QUESTIONS FOR OBSERVATIONAL STUDY DESIGN

COMMUNITY OF PRACTICE FOR STATISTICS
OBSERVATIONAL WORK GROUP: Jon R. Sobus, Leader (NERL), Najwa Haykal-Coates (NHEERL), James L. Crooks (NHEERL), Andrew J. R. Gillespie (NERL), Janet A. Nestlerode (NHEERL), Alan Vette (ACE), Timothy J. Wade (NHEERL)

August 6, 2014

These questions and criteria are focused on study design and best practices to enhance reproducibility of science and have been developed for use in study planning activities. The intended audience includes scientists, branch chiefs, and quality assurance managers for guidance in preparing and reviewing QAPPs and research protocols.

1. **Study Overview**
    a. What is the research problem?
    b. What is (are) the study objective(s)?
    c. What are the research questions that will be answered to address the objective(s)?
    d. What is the population of interest, and how does the study of this population address the research questions?

2. **Study Design and Data Elements**
    a. What type of study is planned (e.g., pilot study, exploratory study, or confirmatory study)?
    b. What is the study design (e.g., cross-sectional, case-control, time-course, prospective)?
    c. Does the study propose the use of existing data?
    If Yes
        i. What is the source of existing data?
        ii. What was the purpose and scope of the original study?
        iii. How does the proposed use of the existing data deviate from the purpose and scope of the original study?
        iv. What is the justification for a new analysis of the existing data?
    d. What is the unit of analysis (i.e., population sample)?
        i. What is the sampling method (e.g., convenience, random, stratified, cluster)?
        ii. How will the sampling method be considered in the statistical analysis (e.g., sample weighting)?
        iii. Can statistical results based on the unit of analysis be extrapolated to the population of interest?
    e. What are the units of observation (i.e., measurement variables)?
        i. What primary variables (i.e., the key dependent variables and independent variables) are included in this study?
        ii. What covariates will be examined?
        iii. What is the variable type (e.g., continuous, categorical) and measurement unit (e.g., milligrams per liter) for each primary variable and covariate?
        iv. What are the methods for evaluating and transforming data distributions?

3. **Statistical Analyses**
    a. Is a statistical analysis plan included?
    b. What statistical analyses will be performed?
        i. What statistical parameters will be estimated (e.g., mean, standard deviation, regression coefficient)?
        ii. What statistical relationships (between measurement variables) will be quantified?

iii. What statistical hypotheses will be tested and what hypothesis tests will be performed?
iv. What are the assumptions behind the statistical analyses (e.g., normal distribution)?
v. What software and procedures (e.g., SAS/Proc GLM, R-Chemometrics) will be used?
vi. How will statistical models be constructed and evaluated, and what is the method for selecting between competing models?
vii. How will uncertainty be quantified and reported?
viii. How will the results of statistical analyses be interpreted with respect to the research questions?
ix. How will the results of statistical analyses be presented (i.e., types of equations, tables and figures)?

4. Sample Size

a. Will a power analysis be performed?

If Yes

i. How will the power analysis be performed?
ii. What are the assumptions for the power analysis and why are they necessary?
iii. What is considered acceptable power for the study?
iv. What is the pre-specified effect size and why was this chosen?
v. How does the sample size support the power level and effect size?
vi. For which variables will repeated measurements be collected?
vii. How will the number of repeated measures and the correlation among them be taken into account in the power analysis?

If No

i. Why is a power analysis not applicable to the proposed study?

5. Bias

a. What are the potential sources of bias (e.g., sampling bias, measurement bias)?
   i. How will bias be identified?
   ii. What approaches will be used to minimize or eliminate this bias?
b. What will be done to avoid reporting bias and publication bias?

6. Data Inclusion/Exclusion

a. What are data inclusion and data exclusion criteria?
b. What are the proposed methods for detecting and treating outliers and spurious values?

7. Data Analysis Considerations

a. How will missing data be coded in the data set and treated in statistical analyses?
b. How will covariates be evaluated as confounders or effect modifiers/interactions?
c. How will values below a detection limit/quantitation limit be represented?
d. Are negative measurement data possible (e.g., as a result of background correction if applicable) and how will these data be used?

8. Multiple Comparisons

a. Will multiple test correction (i.e., an adjustment for multiple comparisons) be performed?
   i. If yes, what method will be used?
   ii. If no, what is the explanation for not performing a multiple test correction?
b. Will subsets of the population sample be extracted for further testing? If yes, which subsets and for what purpose?

9. Data Management and Sharing

a. How will the data be archived (e.g., laboratory notebooks, digital backups)?
b. For data stored in digital-only format, what steps will be taken to ensure that the data are accessible using non-proprietary or open-source software?
c. What database will be used and why?

Additional Resources (note that material in Wikipedia is not peer-reviewed)

Keywords	Links
confirmatory study	http://en.wikipedia.org/wiki/Confirmatory_trial
confounders	http://en.wikipedia.org/wiki/Confounding http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704-EP713_Confounding-EM/BS704-EP713_Confounding-EM_print.html
covariates	http://en.wikipedia.org/wiki/Covariate
data inclusion/ exclusion	http://en.wikipedia.org/wiki/Inclusion_and_exclusion_criteria
dependent variables	http://en.wikipedia.org/wiki/Dependent_and_independent_variables
detection limit	http://en.wikipedia.org/wiki/Detection_limit
distributions	http://en.wikipedia.org/wiki/Probability_distribution http://en.wikipedia.org/wiki/Normal_distribution http://en.wikipedia.org/wiki/Log-normal_distribution
effect modifiers	http://www.ams.ac.ir/AIM/NEWPUB/12/15/9/0014.pdf
effect size	http://en.wikipedia.org/wiki/Effect_size
exploratory study	http://en.wikipedia.org/wiki/Exploratory_research
hypothesis tests	http://en.wikipedia.org/wiki/Statistical_hypothesis_testing
independent variables	http://en.wikipedia.org/wiki/Dependent_and_independent_variables
measurement bias	http://en.wikipedia.org/wiki/Information_bias_(epidemiology) http://pubs.rsc.org/en/content/articlepdf/1997/an/a704789d
measurement unit	http://en.wikipedia.org/wiki/Units_of_measurement
missing data	http://en.wikipedia.org/wiki/Missing_data http://www.stat.columbia.edu/~gelman/arm/missing.pdf
multiple comparisons	http://en.wikipedia.org/wiki/Multiple_comparisons_problem
multiple test correction	http://en.wikipedia.org/wiki/Multiple_comparisons_problem http://udel.edu/~mcdonald/statmultcomp.html http://en.wikipedia.org/wiki/False_discovery_rate
outliers	http://en.wikipedia.org/wiki/Outlier
pilot study	http://en.wikipedia.org/wiki/Pilot_experiment http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3081994/
population	http://en.wikipedia.org/wiki/Statistical_population
power	http://www.ma.utexas.edu/users/mks/statmistakes/power.html http://udel.edu/~mcdonald/statsampsize.html http://www.stat.columbia.edu/~gelman/stuff_for_blog/chap20.pdf

power analysis	http://www.ats.ucla.edu/stat/seminars/Intro_power/default.htm http://www.ats.ucla.edu/stat/gpower/indepsamps.htm http://www.gpower.hhu.de/en.html http://www.ats.ucla.edu/stat/dae/
publication bias	http://www.ma.utexas.edu/users/mks/statmistakes/filedrawer.html
quantitation limit	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2556583/
reporting bias	http://en.wikipedia.org/wiki/Reporting_bias
sample size	http://en.wikipedia.org/wiki/Sample_size http://en.wikipedia.org/wiki/Statistical_power#Software_for_Power_and_Sample_Size_Calculations
sampling bias	http://en.wikipedia.org/wiki/Biased_sample
sampling method	http://en.wikipedia.org/wiki/Sampling_(statistics)
statistical analyses	http://en.wikipedia.org/wiki/Statistics
statistical hypotheses	http://en.wikipedia.org/wiki/Statistical_hypothesis_testing
statistical models	http://en.wikipedia.org/wiki/Statistical_model
statistical parameters	http://en.wikipedia.org/wiki/Statistical_parameter
statistical relationships	http://en.wikipedia.org/wiki/Correlation_and_dependence
study design	http://en.wikipedia.org/wiki/Study_design
subsets	http://en.wikipedia.org/wiki/Statistical_population
uncertainty	http://webpages.ursinus.edu/lriley/ref/unc/unc.html
unit of analysis	http://en.wikipedia.org/wiki/Unit_of_analysis
unit of observation	http://en.wikipedia.org/wiki/Unit_of_observation
variable type	http://en.wikipedia.org/wiki/Statistical_data_type
weighting	http://www.applied-survey-methods.com/weight.html