

# Achieving Data Quality in an Observational Study

# Presenters



**Jean Siebenaler, MD**  
**Medical Director,**  
**Innovus Late Phase**



**Andrea Spannheimer**  
**Vice President,**  
**Innovus Late Phase International**

# Agenda

- Why is data quality important in observational studies?
- Important data quality considerations by study phase
  - Planning & design/analysis phase
  - Site management in observational studies
    - Training
    - Monitoring
  - Data collection and management phase
  - Reporting & publication phase
- Guidance documents
- Summary

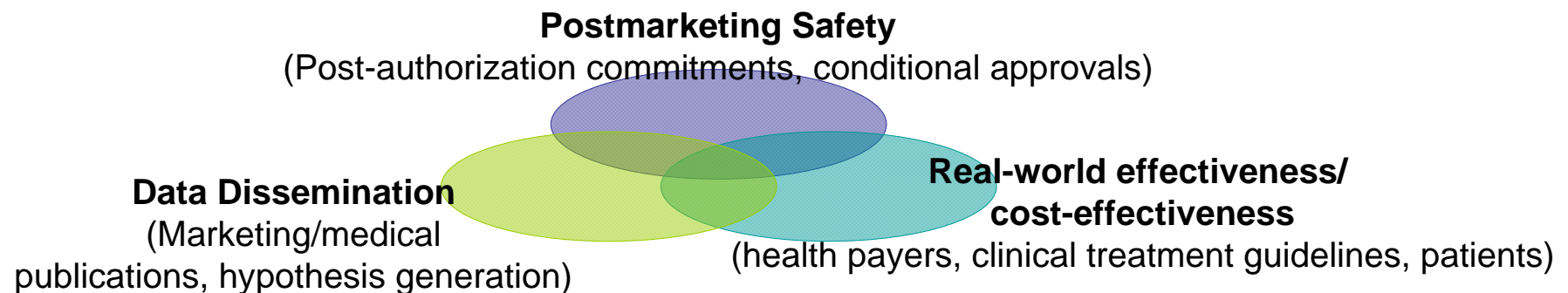
# Why is Quality Important in Observational Studies?

- Observational studies are typically designed to assess variables under real-world conditions.
- The real world is messy!



# Why is Quality Important in Observational Studies?

- Observational study results are increasingly used for objective decisionmaking on multiple levels by different stakeholders;
- Requires ability to draw valid conclusions from study data



# Why is Quality Important in Observational Studies?

- Observational studies today need to be scientifically sound
  - Data bias must be minimized with an appropriate study design and data analysis methodology;
  - Data must be authentic, complete and valid; and
  - Data reporting must be transparent.
- Basic principles of observational study data quality are laid down in corresponding guidance documents, recommendations/codes of authorities, and institutes/medical societies.
- This presentation will outline the considerations required to obtain quality data in observational studies, especially site-based studies.

\*

# Study Planning & Design Phase

Jean Siebenaler, MD  
Medical Director,  
Innovus Late Phase

**INNOVUS**<sup>™</sup>  
Late Phase Research

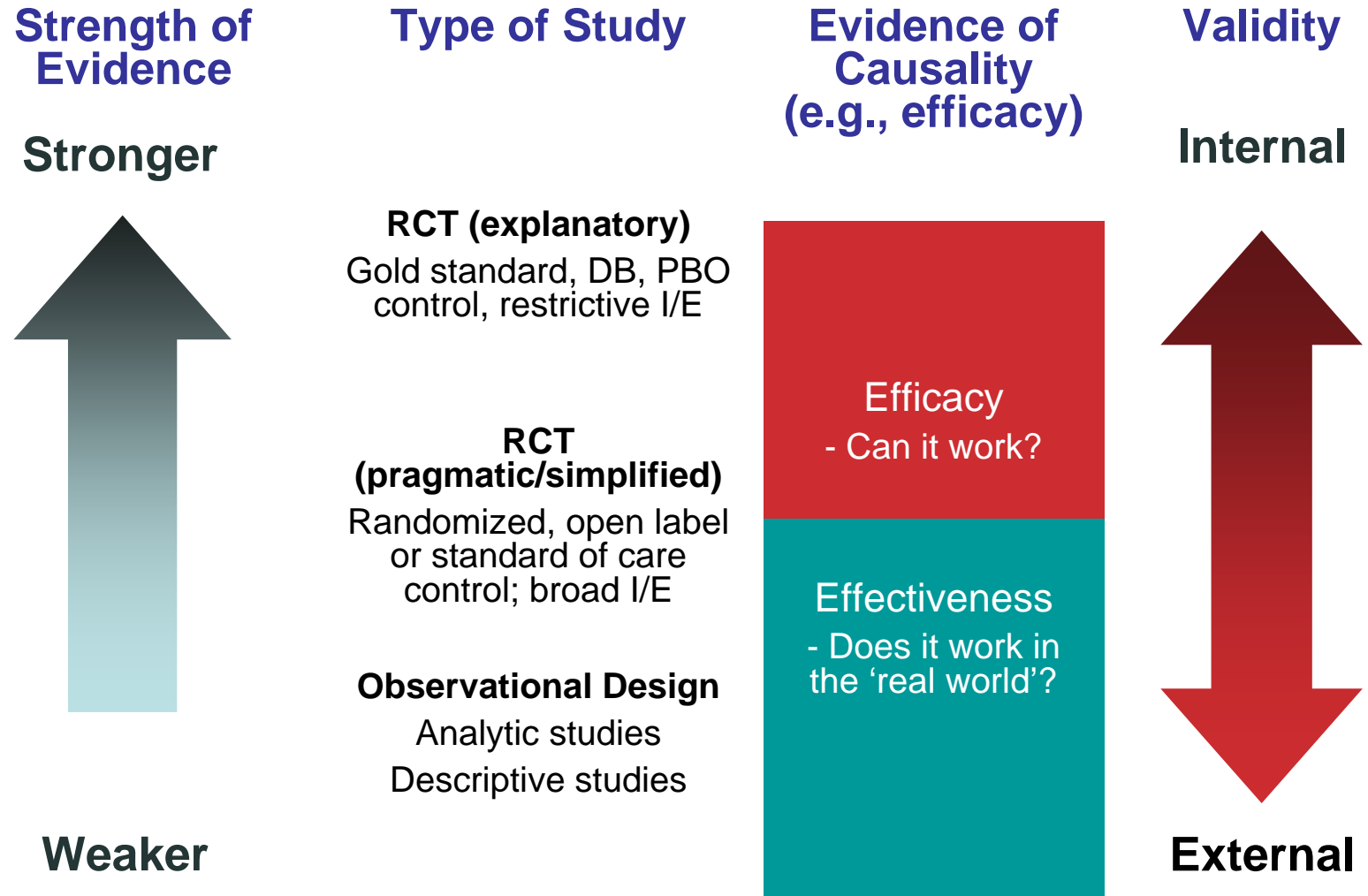
# Data Quality Starts with a Good Research Question

- Is your research question interesting and relevant to the intended audience (clinicians, patients, payers, regulators, etc.)?
- Will your research question add to the current body of knowledge in the area you are studying?
  - Confirm or refute prior data?
  - Provide new data?
- Is your research question ethically sound?
- Risks with poor research questions:
  - Inability to get stakeholder interest for funding
  - Inability to get regulatory or ethics committee approval
  - Inability to gain physician and/or patient interest for study recruitment
  - Inability to publish data in high-impact peer reviewed journals

# Is an Observational Study Your Best Choice?

- Does your research question require that you **intervene or assign** an exposure and then observe the outcome(s) of that intervention or assignment?
  - Decision: You want a **controlled clinical trial**
    - Typically randomized (i.e., RCT)
    - An analytic design (i.e., hypothesis-testing) usually selected to answer more narrowly focused research questions regarding ‘cause and effect’
    - Requires maximal internal validity at the expense of external validity (i.e., generalizability)
      - Exception: Pragmatic RCTs also require data to be generalizable as much as possible
- Do you wish to observe outcomes without intervening or assigning an exposure variable? Do you wish to primarily assess associations between variables?
  - **Decision: You want an observational study design**

# Evidence of Causality Hierarchy

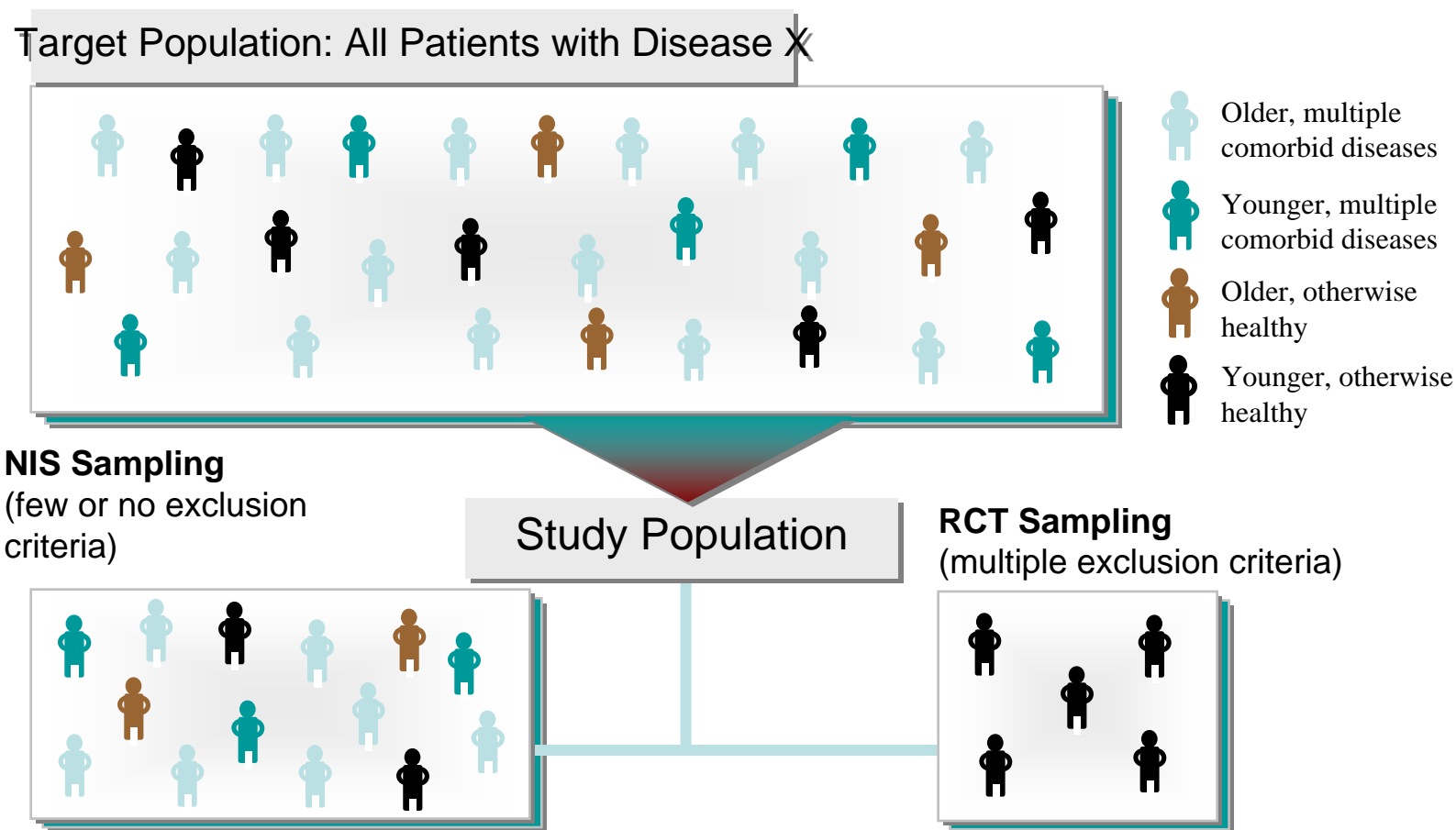


# Define the Population of Interest

- Define the **target population** that is appropriate to your research question(s)
  - It should reflect the larger population to which the study results will be generalized
  - Defined by clinical and demographic characteristics (i.e., I/E criteria) that are carefully selected to minimize selection bias and maximize validity
- Define the **accessible population**
  - Representative subset of the target population that is feasible to study based on geographic location/time
- Define the **intended study population**
  - Representative subset of the accessible population based on appropriate yet feasible **subject sampling methodologies**

# Site Selection and Patient Recruitment

- Goal is to feasibly recruit a diverse patient sample that represents the larger target population (focus on external validity)



# Site Selection and Patient Recruitment

- Sample broad range of sites that manage a broad spectrum of the target population
  - Consider site selection based on site characteristics if they will impact data results (e.g., physician specialty, geographic location, insurance mix, academic vs. community-based, etc.)
- Design a protocol that maximizes chances of site and patient study participation
  - Minimal I/E criteria that determine subject eligibility
  - Minimizes data collection burden on the site
  - Fits well into practice flow and patient management

# What Do You Want to Measure?

- **Choose the exposure and outcome data variables** needed to answer your observational study research question(s) in your accessible population
  - In observational studies, there are typically multiple exposure variables and several outcome variables
- Depending on your research question(s), **determine whether the protocol will mandate sites to collect study-related data** or whether you will take a ‘hands off’ approach and collect data only if it has been collected as routine standard-of-care (i.e., a registry approach)
  - Decision may affect regulatory requirements and IRB/ethics committee decision timelines, especially in the EU

# What Do You Want to Measure?

- Determine the types of **data measurement scales** since they affect your choice of statistical analysis methodology that ultimately affect quality of the data
  - Categorical – limited to descriptive and simple analyses (e.g., counts, proportions, Chi-square, etc.)
    - Nominal – no ordering of positions (e.g., gender, race-ethnicity)
    - Ordinal – ordering of positions (e.g., high, medium, low)
  - Continuous or discrete
    - Values along an arithmetic scale (e.g., age, weight, income)
    - Statistically, the most powerful variables
      - **Ability to calculate descriptive and simple analytics as well as means, standard deviations, t-tests, ANOVA, and more powerful regression analyses**
    - Can always collapse the variables into categories after data collection, if desired

# Maximize Study Validity

- **Minimize random error** with an adequate sample size, reduce sampling error and variability in measurements
  
- **Minimize systematic error (bias)** and ensure data accuracy (i.e., data measures what it should be measuring):
  - **Subject selection bias**
    - When study sample does not represent population of interest
  - **Information bias**
    - Subject bias (e.g., recall bias, untruthful or exaggerated responses)
    - Observer or interviewer bias (e.g., knowledge of outcomes influences data collection)
    - Measurement bias
  - **Confounding**
    - Confounding variable is associated with the outcome and with the exposure but is not an intermediate step in the pathway between exposure and outcome

# Maximize Study Validity

## ■ Reduce subject selection bias

- Use proper sampling methodology
  - Nonprobability sampling (e.g., consecutive sampling)
    - **Requires screening log of all consecutive eligible patients who agree/decline to participate**
  - Probability sampling (requires entire list of site's eligible patient population from which to draw a random sample)
- Avoid low study participation rates, high non-response rates, elevated study drop-out rate
- Number of enrolled patients per site should be limited to avoid clustering of specific variables that may bias results

# Maximize Study Validity

## ■ Reduce information bias

- Blind interviewers/data collectors to subjects' exposure and outcome status when possible
- Provide standardized and comprehensive site training on the protocol and data collection methods
- Use standardized CRF endpoint data collection
  - Clear unambiguous defined variables; avoid free text
  - Minimize recall bias with memory aids if possible and use of validated questionnaires

# Maximize Study Validity

## ■ **Control known confounders**

### – In the design phase

- Restrict study eligibility criteria to exclude or limit those with known confounding variable

- **May limit generalizability of data**

- **May limit ability to acquire sufficient sample size**

- Employ matching of cases and controls on known confounding variable

- **Use sparingly only if analysis phase strategies are not good**

### – In the analysis phase

- Stratification
- Statistical adjustment through multivariate modeling

# CRFs and Patient Diaries

- Collect only what is needed per protocol
  - “Need to have” versus “nice to have” data; prevent data creep
    - Eg. Are physical exam and lab results required in CRF, or only reported as AE if criteria met during routine practice?
  - Don’t mistake this to mean that detail isn’t important. May require significant detail (eg. daily diary of symptoms; complete dosing information for costing)
- Standardize CRFs Standardized interviewer training and endpoint definitions
  - Provide clear and comprehensive CRF completion guidelines
  - Avoid choice of free text and endpoints that have recall bias
  - Consider confounding variables
- Pilot test the CRFs when possible by some sites in various countries
- Keep statistician involved at every stage, including CRF development, data cleaning plan and prioritization strategy (identifying critical data), reviewing data periodically
  - SAP (statistical analysis plan) should be developed before start of the analysis that includes process to handle missing data or implausible data

# What is the Timing of Data Collection?

- Can you collect the data at **one time point** to answer your research question(s)?
  - Cross-sectional design
  - Retrospective cohort design
  - Case-control design
- Do you need to collect data across **multiple time points** in order to answer your research question(s)?
  - Prospective cohort design

# Is Your Observational Study Design Feasible?

- Does the design meet the objectives of your research question(s)?
- Can your design meet your timeline/budget expectations?
- Ensure estimated sample size is feasible given the following:
  - Specific inclusion/exclusion criteria
  - Recruitment timelines
  - Sampling methodology, including need for stratification of cohorts
  - Availability and willingness of patients to participate (if applicable)
  - Expected drop out and survey response rates (if applicable)
- Do you have the right investigators selected in order to meet study recruitment?
- Do you have the appropriate operational resources (i.e., project management, data management, site management, technology) to efficiently conduct the study design you have selected?

# Selecting Study Designs: Methodological and Practical Considerations

Methodological & Practical Considerations	Pragmatic trial randomized	Prospective observational cohort design	Registries and other prosp NIS	Retrospective Database/chart obs studies
Comparative data required for alternative txs	++	+	+	+
Study results require high level of int validity	++	+	-	-
Need to measure small difference in outcomes	++	++	+	-
Need long-term data/results	-	+	++	++
Study results need to be generalizable	+	+	++	++
Need to identify natural trends over time	-	-	++	++
Randomization is an issue because of ethical considerations	-	++	++	++
Outcomes of interest need to be collected directly from patients	++	++	++	-
There are no sufficient historical data to conduct retrospective analysis	++	++	++	-
The budget is limited and it is not feasible to conduct a resource-intensive study	-	-	+	++
The study results are needed quickly	-	-	-	++

Adapted from Michael Drummond

# Observational Study Protocol Elements (1):

- Motivation for performing the study
- Scientific objectives; formulation of one (or more) precise questions and working hypotheses
- Study type
- Selection procedure for treating physicians
- Selection and recruiting process of study participants
- Study population, in- and exclusion criteria
- Sample size justification
- Definition and determination of primary variables
- Measures for assuring data protection and ethical principles
- Definition of evaluation parameters and description of their relevance
- Possible confounding factors and description of measures to detect and control them
- Possible bias

Theobald K & GEP & AGES & ENCEPP

# Observational Study Protocol Elements (2):

- Data collection procedure
- Duration of evaluation
- Concept of archiving
- Data processing and collation
- Description of systems and procedures for collection and documentations of AEs
- Coding process of ADR/SAEs and concomitant drugs
- Timeline
- Description of measures for quality assurance
- Description of the statistical methods
- Descriptions of responsibilities
- Reporting procedures including statistical and medical evaluation and plans for publication

# Site Management in Observational Studies

Andrea Spannheimer  
Vice President, Global Late Phase Research,  
Innovus

# Customized Site Training & Support

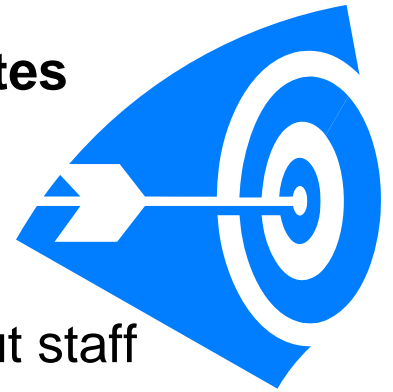


**Professional research sites** (specialized in clinical trials; often recruit external patients into studies)

**Research-experienced practicing sites**  
(Recruit patients from within practices, have research infrastructure in place)

**Research-aware sites**  
(have performed studies in past, but staff is put together on an ad-hoc basis)

**Research-naive sites**  
(no study experience)



- Important to understand the research experience and staff levels of your sites
- Practicing sites often don't have much research experience
- It's important to customize the site training and level of support needed to conduct a high-quality study

# Customized Site Training & Support

- Keep process simple (especially for research-aware and research-naive sites)
- Decrease study's operational burden on the sites
- Site initiation and training
  - 1:1 f-2-f initiation/training
  - 1:1 telephone initiation/training
  - Via investigator meeting
  - Via Webex sessions

# Monitoring

- Monitoring is an important component of quality assurance in the conduct of studies
  - Ensure patient safety
  - Ensure data quality
- Traditionally („gold standard“) for RCTs every 4-8 weeks on site, 100% SDV
- Observational studies
  - Large numbers of sites and patients often require a more cost-effective, targeted monitoring approach
  - Often hybrid monitoring is performed
    - A blend of on-site monitoring and remote monitoring (i.e., monitoring via telephone/email, etc.)
    - Can either be done on a random sampling of a pre-defined percentage of patients or sites or via a triggered approach

# Key Considerations for Determining Monitoring Strategies



**The balance point to be considered at the project level include:**

- Therapeutic Area
- Study Endpoints
- Study Type
- Stage of Development
- Study Complexity
- Study Size
- Regulatory Considerations
- Purpose of the study

**\*ICH/GCP section 5.18.3 – The Extent and Nature of Clinical Monitoring**

# Monitoring

- **Remote / central monitoring:** Monitoring via telephone/email etc.
  
- **Hybrid monitoring:**
  - A blend of on-site monitoring and remote monitoring
  - Can either be done on a random sampling of pre-define % of patients or sites or via triggered approach
  
- **Triggered monitoring:**
  - Pre-defined events trigger a monitoring visit. (e.g. data quality, subject enrollment, safety signals)
  - CTTI Survey of current Monitoring Practice: Factors likely to trigger a site monitoring visit
    - Number of protocol deviations
    - Suspected fraud
    - Rate of enrollment
    - Missing CRFs
    - Incidence of adverse events

# Adaptive Monitoring

- Basis Monitoring

Risk classification of the trial determines frequency of visits and percentage of patients and items for SDV

**Plus**

- Trial specific monitoring components

provisions to avoid or minimize specific risks derived from a thorough analysis of the trial

- Detailed guidelines listing possible „critical points“ of a study with respect to Safety, Rights and confidentiality, Error-prone procedures, Critical interfaces, Sources of variance, Sources of bias
- Provide corresponding lists of suitable provisions to prevent the specific problems identified

**Plus**

- For cause monitoring

On site visits schedules in case of irregularities detected by close central monitoring

# Source Data Verification (SDV)

- Definition of a randomly selected % of sites or patients with 100% SDV
- Definition of key data that need to be source data verified for all or a certain % of patients
- Important: quality input. Define upfront in the monitoring plan what happens if monitoring shows low data quality: e.g.
  - Additional training
  - SDV % increased
- Data that are recommended to be verified:
  - Consent
  - Eligibility criteria
  - Primary and secondary objective parameters
  - Safety data

# Data Collection & Management Phase

Andrea Spannheimer

# Data Management

- Offer a data collection/management system suited to the needs of the sites and patients
  - EDC, paper, or hybrid
  - May accommodate multiple data sources (physician data via EDC, questionnaires via paper, third party data source)
  - Simple & intuitive to accommodate research naïve or resource limited sites & Easy to learn
- Review data early and often to identify trends and understand the data being collected
  - Use monitors / Site Management Centre to provide specific training in problem areas
- Targeted data cleaning, plausibility checks, queries

# Missing Data

- Use strategies for data that can't be captured later if missed, or will not be found in the chart. For example:
  - Quality of Life questionnaire: trigger for sites to review for completeness before patient leaves; automated email reminders for patients to enter directly in EDC system
  - Health economic data / Health resource utilization: provide memory aid for patient to track resource use since last visit (eg. # visits to physiotherapy, time loss from work, cost of transportation, etc.)

# Analysis

- SAP (statistical analysis plan) should be developed before start of the analysis. It should include also handling of missing data or implausible data
- Use appropriate methods for observational studies

# Reporting & Publication Phase

# Reporting & Publication Phase

- Develop a publication and communication plan early in the process
  - Publish a timely summary of the results after the completion of the study
  - At least a short report of the results should be distributed to the participants
- Study report should be traceable, transparent and objective
- Follow STROBE statement for a checklist of items that should be addressed in articles reporting on cohort, case-control and cross-sectional observational studies
  - 22 items essential for good reporting of observational studies:
    - Article's title (item 1)
    - The introduction (item 2 and 3)
    - Methods (item 4-12)
    - Results (item 13-17)
    - Discussion (item 18-21)
    - Other information (22)
  - 18 items are common to all 3 designs, 4 items (6,12,14, 15) are design-specific

Von Elm e, Altman DG, Egger M, et al (2007) PocockThe Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. PLoS Med 4 (10): e296.

# Quality Assurance & Recommendations in Observational Studies

# Quality Assurance in Non-interventional Studies

- Employ SOPs specifically designed for managing the planning, operational, and evaluation processes involved in non-interventional studies (NIS);
- Keep the study staff/project management team apprised of legal and regulatory requirements and recommendations for NIS;
- Ensure that the protocol and study-related management processes do not interfere with the non-interventional premise of NIS;
- Implement a quality plan that describes which quality control and quality assurance activities must be conducted.

Theobals K, Capan M, Herbold M, Schinzel S, Hundt F: Quality assurance in non-interventional studies. *GMS Ger Med Sci* 2009; 7: Doc 29.

# Relevant Guidance Documents / Guidelines Recommendations / Codes (1)

- International society for Pharmacoepidemiology (ISPE). Guidelines for Good Pharmacoepidemiology Practice (GPP) [issued 1996, revised August 2004 & April 2007]. [https://www.pharmacoepi.org/resources/guidelines\\_08027.cfm](https://www.pharmacoepi.org/resources/guidelines_08027.cfm)
- International Epidemiological Association (IEA). Good Epidemiological Practice (GEP) – IEA Guidelines for proper conduct of epidemiological research. November 2007
- Arbeitsgruppe Epidemiologische Methoden der Deutschen Arbeitsgemeinschaft für Epidemiologie (DAE). Leitlinien und Empfehlungen zur Sicherung von Guter Epidemiologischer Praxis (GEP). April 2004.  
[www.gmds.de/publikationen/1b\\_leitlinienundempfehlungen\\_April2004.pdf](http://www.gmds.de/publikationen/1b_leitlinienundempfehlungen_April2004.pdf)
- Rules for non-interventional studies and financial support for healthcare quality registries. ILF Policy 2010:1 October 2010. SID3 (16)
- EFPIA Code of practice
- DHHS Guidance for Industry: Good PV Practice & Pharmacoepidemiologic assessment (March 2005)
- AHRQ: Users guide to Registries for Evaluating Patient Outcomes
- European Network of Centers for Pharmacoepidemiology and Pharmacovigilance (ENCePP): Code of Conduct Revision 1 (2010): [www.encepp.eu](http://www.encepp.eu)
- ISPOR: taskforce prospective observational clinical studies Good research.  
[www.ispor.org](http://www.ispor.org)

# Relevant Guidance Documents / Guidelines Recommendations / Codes (2)

- Verband Forschender Arzneimittelhersteller e.V: (VFA). VFA-Empfehlungen zur Verbesserung der Qualität und Transparenz von nicht-interventionellen Studien [Januar 2007] <http://infomed.mds-ev.de/sindbad.nsf/9341fb3978976712c1256ed600524d6b/f99eec1f4951d5c6c12572c800335eee?OpenDocument>
- Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM); Paul-Ehrlich-Institut (PEI). Gemeinsame Empfehlungen des BfArM und PEI zur Planung, Durchführung und Auswertung von Anwendungsbeobachtungen 2010.  
[http://www.pei.de/nr\\_154718/DE/infos/pu/02-klinische-pruefung/klin-pruef-awb/klin-pruef-awb-node.html?\\_nnn=true](http://www.pei.de/nr_154718/DE/infos/pu/02-klinische-pruefung/klin-pruef-awb/klin-pruef-awb-node.html?_nnn=true)
- BfArM: recommendations for planning and conduct of AWB, November 1998
- ABPI: Code of Practice for Pharmaceutical Industry 2001; the association of British Pharmaceutical Companies
- AGES: Scientific Guidance for the Conduct of Non-interventional Studies (NIS) in Austria [http://www.basg.at/uploads/tx\\_basginfobox/L\\_Z61\\_Guidance\\_NIS\\_AT.pdf](http://www.basg.at/uploads/tx_basginfobox/L_Z61_Guidance_NIS_AT.pdf)
- Sickmüller B, Breitkopf S. Points to Consider zu Anwendungsbeobachtungen. Empfehlungen des Bundesverbands der Pharmazeutischen Industrie zur Durchführung von Anwendungsbeobachtungen. Pharm Ind. 2009; 71(5): 764-9.

# Summary of Guidelines and Recommendations

- Study with scientific purpose and a written protocol/study plan with a SAP
- Contract between sponsor and healthcare professional
- Patient information and Informed Consent
- Data protection laws & rules must be followed
- Ethics Committee review
- Information about planned observational study in public accessible registries
- Planning, supervision, analysis, quality assurance and budget responsibility should belong to Medical /Scientific department of the pharmaceutical company
- Reasonable remuneration in relation to the amount of work
- No inducement to prescribe sponsor's product
- Appropriate SOPs and training of relevant staff
- Process for quality assurance
- Publication of results
- Archiving

# Summary

- Design a scientifically valid study design/objectives with simplicity and feasibility of success
  - Ensure a clear delineation of roles and responsibilities between marketing/sales and medical affairs
  - Ensure all involved parties understand the difference between observational study design/conduct and clinical trial design/conduct
  - Focus on collecting essential data and avoid fishing expeditions
- Use appropriate support tools and materials (CRFs; monitoring, data management, statistical analysis) that reflect specific needs and limitations of observational studies
  - Clear and unambiguous CRFs and patient diaries
  - Appropriate site monitoring strategy
  - SAP that accounts for missing data, adjusts for confounders, and employs observational study analytic methodologies
  - Targeted data cleaning process
- Develop a customized QA plan
- Set-up the publication plan very early in the process and ensure that design and operational strategies account for important publication timelines/dates and objectives
- Keep it simple !

## Contact information:

U.S./Canada: 866-306-1321

International: 612-234-9015

Email: [engage@innovus.com](mailto:engage@innovus.com)

