Records

Advantages	Disadvantages
<ul style="list-style-type: none">• Readily available and contain much useful demographic and clinical information	<ul style="list-style-type: none">• Often times data contained in medical records are non-standardized and inconsistently collected and recorded
<ul style="list-style-type: none">• Can be linked to other follow-up information sources	<ul style="list-style-type: none">• Information is often incomplete and/or missing
<ul style="list-style-type: none">• Can be used to characterize the medical history and clinical course of hospitalized and outpatient individuals	<ul style="list-style-type: none">• Independent checks on validity and/or reliability are atypically performed
<ul style="list-style-type: none">• Can provide data on medication intensity and duration	<ul style="list-style-type: none">• Information on etiologic or prognostic factors of importance is often either not obtained or asked about or recorded in a standardized manner

Data Collection Approaches in Clinical Research

Advantages & Disadvantages

Biologic Data

Advantages	Disadvantages
<ul style="list-style-type: none">• May provide novel insights into underlying disease pathophysiologic processes	<ul style="list-style-type: none">• Need to be collected under standardized conditions with considerable attention to detail
<ul style="list-style-type: none">• Can serve as an important endpoint of relevance	<ul style="list-style-type: none">• Ongoing quality control procedures needed
<ul style="list-style-type: none">• Can be linked to other sociodemographic, medical history, and clinical data to obtain insights into disease occurrence and prognosis	<ul style="list-style-type: none">• Need to consider impact of possible biologic circadian variation for purposes of timing and frequency of data collection efforts
	<ul style="list-style-type: none">• May need collection of multiple measures at baseline to adequately profile subsequent changes

Have as many details as needed and define the variables in your form as needed

Sexually Transmitted Disease Program (ID) Client ID Number: ___/___/___
Clinic Patient Intake Form

(spa) SPA: 1 2 3 4 5 6 7 8 (Fictitious Data—DO NOT QUOTE)
(See STD Annual Report for SPA names:
<http://publichealth.lacounty.gov/phcommon/public/AtoZTopics/AtoZpubDisplayAll.cfm?alpha=S>)

Last Name: _____ First Name: _____

(gender) Gender: Male (M) Female (F) (dob) Date of Birth: ___/___/___

(race) Race/Ethnicity: Hispanic/Latino/Chicano (1)
 White (non-Hispanic) (2)
 African American/Black (3)
 Asian (4)
 Other/Mixed Race/Ethnicity (5)
 Refused (6)

(clinic) Skid Row Location: Hope (1)
 Olive (2)

(gsexpart) Gender of Sexual Partner: Male (1) Female (2) Both (3)

(csw) Commercial sex worker Yes (1) No (2)
(highrisk) Engages in high risk behaviors Yes (1) No (2)
(analsex) Engages in anal sex Yes (1) No (2)
(drugsex) Used drugs during sex Yes (1) No (2)
(internet) Gets sex partners from the internet Yes (1) No (2)
(oralsex) Gives/receives oral sex Yes (1) No (2)
(msexprt) Multiple sex partners Yes (1) No (2)

(datetest) Date Tested: ___/___/___

(chl) Chlamydia infection Yes (1) No (2)
(HIV) HIV+ Yes (1) No (2)
(gc) Gonorrhea infection Yes (1) No (2)
(trich) Trich Yes (1) No (2)

Counselor: _____ Location: _____ Date: ___/___/___

Version Control

<p>Specify details of DNA restrictions:</p> <hr/> <hr/> <hr/> <hr/>	
<p>ARIC PROTOCOL 2. Cohort Component Procedures Version 6.0 Visit 4, VERSION 4.0 July 1997</p>	

Current version and date should be on the first page. It is also recommended that it be on the header/footer of the document and appear on every page.

Improper Data Collection Leads to...

- Inability to answer research questions accurately
- Inability to repeat and validate the study
- Distorted findings resulting in wasted resources
- Misleading other researchers to pursue fruitless avenues of investigation
- Compromising decisions for public policy
- Causing harm to human participants and animal subjects

Data Collection Problems Requiring Immediate Action

- Errors in individual data items
- Systematic errors
- Violation of protocol
- Problems with individual staff or site performance
- Fraud or scientific misconduct



Clinical Research Documentation

"If it wasn't documented, it wasn't done."



Collecting Valid Data

Case Vignette: Collecting Valid Data



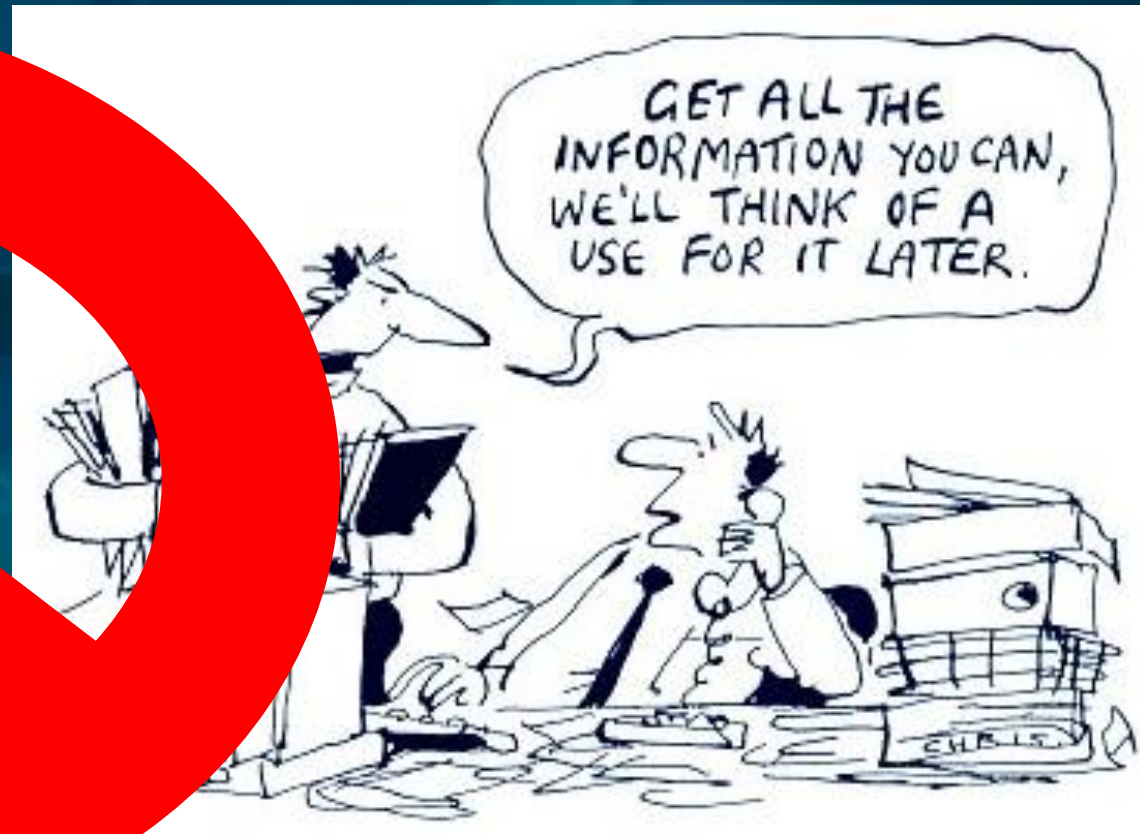
Part of the data collection methodology for Dr. Smith's study includes distributing a 12-page self-administered questionnaire to participants; they must fill out and initial each page of the questionnaire to confirm completion.

One day on his way home from conducting an interview with a subject, the Research Assistant, Joel, needed to write directions for a friend and he reached in his bag and grabbed the first piece of paper that he could find. Joel accidentally ripped the back page off of one of the completed questionnaires to write the directions, which he then gave to his friend. He didn't realize this until a few hours later, when he was reviewing the data that he had collected that day.

Joel thought that he remembered the participant's answers on the last page of the survey, since they were mostly demographic questions.

What should Joel do?

How Much is Too Much?



Data Collection Methods

What's Needed for...

Questionnaires

- 1) A “sampling frame” from which you can choose a representative (or randomized) sample
- 2) A survey instrument that has been vetted
- 3) Invitation to encourage participation
- 4) A means of dissemination (mail, email, telephone)
- 5) Method of following-up with non-respondents
- 6) A system for creating and managing a database of survey responses
- 7) A plan and method for analyzing results

...Interviews

- 1) A method to identify and recruit the people you want to interview
- 2) An interview protocol with primarily or exclusively open-ended questions
- 3) Skilled interviewers
- 4) A way to record the interviews and either summarize or transcribe them
- 5) A method to analyze the results of all your interviews. This may include a qualitative data analysis software program
- 6) Staff who have skills in qualitative data analysis

Communication/Training

Case Vignette: Communication



A few weeks after Dr. Smith added the new questions to the self-administered questionnaire, it occurred to the Research Assistant, Heather, that the data collection methodology could be changed slightly. She realized that the first questionnaire that was administered to subjects (a survey on attitudes) now included information that provided answers to the questions on a subsequent questionnaire (a knowledge pre-test).

Heather realized that it would make much more sense to administer the knowledge test **before** the attitude questionnaire.

How should Heather proceed?

Explicit Script in Data Collection Form

Interviews

Q1 Interviewer name:

Q2 REC ID:

Q3 Phone call date:

Q4 Phone call time:

Q5 Hello, may I please speak with XXX? [If the person is not available ask when a good time to call would be.] I am calling from Boston Medical Center. My name is XXX and I'm calling as a follow-up from the appointment you had on [appointment date] with Dr. XXX. You are being asked to voluntarily participate in a research study. We are studying how our patients are using resources in the hospital and the community to help with their basic needs like food, education, employment. You are someone that can help us understand this better. If you agree, we will ask you to share your opinion on some brief questions that will take about 10-15 minutes, we will not record any identifiable information, such as your name, on the survey itself. We will label your survey answers with a unique number, and this number will be linked to your identifiers in a separate document. Do you have a moment to share your opinion?

No [If no, ask if there is another time the patient might be available to share their opinion.] (1) _____

Yes (2)



Activity

Social Determinants of Health Pilot Study

- Objective #1:** The overall aim of the project is to examine the feasibility and potential impact of screening and referral for unmet SDOH needs on chronic disease clinical outcomes among new BMC adult primary care patients with diabetes, hypertension, and/or depression. We also aim to better characterize the SDOH burden among adult primary care patients.
- Question:** Based on the objective above, what variables (i.e. A1C, age, etc.) do you think would be of interest?









Screening and referral tool →

BMC Health System
THRIVE SCREENER
Addressing Social Determinants of Health










Place Patient Sticker Here

Please fill this out and give to the medical assistant when you are called into the exam room. Your answers will help your care team take better care of your health and connect you with resources. Thank you!

Please check "✓" your answers:

	What is your living situation today?	<input type="radio"/> I have a steady place to live <input type="radio"/> I have a place to live today, but I am worried about losing it in the future <input type="radio"/> I do not have a steady place to live (I am temporarily staying with others, in a hotel, in a shelter, living outside on the street, on a beach, in a car, abandoned building, bus or train station, or in a park)
	Within the past 12 months, the food you bought just didn't last and you didn't have money to get more.	<input type="radio"/> Often true <input type="radio"/> Sometimes true <input type="radio"/> Never true
	Within the past 12 months, you worried whether your food would run out before you got money to buy more.	<input type="radio"/> Often true <input type="radio"/> Sometimes true <input type="radio"/> Never true
	Is this an emergency, do you need food for tonight?	<input type="radio"/> Yes <input type="radio"/> No
	Do you have trouble paying for medicines?	<input type="radio"/> Yes <input type="radio"/> No
	Do you have trouble getting transportation to medical appointments?	<input type="radio"/> Yes <input type="radio"/> No
	Do you have trouble paying your heating or electricity bill?	<input type="radio"/> Yes <input type="radio"/> No
	Do you have trouble taking care of a child, family member or friend?	<input type="radio"/> Yes <input type="radio"/> No
	Are you currently unemployed and looking for a job?	<input type="radio"/> Yes <input type="radio"/> No
	Are you interested in more education?	<input type="radio"/> Yes <input type="radio"/> No

Please check "✓" the resources you want help with:

Housing / Shelter	Food	Paying for Medicine	Transportation to medical appointments	Utilities	Child care / Daycare	Care for elder or disabled	Job search/ training	Education
								
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

I do not want to answer these questions

© 2017 Boston Medical Center

Variables We Collected

Variables

City/State

Language

Marital Status

Provider Name

Education

Age, Sex

Insurance Carrier

Visit date/time

Lab date/time

Screening questions

Clinical: LDL, HDL, A1C, Total Cholesterol, Blood Pressure

Eligibility: has Depression? Diabetes? Hypertension?

...Objective #1 Continued

- **Objective #1:** The overall aim of the project is to examine the feasibility and potential impact of screening and referral for unmet Social Determinants of Health (SDOH) needs on chronic disease clinical outcomes among new BMC adult primary care patients with diabetes, hypertension, and/or depression. We also aim to better characterize the SDOH burden among adult primary care patients.
- **Questions:**
 - 1) What platform (i.e. Excel) would you consider using to collect/store your data?
 - 2) Where could you find the data needed to answer your research question?

Social Determinants of Health Pilot Study

- **Objective #2:** Using a screening and referral process to address unmet SDOH needs in a subgroup of patients with hypertension (HTN), diabetes (DM) or depression, we sought to examine patients' connection to community and hospital resources.

Question: Based on the objective above, what is a data collection method, or methods, that can be employed?

Three Key Reminders

- 1) Develop a data collection and training plan as needed
- 2) Follow your IRB protocol!
 - a) Study retention/closeout of data
- 3) Keep data secure
 - a) Protected Health Information (PHI) and stewardship of your data

Questions or comments?

