DATA MANAGEMENT PRINCIPLES

DEVERICK J. ANDERSON, MD, MPH, FSHEA, FIDSA

DIRECTOR, DUKE CENTER FOR ANTIMICROBIAL STEWARDSHIP AND INFECTION PREVENTION

SHEA ANTIMICROBIAL STEWARDSHIP RESEARCH WORKSHOP – NOVEMBER 2016

Duke University School of Medicine

dicon.medicine.duke.edu
dason.medicine.duke.edu

Duke Center for Antimicrobial Stewardship and Infection Prevention

Disclosures

Research grants – AHRQ, NIH/NIAID, CDC

Royalties – UpToDate

Disclosures

Duke Center for
Antimicrobial Stewardship
and Infection Prevention

Outline

Background

Quality control/Quality assurance (QC/QA)

- The details are important
- Throughout study execution
  - Planning
  - During
  - Post

Two points of emphasis for today

- Internal validity
- Data validation

Take home points

Background

Data is fundamental in epidemiological and stewardship research

- Cause and effect

Researchers faced with the inevitable question:

- DO I BELIEVE WHAT I SEE??

ANSWER: Depends on quality of data

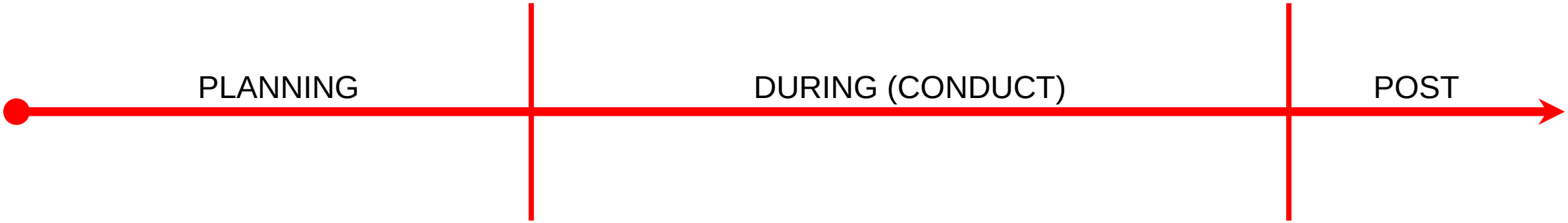
Data journey mirrors the study journey

Caveat: no such thing as a perfect (error- or bias-free) study

- Goal: minimize error and bias to greatest extent possible

QUALITY ASSURANCE

QUALITY CONTROL



DATA

Quality Control & Quality Assurance

Manufacturing

- QC – inspect products at the end of the manufacturing line and remove substandard products
- QA – improve all procedures to improve overall quality of the products
  - Focus on process not product

Research

- QA – practices to minimize systematic bias implemented **before** data collection
- QC – practices to minimize bias **during** and **after** data collection (correct mistakes identified)

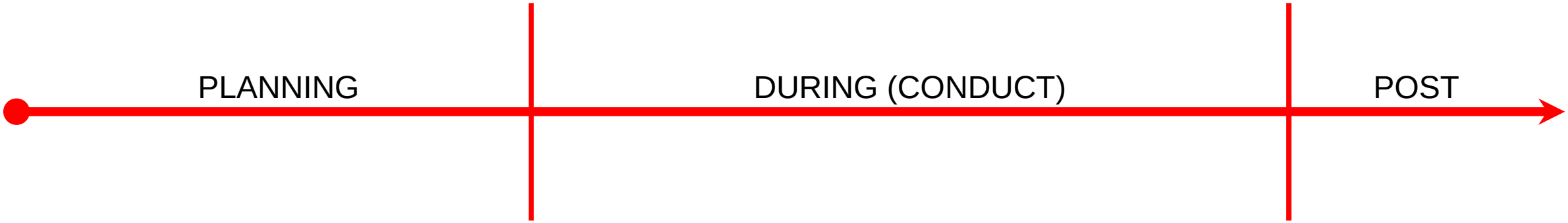
Data management is a part of QC/QA procedures

- Most literature related to clinical trials

QUALITY ASSURANCE

QUALITY CONTROL

Protocol development
Documentation
Personnel: training/certification
Ethics (IRB)



Data collection tools
Data validation planning
Data management planning
Pilot data collection?

DATA

Study Protocol

Outlines all the steps of the study process **before** the study begins

- QA/QC procedures
- Data management
- Statistical analysis plan

Time-consuming and burdensome

But worth it

- Use as the “map” for your journey

Study Protocol

Study objectives

Outcomes
- Primary
- Secondary

Study design

Populations (participants)
- Inclusion
- Exclusion

Variables

Data collection tool

Data collection strategy

Data validation steps

Statistical analysis plan
- Sample size and power

Internal Validity (vs. External Validity)

Internal validity – how well was the study performed?

- Study execution
- Steps to limit bias/confounding
  - Systematic bias

External validity – do results apply to other settings?

- Generalizability
- Repeat process (and get same results)

Schweizer et al. ICHE 2016; 37:1135.

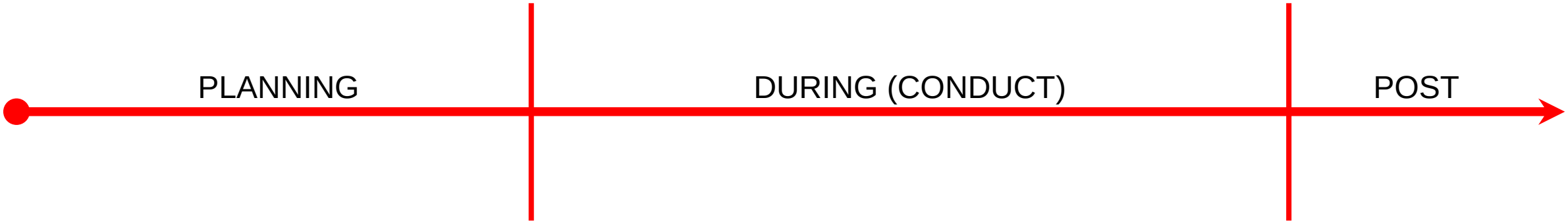
QUALITY ASSURANCE

QUALITY CONTROL

Protocol development
Documentation
Personnel: training/certification
Ethics (IRB)

Documentation
Personnel: training/certification

Writing



Data collection tools
Data validation planning
Data management planning
Pilot data collection?

Data management plan
Data collection
Data validation
Interim analyses?

Analysis
Data storage

DATA

Data Management Plan

How do you turn “raw” data to analyzable, “valid” data?

Errors can (and do occur) at every step

- Primary data
- Data extraction
  - Electronic data transfer
  - Transcription/entry into a database
- Processing (coding), storage
- Analysis

Data Management Plan - Tips

Identify data sources

- Familiarize yourself with type(s) of data available
- Manual collection
- Backup ALL raw data

Create data dictionary

- Train data abstractors

Develop data collection tools

Develop electronic database

- Data entry – predefined choices
  - MINIMIZE FREE TEXT
- Relational – need identifiers to connect databases
  - Unique to subject but present in all databases

Pilot tools and methods

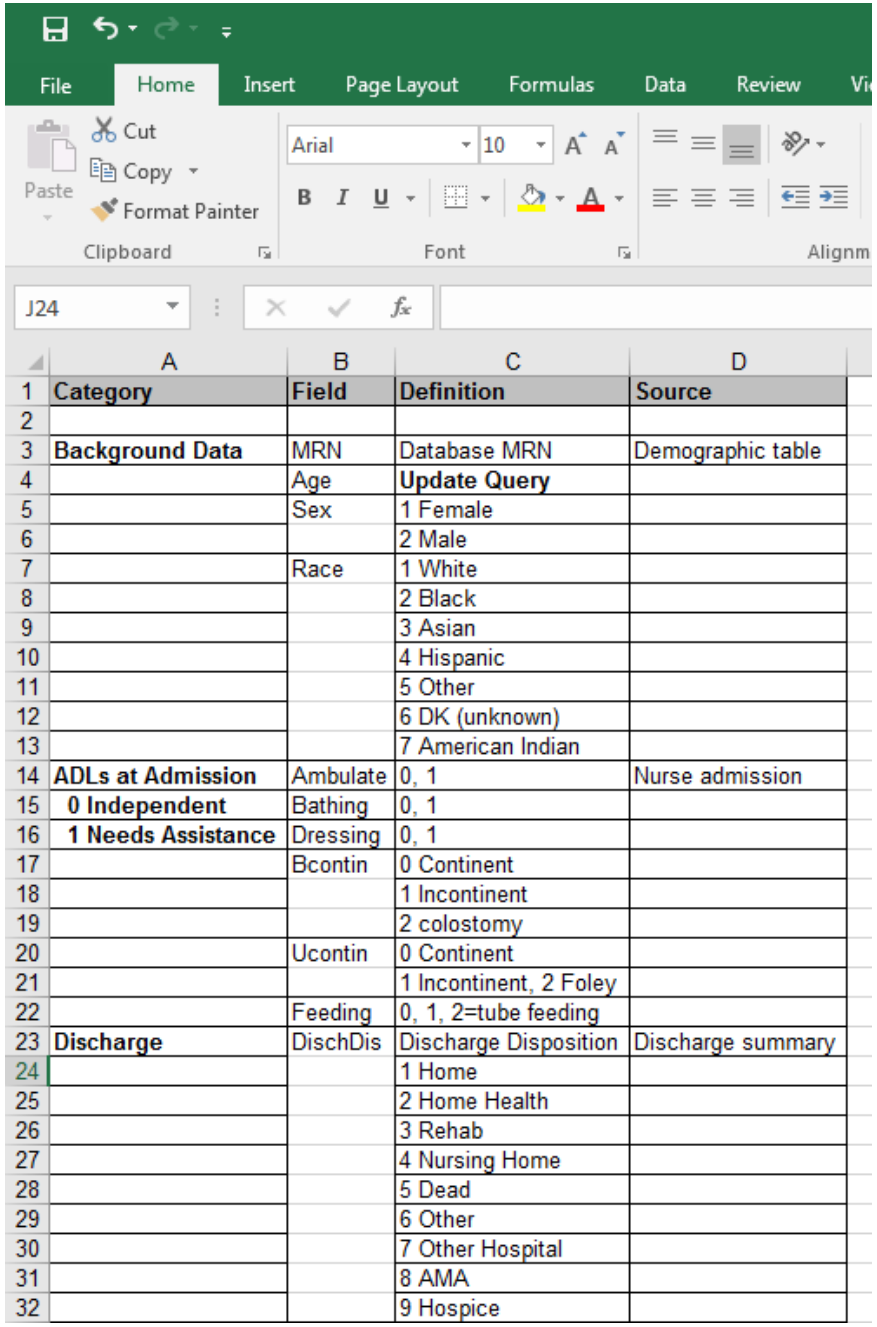
- Modify

Collect data

Clean and validate data

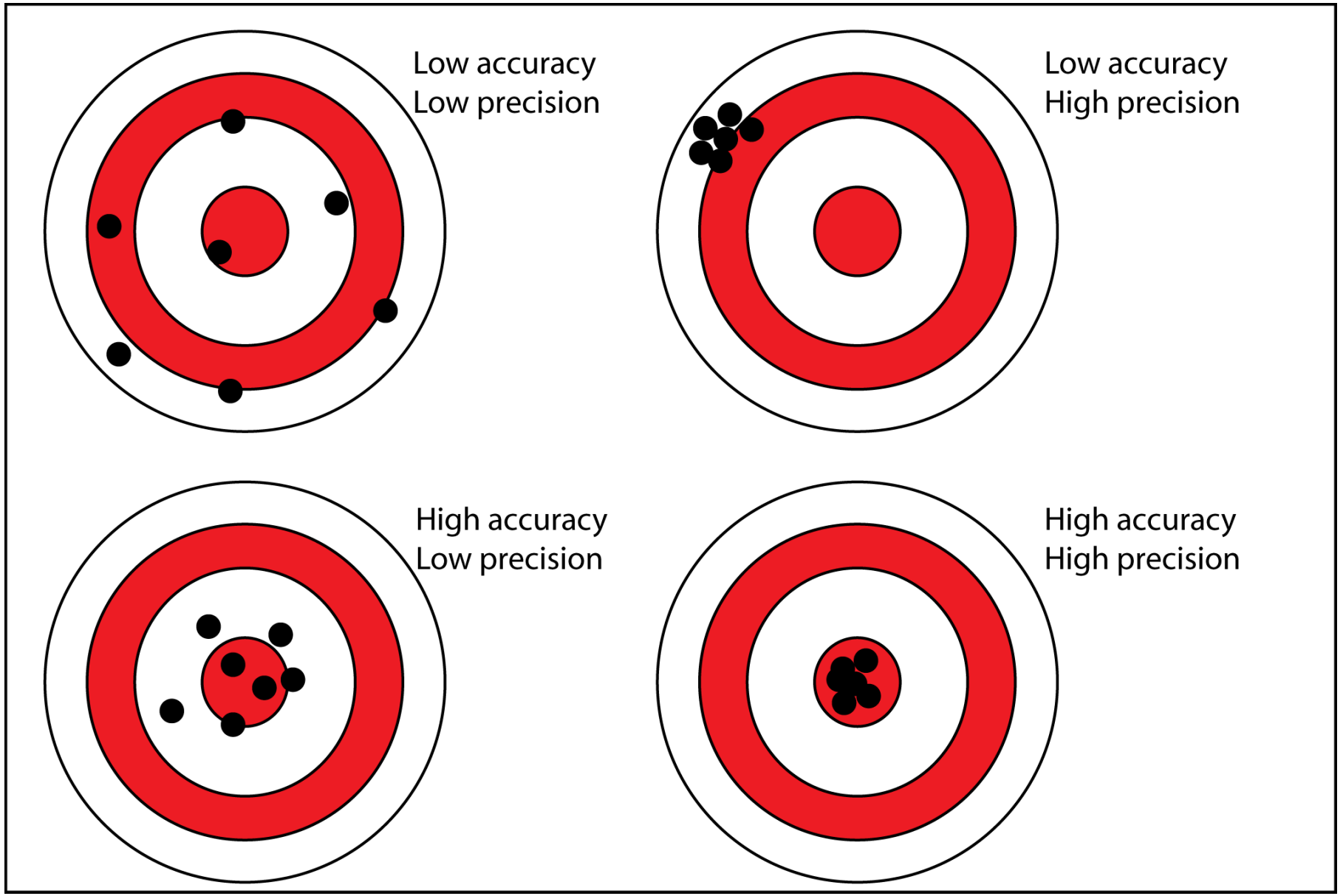
Outline security steps

Data Dictionary



	A	B	C	D
1	**Category**	**Field**	**Definition**	**Source**
2				
3	**Background Data**	MRN	Database MRN	Demographic table
4		Age	**Update Query**	
5		Sex	1 Female	
6			2 Male	
7		Race	1 White	
8			2 Black	
9			3 Asian	
10			4 Hispanic	
11			5 Other	
12			6 DK (unknown)	
13			7 American Indian	
14	**ADLs at Admission**	Ambulate	0, 1	Nurse admission
15	**0 Independent**	Bathing	0, 1	
16	**1 Needs Assistance**	Dressing	0, 1	
17		Bcontin	0 Continent	
18			1 Incontinent	
19			2 colostomy	
20		Ucontin	0 Continent	
21			1 Incontinent, 2 Foley	
22		Feeding	0, 1, 2=tube feeding	
23	**Discharge**	DischDis	Discharge Disposition	Discharge summary
24			1 Home	
25			2 Home Health	
26			3 Rehab	
27			4 Nursing Home	
28			5 Dead	
29			6 Other	
30			7 Other Hospital	
31			8 AMA	
32			9 Hospice	

Duke Center for
Antimicrobial Stewardship
and Infection Prevention



https://sites.google.com/a/apaches.k12.in.us/mr-evans-science-website/accuracy-vs-precision

Data Validation

Multistep process to ensure data collected represent “truth”

- Improve “accuracy”

Approach depends on type of data

Requires some type of “gold standard”

Commonly used strategies for manual abstraction:

- Multiple reviewers
- Random sample
- Key variables
- Check completeness of data collection

Data Validation - Datasets

Large datasets still need to undergo validation

Can use some of the same strategies

- Completeness of data

Additional strategies

- All variables present
- Error checking (“out of range”)
  - Dates
- New variables (drug names?)

Think about perspective

- Review of data already in dataset confirms that what you have may be accurate
- But, doesn’t confirm that ALL data are present

KEY POINT: these datasets weren’t created for your research project!

Data Validation - Datasets

DATASET

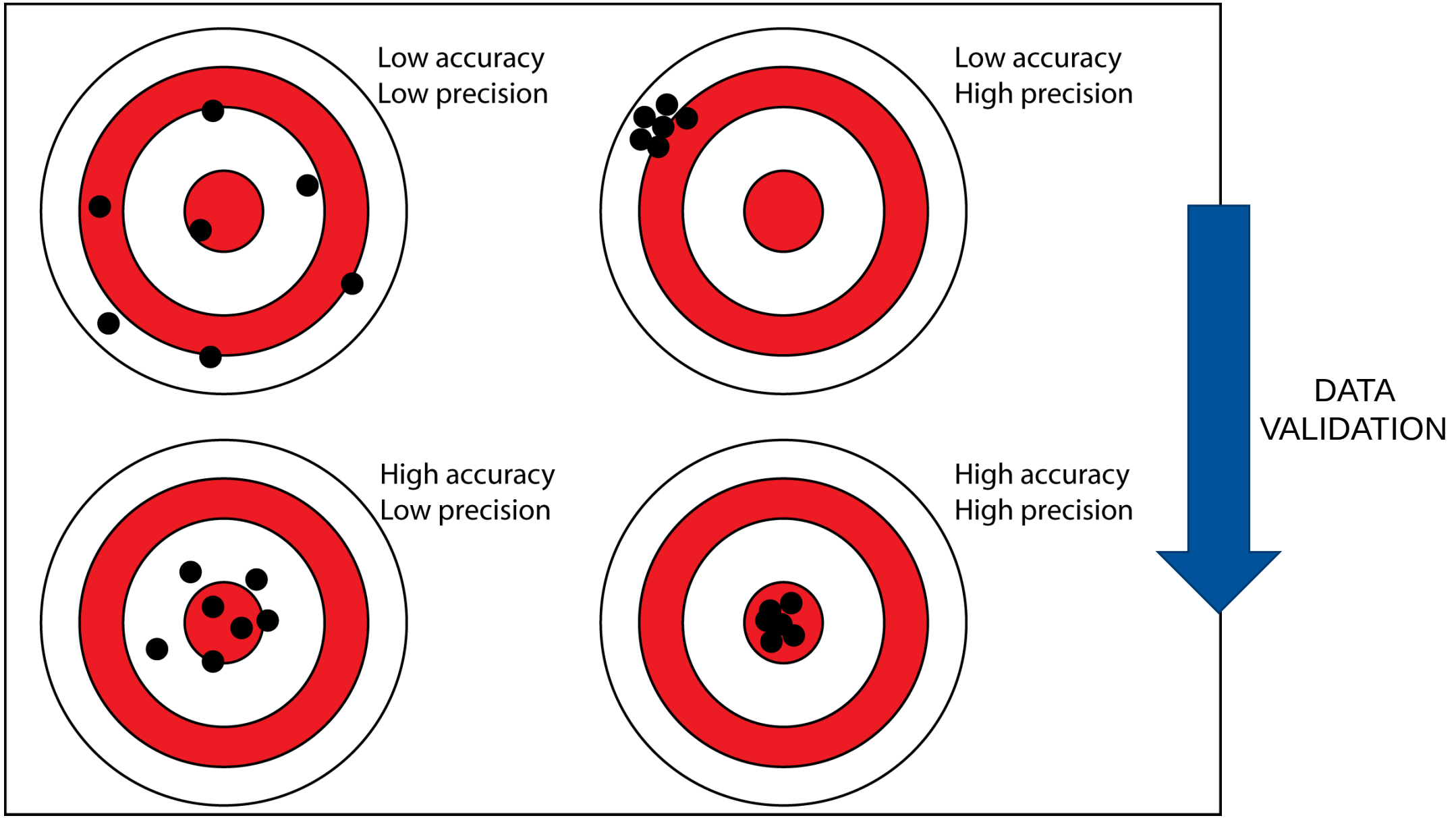
Table 3. Antimicrobial Agents and Routes Captured in Sample eMAR File.

Drug	On Report	On Report NF-	Not Used	Missing	Route Validation	Drug	On Report	On Report NF-	Not Used	Missing	Route Validation
Acyclovir	☐	☐	☐	☐	IV Y/N PO/VT Y/N	Fidaxomicin	☐	☐	☐	☐	
Amantadine	☐	☐	☐	☐		Fluconazole	☐	☐	☐	☐	
Amikacin	☐	☐	☐	☐	IV Y/N Inhaled Y/N	Foscarnet	☐	☐	☐	☐	
Amoxicillin	☐	☐	☐	☐		Fosfomycin	☐	☐	☐	☐	
Amoxicillin/ Clavulanate	☐	☐	☐	☐		Ganciclovir	☐	☐	☐	☐	
Amphotericin B	☐	☐	☐	☐		Gemifloxacin	☐	☐	☐	☐	
Amphotericin B liposomal	☐	☐	☐	☐		Gentamicin	☐	☐	☐	☐	IV Y/N Inhaled Y/N
Ampicillin	☐	☐	☐	☐		Imipenem/ Cilastatin	☐	☐	☐	☐	

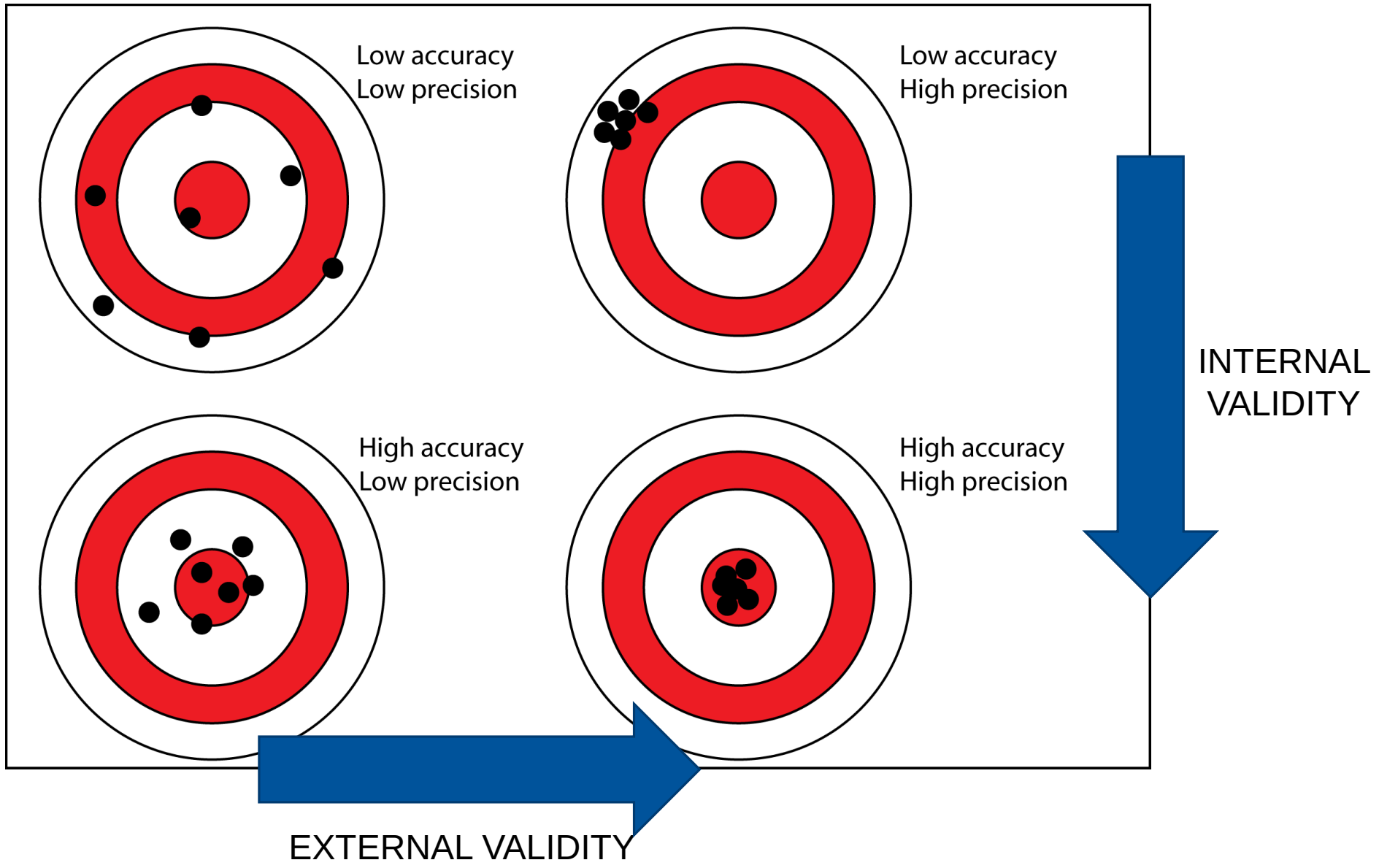
RANDOM PATIENT SAMPLE

Table 4. Manual Validation of Patient Records as compared to sample eMAR file.

	Patient MRN	Date(s)	Unit	Comments
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				



https://sites.google.com/a/apaches.k12.in.us/mr-evans-science-website/accuracy-vs-precision



https://sites.google.com/a/apaches.k12.in.us/mr-evans-science-website/accuracy-vs-precision

Beware the Preexisting Database

Key consideration in study design – prospective vs. retrospective

- Retrospective study designs more prone to various types of bias

Some advantages

- Decrease time/effort
- Availability
- Limited/de-identified

Just because data exist, doesn't mean should be used for your study

- Incomplete
- Not validated

Drees et al. ICHE 2016; 37:1278.

Preexisting Data – Surveillance Data

Statewide review of CLABSI surveillance data in Connecticut

Trained reviewers from DPH acted as “gold standard”

- Reviewed positive blood cultures from 30 hospitals

Results: >50% underreporting of CLABSI

	CT hospital reports to the National Healthcare Safety Network		
CT DPH reviewers	**CLABSI**	**No-CLASBI**	**Total**
CLABSI	23	25	48
No-CLABSI	4	424	428
Total	27	449	476

Backman et al. AJIC 2010;38:832-8.

Preexisting Data – Surveillance Data

Similar study in Oregon

Largely same results, but variation across hospitals

Change in CLABSI incidence after validation	No. (%)[a] of hospitals
Decreased by 0.70	1 (2)
No change	33 (75)[b]
Increased by 0.01–0.50	2 (5)
Increased by 0.51–1.00	2 (5)
Increased by more than 1.00	6 (14)[c]
Total	44 (100)

Oh et al. ICHE 2012;33:439.

Preexisting Data – Billing Data

Review of CLABSI data from 3 hospitals

- Surveillance (IC) vs. billing (ICD-9, used for HAC)

Variable	No. (%) of cases	Sensitivity, %	PPV
Overall (n = 890)		14	55
Concordant	112 (13)	...	...
IC only	686 (77)	...	...
HAC only	92 (10)	...	...

Moehring et al. ICHE 2013;34:238-44.

Preexisting Data – Administrative Data

Pharmacy administrative databases different from administration databases (eMAR)

- Cost/purchasing

32 units in Canada

- Compared DDD from pharmacy system to DDD from eMAR

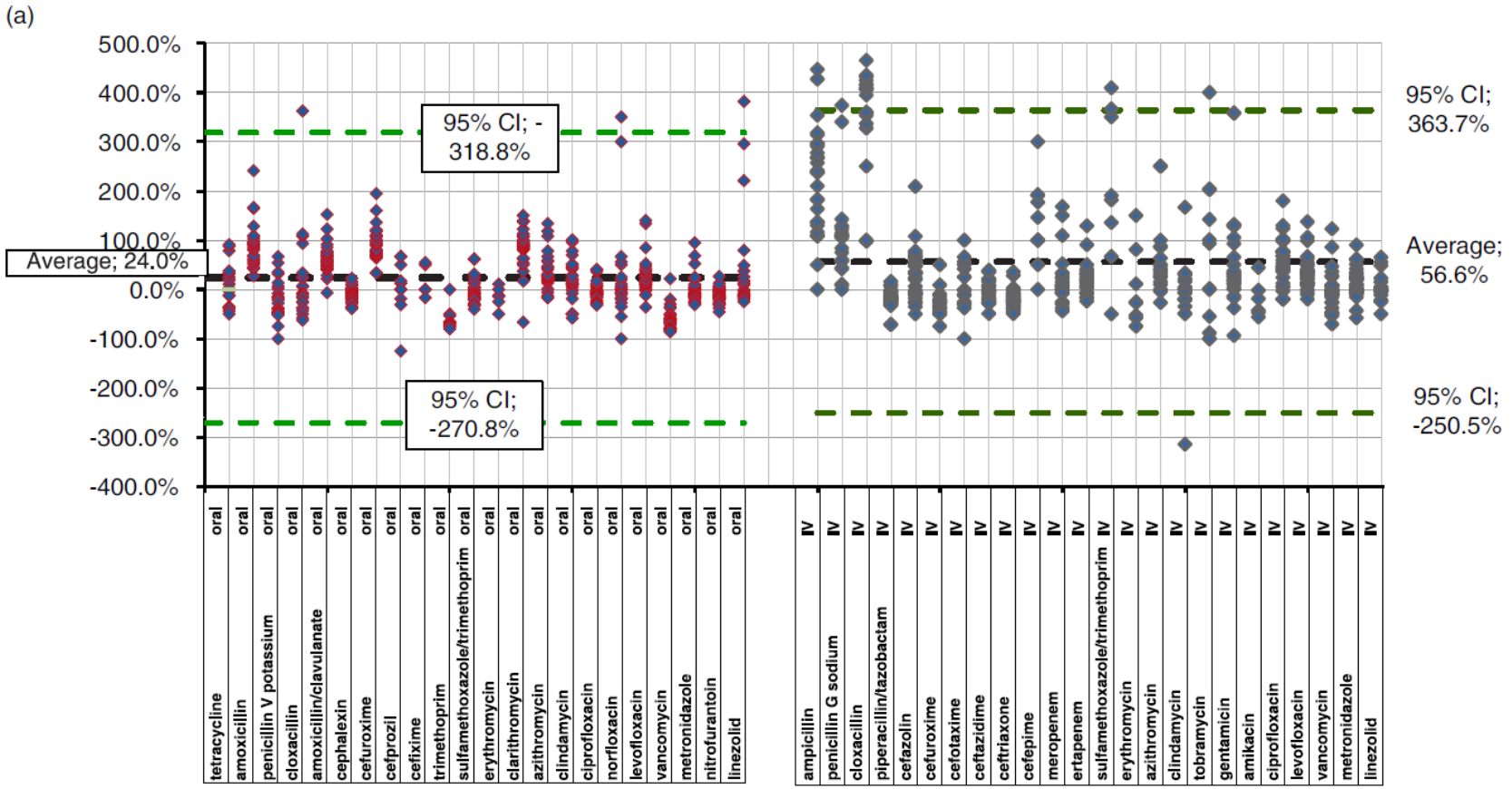
Dalton et al. ICHE 2015;36(6):688-94.

Pharmacy DDD – eMAR DDD

Average differences:

24% for PO abx

57% for IV abx



Dalton et al. ICHE 2015;36(6):688-94.

Duke Center for Antimicrobial Stewardship and Infection Prevention

Beware the Preexisting Database

Don’t fit your question to the data, find data that fit your question

Bottom Line: Don’t avoid retrospective research with preexisting dataset, KNOW LIMITATIONS

- Data inaccuracies (“noise”) stable over time?
- Know strategies to improve quality

HYPOTHETICAL EXAMPLE

Stewardship Hypothetical Example

Objective: to determine if restriction vs. post-Rx review leads to better utilization of antimicrobial therapy

Protocol development

- Define interventions
- Eligible patients
- Location
- Statistician

Example – Data Management

Data source: _____

- Obtain utilization data from eMAR
- OTHER?
- Save raw file

Data dictionary – Key variable:

_____

Data collection: _____

Electronic database: _____

- Need identifiers to link datasets

Data validation strategy: _____

Special Scenario – Multicenter Research

Multicenter research ultimately preferred

- Increases external validity

Complexity of data management increased

- Number of centers = number of different ways a process might happen

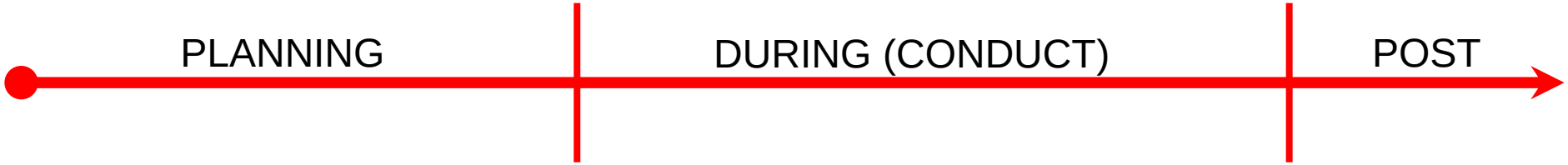
Data management plan developed centrally and distributed to participating centers

- QA/QC
  - Participating centers must perform local QA/QC
  - Central location likely adds an additional layer of QA/QC
    - Data checks
    - Data feedback/reports for participating centers

Central location must have a system to receive data from all participating centers

Take Home Points

Data management involves all the stops on the data voyage for your project



Component of QA/QC

Practical tips to increase internal validity/minimize bias:

- Develop a study protocol
- Write a data management plan
- Perform data validation
- Pay attention to the details

SHEA White Paper Series

RESEARCH METHODS IN HEALTHCARE EPIDEMIOLOGY AND ANTIMICROBIAL STEWARDSHIP

RCT

- Anderson et al. ICHE 2016;37:629.

Quasi-experimental

- Schweizer et al. ICHE 2016;37:1135.

Observational studies

- Snyder et al. ICHE 2016;37:1141.

Mathematical modeling

- Barnes et al. ICHE 2016;37:1265.

Survey and qualitative research

- Safdar et al. ICHE 2016;37:1272.

Administrative and surveillance databases

- Drees et al. ICHE 2016;37:1278.

References

Whitney et al. *Epidemiol Rev* 1998;20:71-80.

Neta et al. Quality Control and Good Epidemiological Practice. In: Handbook of Epidemiology, 2nd Ed. ED: Ahrens and Pigeot.