Good Data Practices
VIREC Cyber Seminar Series

September 9 – 13, 2013

Acknowledgements

- Initial concept development work by Michael Berbaum, PhD, University of Illinois at Chicago
- Review of existing public online course content and resources by Margaret Browning, PhD and Linda Kok, MA, VA Information Resource Center
- Examples from ongoing and completed research supported by the VA, NIH, and PCORI
- Feedback from colleagues, especially Elizabeth Tarlov, Kevin Stroupe, Matt Maciejewski, and Laurel Copeland

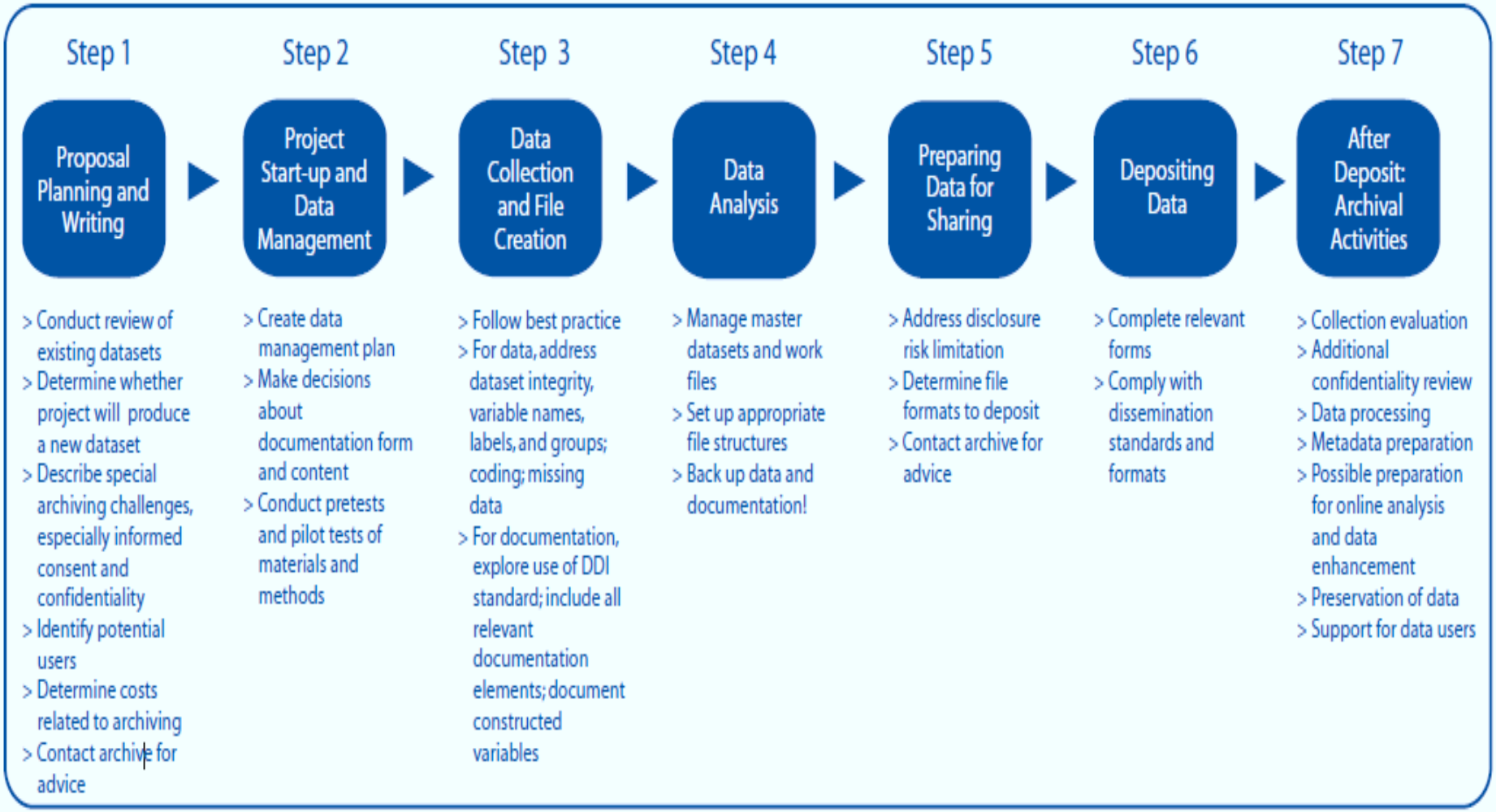
VIReC
RESEARCHERS' GUIDE TO VA DATA

Good Data Practices Series Overview

Mon 9/9/13	1	Early Data Planning for Research - Denise Hynes, PhD, MPH, RN
Tue 9/10/13	2	Managing and Documenting Data Workflow - Denise Hynes, PhD, MPH, RN
Wed 9/11/13	3	Planning for Data Re-use - Linda Kok, MA
Thu 9/12/13	4	Research Application - Laurel Copeland, PhD
Fri 9/13/13	5	Research Application - Matthew Maciejewski, PhD

Not about…

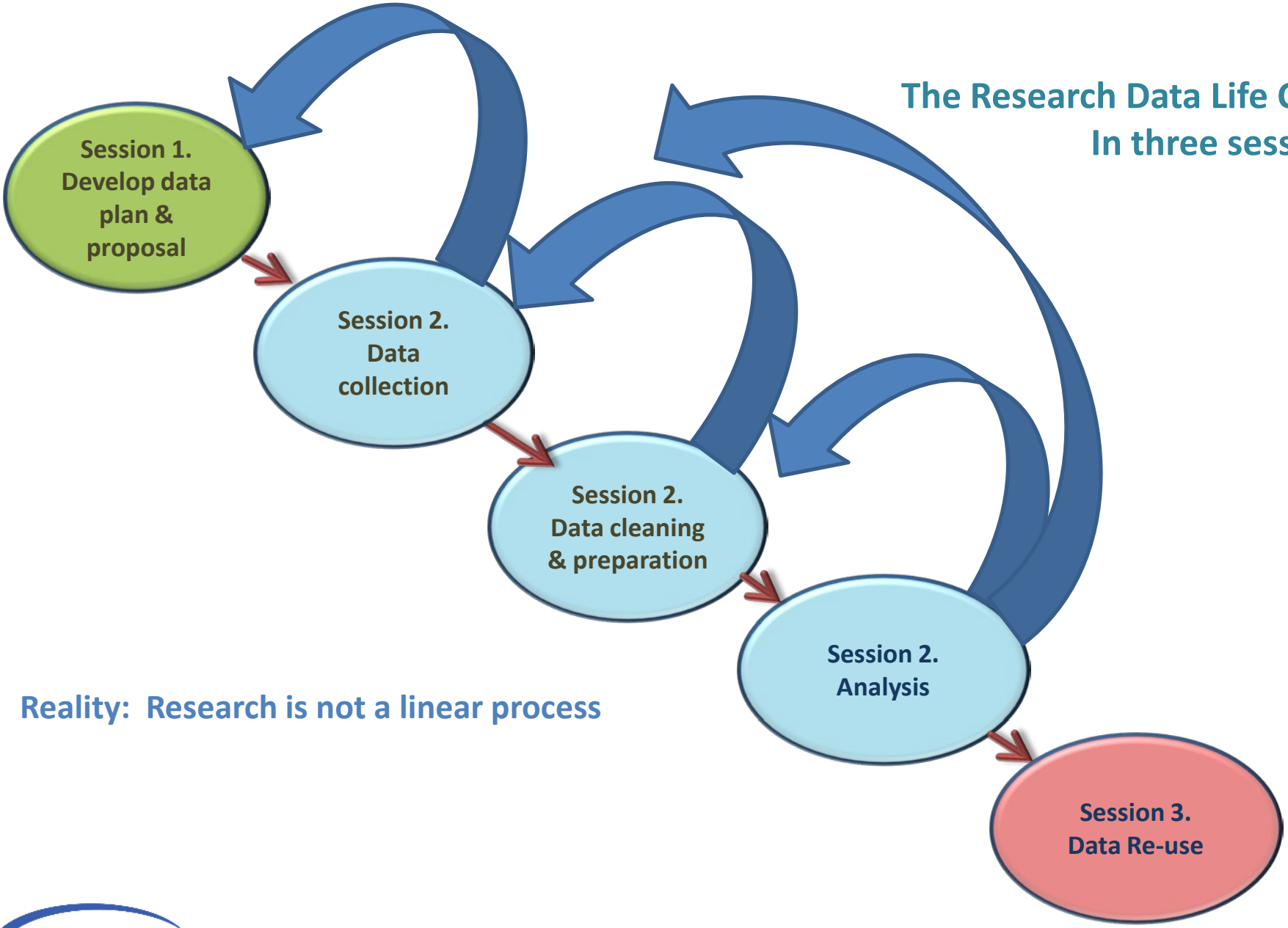
- How to design a research study
- How to execute each step of the research study
- How to structure data
- How to analyze data
- How to get funding ☺



VIReC RESEARCHERS' GUIDE TO VA DATA

Inter-university Consortium for Political and Social Research (ICPSR). (2009). *Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle* (4th ed.). Ann Arbor, MI. ISBN 978-0-89138-800-5

The Research Data Life Cycle
In three sessions



Reality: Research is not a linear process

VIReC
RESEARCHERS' GUIDE TO VA DATA

Session 1: Early Data Planning for Research

- Importance of data planning
- Factors that influence data needs
- Additional data planning for IRB submission
- Data planning checklist

Poll Question #1

- What is your primary research role?
  - Investigator
  - Data analyst/programmer or statistician
  - Research coordinator or assistant
  - Student, trainee, or fellow

VIReC
RESEARCHERS' GUIDE TO VA DATA

Poll Question #2

- What would you say is your level of research experience?

  1 (Novice)

  2

  3

  4

  5 (Expert)

Session 1: Outline

- **Importance of data planning**
- Factors that influence data needs
- Additional data planning for IRB submission
- Data planning checklist

Importance of data planning

- Forces you to think clearly about work flow
- Helps identify issues before you begin
- Provides a guide for project team
- Helps you write/refine your protocol
- Serves as a reference for future work

- **Simplifies your life in the long run**

VIReC
RESEARCHERS' GUIDE TO VA DATA

Session 1: Outline

- Importance of data planning
- **Factors that influence data needs**
- Additional data planning for IRB submission
- Data planning checklist

Factors that influence data needs

- Research question
- Study design
  - Objectives, aims, hypotheses
  - Independent and dependent variables
  - Planned manuscripts
- Available data
- Feasibility testing

Factors that influence data needs

- **Research question**
- Study design
  - Objectives, aims, hypotheses
  - Independent and dependent variables
  - Planned manuscripts
- Available data
- Feasibility testing

Example:
"FINER" criteria for a good research question

Feasible	• Adequate number of subjects • Adequate technical expertise • Affordable in time and money
Interesting	• Getting the answer intrigues investigator, peers and community
Novel	• Confirms, refutes or extends previous findings
Ethical	• Consistent with institutional review board standards
Relevant	• To scientific knowledge • To clinical and health policy • To future research

From: Hulley S, Cummings S, Browner W, et al. Designing clinical research. 3rd ed. Philadelphia (PA): Lippincott Williams and Wilkins; 2007.

Example:
"D-FINER" criteria for a good research question

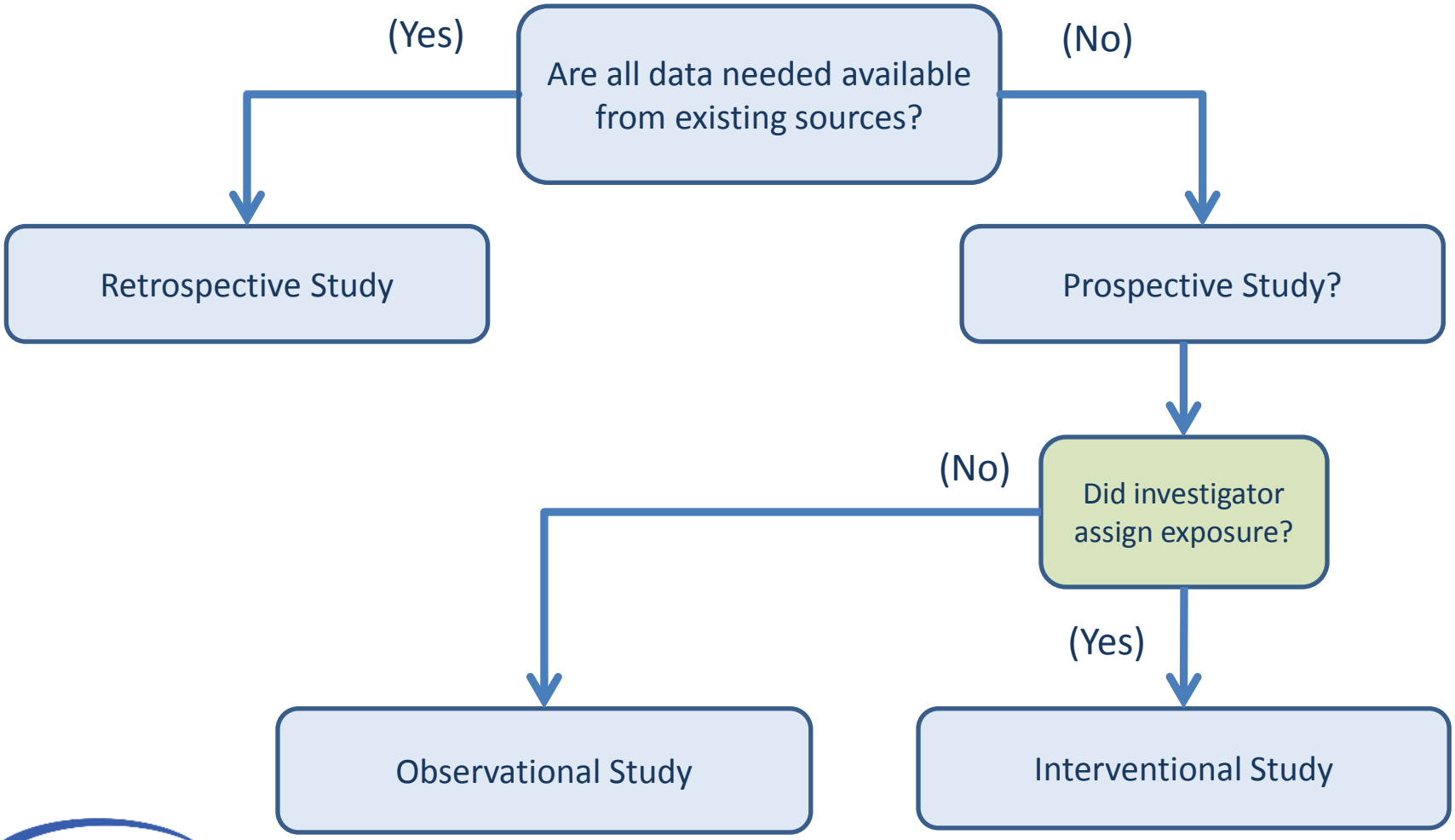
Feasible	• Adequate number of subjects • Adequate technical expertise • Affordable in time and money • Practical data plan
Interesting	• Getting the answer intrigues investigator, peers and community • Makes use of data in new ways
Novel	• Confirms, refutes or extends previous findings • Produces new data that may be useful to others
Ethical	• Consistent with institutional review board standards • Data plan protects privacy of subject data
Relevant	• To scientific knowledge • To clinical and health policy • To future research • To future data sharing

From: Hulley S, Cummings S, Browner W, et al. Designing clinical research. 3rd ed. Philadelphia (PA): Lippincott Williams and Wilkins; 2007.

Factors that influence data needs

- Research question
- **Study design**
  - Objectives, aims, hypotheses
  - Independent and dependent variables
  - Planned manuscripts
- Available data
- Feasibility testing

Factors that influence data needs



ICPSR | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

A PARTNER IN SOCIAL SCIENCE RESEARCH

Guide to Social Science Data Preparation and Archiving

Best Practice Throughout the Data Life Cycle

4th edition

VIReC RESEARCHERS' GUIDE TO VA DATA

Factors that influence data needs

- Research question
- Study design
  - **Objectives, aims, hypotheses**
  - **Independent and dependent variables**
  - Planned manuscripts
- Available data
- Feasibility testing

Factors that influence data needs

- Data to support your research objectives/aims/hypotheses
  - What type of data? PHI or non-PHI?
  - How much data?
  - Unit of analysis
    - One record or multiple records per
      - Event
      - Subject
      - Encounter, claim, inpatient stay
      - Facility, VISN
      - County, zip code
  - Temporal aspects
  - Measurement frequency
  - Informatics tools

Table 4 ■ Informatics Strategies and Data Sources of Selected Veterans Affairs Quality Enhancement Research Initiative Disease-Focused Service Directed Projects*

QUERI Disease Focus Research Project Title	Research Objectives	Data Collection Approaches	Informatics Tools
IHD: Translation plan to improve lipid management for patients with IHD	(1) Improve health of veterans with IHD in VISN 20 via increases in lipid level measurement, appropriate use of lipid-lowering agents, number of patients with low-density lipoprotein levels at guideline recommendation (2) Further evaluate and refine intervention from Lipid Measurement and Management System pilot in VISN 20 using continuous quality improvement (3) Develop and evaluate Lipid Clinical Reminder (4) Design, test, and implement national data reporting system for lipid measurement and management	Audiotaped interviews with clinicians; vitals, laboratory data, recruitment data, and clinical reminders retrieved from existing databases	Lipid clinical reminders created and implemented in CPRS; vitals, laboratory data, recruitment, and clinical reminders data retrieved via CPRS

Citation: Hynes, DM, et al., Informatics Resources to Support Healthcare Quality Improvement Research in the Veterans Health Administration. *J of the American Medical Informatics Association.* Sep/Oct 2004;11(5):344-350.

Table 4 ■ Informatics Strategies and Data Sources of Selected Veterans Affairs Quality Enhancement Research Initiative Disease-Focused Service Directed Projects*

QUERI Disease Focus Research Project Title	Research Objectives	Data Collection Approaches	Informatics Tools
MH: MH QUERI translation plan to improve antipsychotic treatment	(1) Compare QUERI-recommended intervention to basic intervention regarding medication management in schizophrenia (2) Prepare MH QUERI translation package for pilot test of national rollout strategy in two VISNs (3) Facilitate a pilot of the QUERI-recommended intervention in two VISNs	Chart reviews to determine medication adjustment needs pre- and post-intervention; survey for clinicians regarding guidelines and use of informatics tools	Clinical reminders created and implemented within CPRS for medication prescribing practices; VistA Structured Query Language
SCI: Increasing influenza vaccination rates in the veteran SCI population	(1) To increase influenza vaccination rates and decrease morbidity and mortality due to respiratory infection in veterans with SCI at eight pilot sites	Vaccination rates: patient surveys and data supplied by the Office of Quality and Performance External Peer Review Program chart review program; hospital discharges and visits for respiratory infections–NPCD	NPCD; Spinal Cord Dysfunction Registry database

QUERI = quality enhancement research initiative; IHD = ischemic heart disease; VISN = Veterans Integrated System Network; MH = mental health; SCI = spinal cord injury; CPRS = Computerized Patient Record System; NPCD = National Patient Care Database.

Factors that influence data needs

- Research question
- Study design
  - Objectives, aims, hypotheses
  - Independent and dependent variables
  - **Planned manuscripts**
- Available data
- Feasibility testing

Factors that influence data needs

- Research question
- Study design
  - Objectives, aims, hypotheses
  - Independent and dependent variables
  - Planned manuscripts
- **Available data**
- Feasibility testing

Factors that influence data needs

- Available data
  - Conduct review of existing datasets
  - Determine whether project will produce a new dataset
  - Anticipating future sharing, consider special archiving challenges, especially informed consent and confidentiality

Factors that influence data needs

- VA data available for your research
  - Sources of information about VA data
    - VHA Data Portal
    - VA Information Resource Center (VIReC)
    - HSRData-L Listserv
    - Health Economics Resource Center (HERC)
    - Corporate Data Warehouse (CDW) Metadata Report
    - Patient Care Services (PCS)
    - Pharmacy Benefits Management Services (PBM)
    - OIA HIG Health Data Quality Program

Factors that influence data needs

- Access to VA data
  - How do I request VA data?
    - VHA Data Portal Data Access Pages
      - Research & operations data requests
      - DART – NDS managed data
    - VIReC – VA/CMS data
    - PBM – Pharmacy data

- VIReC Database and Methods series, Research Access to Data November 4, 2013

Welcome to the VHA Data Portal!

The one-stop-shop for data users' needs.

The VHA Data Portal is a collaborative effort among the following program offices to provide a central gateway to information about VHA data:

- **VHA National Data Systems (NDS)**
- **VA Information Resource Center (VIReC)**
- **VA Informatics and Computing Infrastructure (VINCI)**
- **VHA Data Quality Program**

Our mission is to promote a knowledge-sharing culture that supports the needs of VHA data users. The portal integrates information from multiple sources into a common format and single location to promote a comprehensive knowledge base and to facilitate a positive end-user experience.



News

NDS Data Access Update

The new NDS Healthcare Operations form is here! The form streamlines the access request process for NDS Healthcare Operational requests. Visit the Operations Access page for more information and to view the new form.

VINCI Wants Your Feedback!

Please take the VINCI Customer Satisfaction Survey and provide us with your feedback about VINCI.

Upcoming Events

BISL Live Meeting Training of the Month

August 28: Part 3 of 3 SAS Grid Analytics Advance

September 25: Part 3 of 3 SAS Grid Analytics Advance

VIReC Cyber Seminars

September 17: Using Patient-Facing Kiosks to Support Quality Improvement at Mental Health Clinics

September 9-13: Good Data Practices Mini-Series

Popular Links

- Launch VINCI Workspace
- VINCI Collaboration Sites
- VIReC Intranet Site
- CDW SharePoint Site
- VHA Data Quality Program Intranet Site
- Corporate Databases Monograph
- HSRData-L Listserv
- DART Overview and Forms
- Launch DART Application
- RAMP

VA Information Resource Center (VIReC)

VIReC Home
VA/CMS Home
About Us
New Users of VA Data
Data Transition to CDW
News & Updates
FAQs
Acronyms
HelpDesk

Data Issues Brief, July 2013
Upcoming Cyber Seminar
Medical Care Journal: HIT in VHA Research
VIReC News & Updates
Data Transition to CDW

At a Glance

Introduction to VIReC and VA data: Learn about VIReC's role in VA research and how to navigate our website.

Data Issues Brief: VIReC's monthly newsletter provides researchers current news and updates.

HSRData-L Listserv: Join our virtual community of VA researchers who share knowledge and experiences about VA data and information systems.

VA/CMS Data for Research Project: VIReC serves as the data custodian for Centers for Medicare and Medicaid Services (CMS) data for research use in the VA.

Journal Supplements: VIReC collaborates with peer reviewed journals to publish supplements on relevant VA data and informatics research topics.

Requesting and Accessing Data

Data Access Tools: Information on data access tools, including applications and analytic workspaces.

Data Access and Request Guide: A complete list of access guidelines for the data sources most commonly used by VA researchers.

Preparatory to Research: An overview of preparatory to research data use regulations, including access information.

Data Access Request Tracker: Introduction to DART including, how to register and submit requests and resources for using the online application.

DART Metrics: View the current processing report for DART requests available from NDS (see current DART metrics).

Resources for Researchers

Data Sources and Data Topics: Select a specific data source or data topic described by VIReC.

Research User Guides (RUGs): Detailed descriptions of select VA data sources, including variable descriptions.

Data Reports: A complete list of technical reports, data investigations, data quality updates, and QUERI reports.

Summary Information: VIReC provides summary information such as Historical Variable Attributes and Variable Frequencies for select data sources.

Publications: A complete list of peer reviewed articles and journal supplements published by VIReC.

Education

VIReC Cyber Seminars: Expert discussion on key issues in clinical informatics and databases and methods for VA researchers.

Toolkit for New Users: A starting point for new users of VA data and experienced researchers using new data sources.

Tutorials: Step-by-step instruction on accessing and using select VA data and information systems.

Presentations: Learn about topics relevant to VA research from experts in the field.

HERC
Health Economics Resource Center

HEALTH ECONOMICS RESOURCE CENTER

HERC Home
News »
Resources »
Data »
- Data Overview
- Average Cost Data
- Decision Support System (DSS)
- Financial Management System (FMS)
- PAID
- Cost Distribution Report (CDR)
- Fee Basis Files
- Fixed Asset Package
- VA Utilization Files
- Database of VA Facilities
- Pharmacy Benefits Management Database
- Datasets not at the VA
- Tabulations
- VHA Labor Costs

Events

ECONOMETRICS CYBERCOURSE

OCT 2
Econometrics Course: Introduction & Identification
Todd Wagner, Ph.D.
REGISTER »

CYBERSEMINARS

SEP 18
Posttraumatic Stress Disorder, Military Sexual Trauma and Preterm Birth – Evidence from 16,000 VA Pregnan
Jonathan Shaw, M.D.
REGISTER »

LEARN MORE »

News

August 15, 2013
"Guidebook for Research Use of PAID Data" has been updated.

August 8, 2013
A new HERC Bulletin (Vol 13, Iss 3) is now available.

July 25, 2013
PROC CONTENTS added for FY12 outpatient average cost data

MORE »

About Us

HERC is a national center located in Menlo Park, CA that assists VA researchers in assessing the cost-effectiveness of medical care, evaluating the efficiency of VA programs and providers, and conducting high-quality health economics research.

MORE »
CONSULTING SERVICE »
FREQUENTLY ASKED QUESTIONS »

Top FAQs

A1. What is cost-effectiveness analysis?

B6. How do VA costs compare to the cost of non-VA providers?

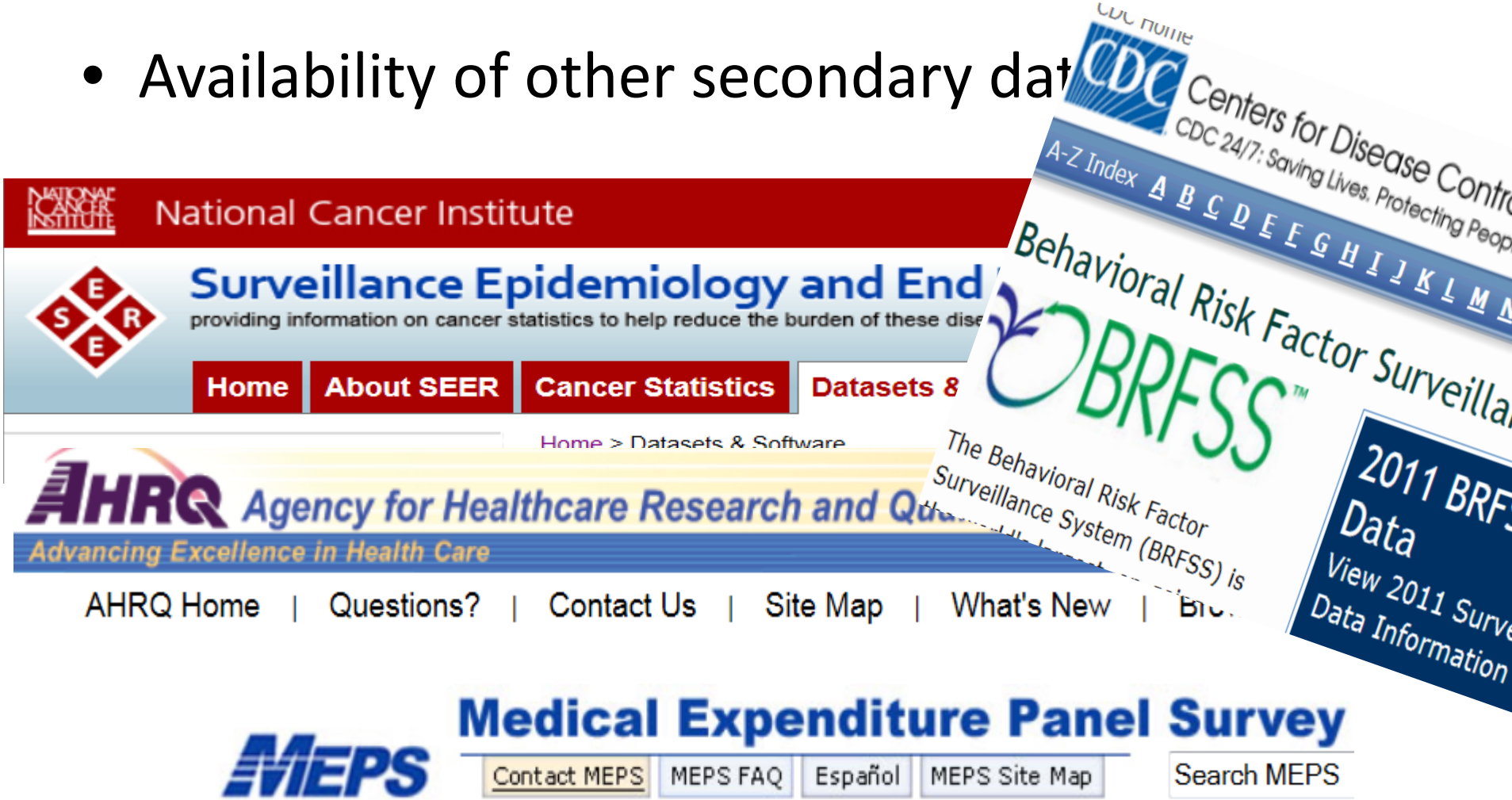
E2. What is retransformation bias, and how can it be corrected?

A3. How Do I Adjust for the Effects of Inflation?

B7. Comparison of VA and Medicare costs

Factors that influence data needs

- Availability of other secondary dat



Comparison Cohorts

Factors that influence data needs

- Planning data collection/acquisition
  - Will you need data directly from subjects?
    - Do you need any special software to collect the data?
  - Do you need site-specific information?
    - Is it available in a public data set or will you need to interview key informants?
  - How much detail do you need about events of interest?
  - How will primary data be integrated with secondary data?

Factors that influence data needs

- Availability of electronic medical record data
  - How many different EMR systems will need to be accessed?
  - What processes do you need to plan for data extractions?
  - How well do you understand the data structures?
    - What you see in a user interface is not always what you get from the operational data store!
  - Is there a data warehouse of EMR data available along with experts to consult?
  - How will you link data with other data sources?

Factors that influence data needs

- Primary data collection
  - Paper/Scanned forms
  - Electronic data entry
  - Clinical data management systems
    - Research Electronic Data Capture (REDCap)
      - » Available soon
      - » Multi-site, web-based data entry within the VA firewall
      - » Data export procedures to SAS, SPSS, STATA and R
      - » A shared library of data collection instruments and forms
  - Clinical trials management systems
  - Electronic data capture /integration with other records

Factors that influence data needs

- Other primary data collection issues
  - One-time or repeated data capture?
  - Length of each collection period
  - Amount of data generated
  - Type of data -- Images?
  - Will data be shared with other sites/collaborators

Factors that influence data needs

- Secondary data considerations
  - Existing data sources needed
  - Identifier
    - Are real SSNs necessary?
  - Amount of data to be collected (primary & secondary)

Factors that influence data needs

- Plan for data integration
  - Data from multiple data sources
    - Primary + secondary data
    - Multiple primary + multiple secondary data from different sources
    - Structured + unstructured data
  - Plan for combining or integrating data
  - Plan for sharing data during and after the study

Factors that influence data needs

- Integrating data and documentation
  - Using integrated software
  - Computer-assisted interviewing

Factors that influence data needs

- Research question
- Study design
  - Objectives, aims, hypotheses
  - Independent and dependent variables
  - Planned manuscripts
- Available data
- **Feasibility testing**

Example:
"D-FINER" criteria for a good research question

Feasible	• Adequate number of subjects • Adequate technical expertise • Affordable in time and money • Practical data plan
Interesting	• Getting the answer intrigues investigator, peers and community • Makes use of data in new ways
Novel	• Confirms, refutes or extends previous findings • Produces new data that may be useful to others
Ethical	• Consistent with institutional review board standards • Data plan protects privacy of subject data
Relevant	• To scientific knowledge • To clinical and health policy • To future research • To future data sharing

VIReC RESEARCHERS' GUIDE TO VA DATA

Adapted from: Hulley S, Cummings S, Browner W, et al. Designing clinical research. 3rd ed. Philadelphia (PA): Lippincott Williams and Wilkins; 2007.

Factors that influence data needs

- Feasibility testing
  - Adequate technical expertise
    - Data management
    - Statistical modeling & analysis
  - Affordable in time and money
  - Manageable in scope
  - Is there an adequate number of subjects characteristics to test the hypothesis?

Factors that influence data needs

- Preparatory to research requirements for feasibility testing
  - **VHA 1200.05 -** Human Subjects Protection
    - Data access is only to develop one specific protocol
    - No IRB or R&D Committee approval needed
    - Access ends once the protocol has been submitted to the IRB and the Research and Development (R&D) Committee for review.
    - Pilot studies are not considered preparatory to research and must have separate IRB approval

Session 1: Outline

- Importance of data planning
- Factors that influence data needs
- **Additional data planning for IRB submission**
  - Data planning checklist

Additional data planning for IRB submission

- Data privacy
- Data security

Additional data planning for IRB submission

- Data privacy plan: How you will protect the subjects' private health information and identity?
  - Data access restrictions
  - Who will have access and how?
  - Handling real & scrambled SSNs
  - Data from external sources
  - Special subjects
    - Deceased Veterans
    - Patients with alcohol, drug, HIV & Sickle Cell information
    - Non-patient participants
- VHA Handbook 1605.1 Privacy & Release of Information
- VHA Handbook 1200.05 Requirements for Protecting Human Subjects in Research

Additional data planning for IRB submission

- Data security plan
  - Storage location
    - VA network server
    - VINCI Workspace
  - Special backup scheduling
  - Plan for data transfers between sites
  - VHA requirements

Session 1: Outline

- Importance of data planning
- Factors that influence data needs
- Additional data planning for IRB submission
  - **Data planning checklist**

Data planning checklist

- Data planning checklist
  - What existing data will you need?
  - Will you need data directly from the subjects?
  - Will you need both?
  - What is the time-frame for all data needed?
  - What are the latest years of those data available?
  - How much data will be generated? Have you planned for server space?
  - How will you link data from different sources?
  - If the data sources do not have a data element in common, is there a third data source available with an element common to each?
  - Is the software you need available?

- Source: MIT Library Guides

VIReC RESEARCHERS' GUIDE TO VA DATA

Data planning checklist

- Formalizing Your Data Management Plan
  - Description of the project
  - Description of the data to be collected
  - Standards to be applied for formats, metadata, etc.
  - Plans for short-term storage and data management: e.g., file formats, local storage and back up procedures, and security
  - Description of legal and ethical issues: e.g., intellectual property, confidentiality of study participants
  - Access policies and provisions: i.e., how will you make data available to others, any restrictions to data reuse, etc.
  - Provisions for long-term archiving and preservation
  - Assigned data management responsibilities: i.e., which persons will actually be responsible for ensuring data management; compliance monitoring over time

UCLA Library Website

http://guides.library.ucla.edu/content.php?pid=385860&sid=3182780&preview=200eac9a6bb823f4503d8413245daed7

Additional Resources

- ICPSR Data Management Plan
  - http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/
- MIT Libraries:
  - http://libraries.mit.edu/guides/subjects/data-management/plans.html
- DMPTool a service of the University of California Curation Center, California Digital Library
  - https://dmp.cdlib.org/about/dmp_about
- Digital Curation Centre, Scotland:
  - DCC Template for a Data Management Plan

QUESTIONS

VIReC
RESEARCHERS' GUIDE TO VA DATA

Contact Information

Denise Hynes, PhD, MPH, RN

VA Information Resource Center

A VA Health Services Research & Development Resource Center working to improve the quality of VA research that utilizes databases and information systems

Hines VA Hospital

VIReC@ va.gov

708-202-2413

Session 1: Early Data Planning for Research

- Recap
  - Importance of data planning
  - Factors that influence data needs
  - Additional data planning for IRB submission
  - Data planning checklist

Session 2: Managing and Documenting Data Workflow

- Preview
  - Getting started
  - Importance of documentation
  - Data management workflow
  - Analysis workflow