SECURITY

COUNTERINTELLIGENCE

OFFICE OF THE NATIONAL COUNTERINTELLIGENCE EXECUTIVE

Guidelines for Personnel Security Research

Recommendations from the IC Behavioral and Social Sciences Research Group

March 2012

This report was principally authored by representatives of:

Central Intelligence Agency (CIA)

Defense Personnel Security Research Center (PERSEREC)

Office of the National Counterintelligence Executive, Special Security Directorate (ONCIX/SSD)

Report endorsed by the IC Behavioral & Social Sciences Research Group[1]

whose members support the following agencies:

Air Force Office of Special Investigations (AFOSI)

Central Intelligence Agency (CIA)

Defense Intelligence Agency (DIA)

Defense Personnel Security Research Center (PERSEREC)

Department of Defense (DoD)

Department of Energy (DoE)

Department of Homeland Security (DHS)

Department of State (DoS)

Federal Bureau of Investigations (FBI)

National Center for Credibility Assessment (NCCA)

Office of the National Counterintelligence Executive (ONCIX)

National Geospatial-Intelligence Agency (NGA)

National Reconnaissance Office (NRO)

National Security Agency (NSA)

Naval Criminal Investigative Service (NCIS)

United States Army

---

[1] The IC Behavioral & Social Sciences Research Group (BSSRG) is a group whose goal is to provide a structured setting for dissemination, consultation, collaboration, and continuing education for behavioral and social scientists tasked with the discovery or application of national security research. The BSSRG is comprised of individuals who are employed as either civilians or contractors throughout the IC. Members of the BSSRG represent themselves as behavioral scientists; their views do not necessarily represent the agencies they support.

Table of Contents

Executive Summary

Personnel security research is conducted to inform decisions about personnel security processes or policies. Sound research findings improve policies and processes and also provide policymakers with a defensible evidence base from which to justify their risk-management decisions. Decisionmakers who use research findings also must be able, however, to verify that the research findings they use were derived from scientifically sound approaches and to understand any limitations of the research findings that may reduce their applicability to the issue(s) at hand. Researchers who support policymakers must also have sufficient information about research studies to interpret the findings and, in some instances, to replicate the research effort.

While U.S. Government agencies strive to conduct sound personnel security research, the degree to which this research adheres to scientific principles of study design, procedures, statistical analysis, and reporting often varies. Although some of this inconsistency is likely a result of the unavoidable challenges inherent in conducting applied research, it also may be a result of the lack of awareness of standard scientific practices of those conducting the research, tasking the effort, or consuming the products of the research effort.

This document, authored by behavioral scientists, was created to serve as a reference for both researchers and policymakers. It provides guidance on scientific approaches to personnel security research, to include information on study design, sampling, data collection, statistical analysis, and the reporting of research findings. This document should be used as a reference by those undertaking research efforts in personnel security and by those seeking to use personnel security research findings to inform policy or process decisions.

Introduction

Personnel security research is conducted by agencies throughout the Federal Government to establish an evidence base that can be used to inform risk management and resource allocation decisions. Research findings augment the collective historical experience of policymakers by providing scientifically defensible evidence to support their decisionmaking. While researchers and consumers of research are often tempted to seek out findings that support their pre-existing notions, soundly conducted research provides objective information that advances our understanding of important issues, either by supporting existing notions or by providing a new perspective on long-held assumptions or difficult policy decisions.

To maximize the value of research findings, research efforts should strive to adhere to accepted scientific approaches in study design, procedures, statistical analysis, and reporting. Unfortunately, this is often not the case. Considerable variability currently exists in the degree to which personnel security research adheres to scientific principles. Research in applied settings faces a number of challenges that often hinder a researcher's ability to adhere strictly to these principles. Those who conduct and consume research should understand and attempt to incorporate accepted practices. Even when research is conducted with the best of intentions, minor inconsistencies in acceptable research methods can undermine the value of that research for use in policy decisions.

This document outlines best practices in personnel security research. It was authored by a working group of research professionals[2] familiar with personnel security research and is designed to serve as a reference for both researchers and policymakers. The document outlines considerations for researchers who conduct studies, as well as considerations for those reading reports documenting research efforts. While the document is targeted for a lay-audience unfamiliar with the nuances of scientific research design and procedures, an appendix to the document provides more detailed information on technical research considerations and examples of potential resources.

Research Design

The design of personnel security research efforts is dictated by the objective of the study, the data available to researchers, and often by the resources available. While different types of studies will necessitate different methodological designs, the first step in any research effort should be the selection of design and analytic approaches that will maximize the value of the study results. This section addresses major design considerations for any study.

Research Goals

The first step in undertaking a research effort is to identify the goal or goals of the study. Each goal should represent a broad statement of what stakeholders should expect to take away at the end of the study. For example, a study designed to determine the value of a particular element of personnel

[2]The authors were a subgroup of the larger IC Behavioral & Social Sciences Research Group (BSSRG). The BSSRG is a group whose goal is to provide a structured setting for dissemination, consultation, collaboration, and continuing education for behavioral and social scientists tasked with the discovery or application of national security research. The BSSRG is comprised of individuals who are employed as either civilians or contractors throughout the IC. Members of the BSSRG represent themselves as behavioral scientists; their views do not necessarily represent the agencies they support.

security investigations might read as follows: *The goal of this study is to determine the value of Element X to the personnel security investigation process for single scope background investigations.* Notice that the goal includes an evaluative component (value), an object of analysis (Element X) and a defined population (single scope background investigations).

While this goal may accurately represent what the researcher is hoping to accomplish, it is vague in terms of how it will be accomplished. Therefore, it may be necessary to further refine the goal by defining additional objectives that address each of the goal's understandable parts. Following on the previous example, objectives for this goal might be: (1) *Determine the efficiency of Element X in single scope background investigations; and (2) Determine the quality of Element X in single scope background investigations.*

In this example "efficiency" and "quality" represent two dimensions of the evaluative component, but both represent terms that must be specifically defined in terms of analysis. Otherwise, the reader is left wondering what efficiency and quality really mean. Research questions may be required when a research objective is too vague to clearly define how the study will actually meet the overall goal. In our example, the following research questions could apply to Objective 1:

- *What is the average time, in days, required to complete Element X?*
- *What is average cost to complete Element X?*

Each research question details specific measures that, when considered together, are intended to achieve the stated research goal. One important element of research questions is that variables (e.g., efficiency) should be "operationalized;" the variables should be defined in terms of how they will be assessed so that the customer or consumer can clearly see how the data will be analyzed to answer each research question, meet each objective, and realize each goal.

Once the research goal has been identified, where possible the researchers may state hypotheses or predictions that will be tested in the study. Generating a hypothesis further clarifies the aim of the research, communicates the assumptions of the researchers, and guides the approach to statistical analysis.

Sampling Strategy

Once the research question has been selected and refined, the researcher must address the issue of sampling. Ideally, the researchers would evaluate all persons or cases of interest to identify the "ground truth" behind every research question. Unfortunately, this is almost always impossible in personnel security research. The IC has well over 100,000 cleared individuals, and examining each of the cases manually would require more time and resources than are available. Therefore, we select a sample, a subset of a population of interest, to evaluate and draw conclusions about the larger population. The most important element of any sample is that it be truly representative of the population. For example, our results may be skewed if we were to examine the incidence of criminal behavior in a cleared population that is 60 percent male and 40 percent female, and we use a sample that is 75 percent male and 25 percent female. The results may reveal more criminal behavior within the sample than is actually present in the larger population.

Researchers use a host of common sampling techniques, and the type of sampling used depends very much on the nature of the study. The following are short descriptions of the sampling strategies most commonly used in personnel security research:

- *Random sampling* is used most frequently and simply means that every member of the population under study has an equal chance of selection.
- *Stratified sampling* is used when the researcher wants to ensure that a certain segment of the population is represented. The research team divides the population into subgroups and randomly selects from those strata.
- *Convenience sampling* is just what the name implies—a population that is readily available to the research team—as in the man/woman on the street.

Ethics in Research with Human Participants

Researchers are responsible for the ethical conduct of their research. When conducting personnel security research with human participants, researchers must implement strategies to reduce any potential harm to participants and to data examined from participants. This includes: obtaining informed consent, anonymizing and de-identifying data, and ensuring data security, integrity and confidentiality. Research also must abide by laws that govern research involving human subjects.[3,4] As such, prior to the initiation of a study, researchers may be required to submit research protocols through the approval process of an Institutional Review Board (IRB) – an ethics review committee designated to review research involving humans in order to protect the rights and welfare of the research participants. In some cases, personnel security research may be exempt from an IRB review. Exemptions may include research involving the collection of existing data when those data were collected in a way that subjects cannot be identified. Research also may be exempted from an IRB review if it is conducted to assess the performance of a program where the results of the evaluation are for official government use only and are not intended for generalized use beyond that program.

Prior to the initiation of a research project, researchers should consult with their department's or agency's IRB, if available, to determine whether the effort is subject to IRB oversight and approval[5], as well as consult with their Offices of General Counsel, and civil liberties, civil rights, and privacy officials to ensure other appropriate safeguards are incorporated into all phases of the research effort.

Sample Size

Determining the sample size needed to adequately represent the population of interest is one of the most common sampling challenges. Researchers should avoid using sample sizes that are

---

[3] The Federal Policy for the Protection of Human Subjects or the "Common Rule" was published in 1991 and codified in separate regulations by 15 Federal departments and agencies (see Title 45, Code of Federal Regulations, Part 46)

[4] Department of Defense. (2011). *Protection of Human Subjects and Adherence to Ethical Standards in DoD-Supported Research* (DoD Instruction No. 3216.02). Retrieved from http://www.dtic.mil/whs/directives/corres/pdf/321602p.pdf. See also the Department of Health and Human Services guidelines on the protection of human subjects, Title 45, Part 46 (45 CFR Sections 46.101-46.123 and 46.304-46.305), as well as Agency regulations that reference these guidelines.

[5] Several federal departments and agencies have additional regulations in place for research involving special populations or for general human subject research. The federal department/agency that conducts or supports research retains final authority for determining whether an institution has complied with its regulations for the protection of human subjects.

needlessly large. No absolute rules exist for determining sample size, and these decisions must be made based on the nature of the study, methodology employed, efficient use of resources, risks to participants, and characteristics of the sample. Two major considerations are common in personnel security research: confidence and statistical power.

Confidence Level and Confidence Interval. Samples must be selected that provide the best representation of the population as possible within the time and resource constraints of a study. The confidence interval is one measure of how well the results represent the actual population. The ***confidence interval*** represents the range of values in which the population-level results are expected to fall. It is represented as a "plus-or-minus" value; the larger the confidence interval, the greater the margin of error we are willing to accept in our results. For example: If a study said 25 percent of background investigations for a Secret clearance have at least one actionable issue and the confidence interval is +/-5%, we would say that between 20 percent (minus 5 percent) and 30 percent (plus 5 percent) of background investigations for a Secret clearance should be expected to have at least one actionable issue.

In a related vein, the ***confidence level*** determines how sure one can be that results will fall within the confidence interval. Based on the previous example, a confidence level of 95 percent would mean that we can be 95 percent sure that the percentage of Secret-level background investigations with at least one actionable issue is between 20 percent and 30 percent.

Sample sizes corresponding to specific levels of confidence will vary based on population size and desired confidence interval. Generally, the larger the sample, the smaller the confidence interval; however, samples over 400 subjects will yield acceptable confidence intervals for most populations.[6]

Table 1. Confidence Levels and Intervals Corresponding to Sample Sizes, by Population Size

	Sample Size Required for:			
Confidence Level	99%		95%	
Confidence Interval	+/- 3%	+/- 5%	+/- 3%	+/- 5%
Population Size				
500	394	286	341	217
1,000	649	400	516	278
5,000	1,350	588	880	357
10,000	1,561	624	964	370
50,000	1,783	657	1,045	381
100,000	1,851	661	1,056	383

Bartlett, J.E., Kotrlik, J.W., & Higgins, C.C. (2001). Organizational research: Determining appropriate sample size in survey research. *Information Technology, Learning and Performance Journal, 19*(1), 43-50.

[6] While a general rule applicable to many research questions common to personnel security, researchers should be mindful of the frequency in which observations of interest occur in the population. For example, if a researcher is interested in studying factors related to espionage, which occurs *infrequently*, a random sample of the population would likely not contain enough instances to conduct statistical analysis or draw meaningful conclusions.

For most samples, confidence intervals and levels can be quickly calculated using existing software programs, and multiple websites also offer easy-to-use sample size and confidence calculators. For more information, see Appendix A.

Statistical Power. Statistical "power" is the extent to which a researcher is able to detect statistically meaningful differences between groups. Power is influenced by a number of factors, one of the most important being the size of the sample. If examining differences between groups is an essential aim of a research effort, the research should perform a "power analysis" to estimate the minimum sample size. If sample size is too low, the experiment will lack the precision to provide reliable answers to the questions it is investigating. If sample size is too large, the research effort may yield statistically significant differences, but the magnitude of these differences may be small. As with confidence interval calculations, power analyses can quickly be calculated using existing software programs or easy-to-use online calculators. To use these tools, researchers will, at a minimum, need to enter the effect size and the rate at which they are willing to obtain false negatives. Researchers should be conservative in estimating the effect size and the rate at which they are willing to obtain false negatives when this information is unknown. For more information, see Appendix A.

Data Collection Methods

Sample size and confidence determinations may vary based on the data collection method used. There are two primary types of data, qualitative and quantitative. ***Qualitative methods*** *are* usually used to collect data on ill-defined constructs that are not conducive to traditional numerical assessment. Qualitative data is often collected by observation and is recorded in narrative form. ***Quantitative methods***, on the other hand, are used to collect numerical data using methods by which the same structured data is collected, to the extent possible, across all observations. Quantitative methods almost always depend on random sampling and allow for statistical analyses to make generalizations about larger populations. Appendix A presents additional details on the different types of qualitative and quantitative data collection methods.

Data Coding

Personnel security research relies heavily on quantitative data collection methods that involve reviews of personnel files or investigations. Therefore, systematic coding procedures are used to quantify information of interest. Common data coding tools include observation sheets, questionnaires, surveys, and file review forms. These tools enable raw information to be recorded systematically and make it subject to statistical analysis. Some information of interest is largely objective, such as age, gender, and marital status; other information is more subjective, such as the extent of drug use.

Researchers attempt to develop clear, consistent, and objective coding procedures that reduce the subjectivity of coder judgments. Researchers often document these procedures in coding manuals that provide instruction for coders and serve as a reference throughout the coding and data analysis processes.

It is important to ensure that coding is conducted in a consistent manner across coders. Therefore, researchers often employ strategies to ensure inter-rater reliability (IRR). IRR is a measure of the degree to which coders agree with one another when assigning ratings or values to information of interest. IRR may refer to the degree to which coders agree with each other, or to the degree to which coders' ratings agree with an established expert rating. Regardless of the approach chosen, researchers should assess IRR throughout the coding process to ensure that coders are coding data using the same standard. (See Appendix A for additional information on ensuring quality and IRR in coding.)

Statistical Analysis Methods

Descriptive Statistics

Descriptive statistics are used to summarize, organize, and simplify collected data. They are used to describe the characteristics of the sample and include demographics (e.g., age, gender), averages, and other elements that are used to summarize the data in the sample. The research team should select appropriate descriptive statistics to describe the characteristics of the sample and the data collected. The following are some of the most common descriptive statistics found in quality personnel security research products.

Inter-Rater Reliability (IRR). As stated above, IRR is most important when analysts are making judgments about factors that might be subject to error. IRR is commonly evaluated by having coders code the same set of data and then calculating the agreement across raters. (See Appendix A for additional information on IRR.)

Central Tendency. Central tendency refers to a way in which data is described in terms of a single score that defines the "most typical" value across all scores in the sample (the distribution). Three measures of central tendency commonly used in personnel security research are the mean, median, and mode.

The ***mean*** is the sum of scores divided by the number of scores—or the average score. The mean is the preferred measure of central tendency because it uses every score in a sample in its calculation and therefore provides a good representation of all scores.

The ***median*** is the score that divides the distribution exactly in half; it is the score that falls directly in the middle of all scores. While the mean and median are often close, if not the same, there are times when they may be very different, and the median may be the preferred measure of central tendency. Examples are when: the sample contains extreme scores that may inflate or deflate the mean; the sample has undetermined values (e.g., null values) making it impossible to calculate a mean for each individual in the sample; a distribution has no upper- or lower-limit, rendering it impossible to place a numeric value on one option; and ordered data is described where the degree of difference between each ranking is unknown.

The ***mode*** is the third measure of central tendency and is the score or category that occurs most often in a distribution. The mode is most often used to describe the shape of a distribution, when the data are categorical (e.g., type of investigation, gender) and when results cannot logically be represented by a fraction.

Variability. Variability is the degree to which scores in a distribution are spread out or clustered together. The ***standard deviation*** is a measure of variability that indicates the standard distance that scores in a distribution (or set of scores) fall from the mean. The standard deviation tells whether scores tend to be clustered closely to the mean or distributed all around it. The larger the standard deviation, the greater the average distance between scores in a sample. Variability allows us to obtain an objective measure of how close together or far apart the scores are in any distribution. One of the primary purposes of measuring variability is that it measures how well an individual score represents the entire distribution or how much error we can expect across a set of scores.

The ***range*** is the difference between the largest and smallest scores in a distribution. Although range provides a description of the lower and upper limits of scores in a sample, it is highly dependent on the extreme values in the distribution and may not provide an accurate illustration of how spread-out the scores actually are.

Distribution. Another type of commonly used descriptive statistic is the ***frequency distribution***. Frequency distributions are simply organized tabulations of the number of individuals or scores in each category of interest. Examples of frequency distributions may be the number of males compared to females in a sample or the number of times a specific behavior or characteristics is observed across participants.

Cross-tabulation is a type of frequency distribution where data are categorized against two or more factors. An example might be the types of investigation conducted (A, B, or C) by investigative service provider (X, Y, or Z), laid out in a 3 by 3 table.

Inferential Statistics

Inferential techniques allow us to study samples and then draw inferences about an entire population. Inferential statistics rely on the concept of statistical significance to determine, for instance, if the difference between two or more groups or between two or more effects of a manipulation are statistically significant. That is, are the results different than what we could expect to see by chance? Commonly used inferential statistics include t-tests, analysis of variance (ANOVAs), correlations, and Chi-square tests. (See Appendix A for a more detailed discussion.) It is incumbent upon the researcher to select the tests that are both most appropriate to answer the questions and that can be appropriately applied to the data available. It is important to note that many of these techniques are valid only if the sample data meets certain assumptions.

Reporting Research Results

Research in the personnel security domain is most often conducted to influence personnel security processes or policies. Researchers often are not able to provide consumers with in-depth presentations on a particular research effort, and it is not always feasible to consult with researchers during the policymaking process. Therefore, the report describing a research effort and its findings is critical. The ultimate product of the research effort is often the research report; therefore, it is the resource that will likely be referenced in the decision-making process. Because of the significance of the research report, it is incumbent upon the authors to document the study clearly and thoroughly and to present the findings candidly.

The research report also serves to ensure the replicability of the research effort. While senior leaders may be hesitant to allocate resources for studies that replicate existing studies, the replication of research efforts provides consumers with greater confidence in evidence-based decisionmaking. A replicated research effort should produce the same or very similar outcomes, even when conducted by different research teams. When efforts at replication do not produce similar outcomes, they often identify factors or variables that likely account for differences and provide consumers with a more nuanced understanding of the issue. Replication also acts as a safeguard against careless, fraudulent, or dishonest research. For all of these reasons, the research report should document the research effort in enough detail so that it can be replicated by other researchers.

The following provides research reporting guidelines that are intended to (1) ensure that researchers address all relevant study information in research reports, and (2) ensure that consumers of any research report (e.g., policymakers, senior executives, and other researchers) are provided sufficient information to understand and use the results of the study.

Report Format & Structure

Behavioral scientists engaged in personnel security research tend to use the report format that is recommended by the American Psychological Association (APA) and detailed in the APA publication manual, although there is no single "correct" research report format. Table 2 below shows the elements of a report using APA format (2010, pp. 41-59). Whatever format is used, however, it is critical that a report documents all relevant aspects of the study. An advantage to using a prescriptive scientific research report format like the APA's is that it ensures that a report addresses the critical issues. This paper outlines many of the same issues highlighted by the APA format, including aspects of research methods, sampling, replicability, research goals or hypotheses, sample selection, methodological approach, data analysis, and data interpretation.

Table 2. Common Elements of an APA-Formatted Research Report

Title
Author's Name & Institutional Affiliation
Abstract[7]
Introduction
Method
Results
Discussion
References
Footnotes
Appendices & Supplemental Materials

Background & Research Goals

When available, previous research findings or policies relevant to the study should be discussed in terms of how they relate or contribute to the current study. Referencing these resources provides

[7] An abstract is a brief, comprehensive summary of the contents of scholarly publications. In government reports, however, the abstract is often substituted by an executive summary. Although similar to abstracts in that they both summarize the contents of the report, executive summaries are written in non-technical terms and contain more comprehensive information, to include a brief statement of the problem addressed, background information, key findings, and main conclusions.

context for the reader to better appreciate the current understanding of a particular area, enables the reader to better interpret the findings of the current study, and provides references for the reader should they want to further expand their knowledge.

Personnel security research is primarily conducted to answer specific questions. As previously stated, the research report should clearly state the goals and any research hypotheses identified at the beginning of the effort. Even in exploratory research efforts aimed at providing baseline or descriptive information about particular issues rather than answering specific questions, the report should specifically state the objective of the effort and should highlight for the reader what new information the study aims to provide.

Reporting on the Research Procedure

The report should summarize the steps taken in the research protocol. This includes: (1) how participants were recruited or how data were sampled; (2) the development and testing of coding systems or other data collection tools, such as questionnaires or surveys; (3) the steps taken to gather data, including instructions provided to participants; and (4) any other steps in the execution of the research effort. The report should provide enough information that the reader could reasonably replicate the study independently.

Reporting on Results

Describing the Sample. The results section of the research report should first describe the sample from which the data were drawn. This information helps the reader understand how similar or dissimilar the sample studied is to the larger population of interest. This, in turn, helps the reader to judge the ability to generalize the findings and to compare findings across similar studies. When describing the sample, the author(s) should, at a minimum, report the major demographic characteristics of the sample, such as age, gender, and race/ethnicity. When relevant to the research question at hand, the report may also describe subgroups or other variables like national origin, level of education, language proficiency, etc.

Finally, researchers sometimes intentionally remove data from their dataset because of data collection errors, technical difficulties, or other reasons that suggest that the data are not valid. Researchers exercise discretion in deleting data and do so only in cases where there is clear justification. When data are deleted, the researcher includes in the report the nature of the data deleted and the process for determining to do so.

Reporting Statistics. After describing the sample, the results section should summarize the data collected (descriptive statistics) and provide the results of any statistical tests used (inferential statistics). The author(s) should report the results in enough detail to allow the reader to understand the data from which conclusions and recommendations are drawn. When describing statistical tests, the report should describe the analysis conducted and provide information that allows the reader to understand the meaning of the reported statistics. Depending on the analytic strategy, this may include group or subgroup sample sizes, frequencies (e.g., percents), measures of central tendency (e.g., mean), measures of variability (e.g., standard deviation), confidence intervals, the type of statistical tests used and corresponding test statistics (e.g., t, chi-square), regression coefficients (e.g., r), effect sizes (e.g., d) and/or statistical significance (e.g., p values).

It is the responsibility of the author(s) to report all relevant results, even those that are not consistent with the hypotheses of the researchers, as well as results that are inconsistent with existing or proposed policies or procedures. Researchers should aid in the interpretation of these data and should perform due diligence by addressing any multiple interpretations that might be possible.

Reporting Limitations

All research efforts have limitations. Studies that employ great scientific rigor to provide insights into the workings of a particular area may be limited in terms of their general application to the operational environment. Conversely, applied research efforts conducted in an operational environment may account for real-world variations but can lack the rigor necessary to draw firm conclusions. The study author(s) should document major limitations in the research report, whether they relate to resource constraints, difficult timelines, the sample available, the data collection strategy employed, the analysis performed, or any other research issues addressed in this document. By clearly documenting these limitations, readers are better equipped to determine their confidence in the findings and the extent to which the findings are applicable to the policy or process issue at hand.

Reporting Recommendations

When backed by sound methodology, research findings provide valuable information to the security policy-making process. With this audience in mind, the research report should highlight the findings most pertinent to these considerations. It is incumbent upon the author(s) of the report, however, to remain objective and to refrain from speaking beyond the data available. In other words, the report should point to the findings most relevant to policy decisions but must not suggest a particular course of action unless the findings clearly suggest one.

Conclusion

Use of standard research methods and practices ensures the validity of research intended to inform personnel security policy. Personnel security policy must be grounded in sound rationale and based on scientifically defensible research. The risks of basing policy on inconsistent or inadequately documented research include the following:

- Time and resources are wasted conducting research and establishing a policy that may need to be reissued.
- Policy decisions cannot be defended in legal proceedings.
- The efficiency, effectiveness, or fairness of the personnel security system is inadvertently reduced.

To avoid these risks, researchers should adhere to best practices for research methods in any research intended to inform a policy decision. The hallmarks of such research include:

- The goals, objectives, and research questions addressed by the study are clearly defined.
- The methods used in any study are documented in sufficient detail to enable replication of the study by a third party.

- A description of the population being studied, demographic characteristics of the sample, and the sample selection process are described in detail.
- The data collection method is explained, and (when applicable) a copy of the coding instrument or survey questionnaire used in the study is provided.
- Documentation includes the operational definitions of every data element collected and a listing of all of the possible values.
- Data quality control procedures used to assess the quality of data used in the analysis are reported.
- The analyses conducted are appropriate to the research objectives of the study.
- Any limitations of the study design or execution and the potential impacts on the validity of the study findings are documented.

Use of the research principles described in this document will help ensure that quality research products are developed to support policy decisions. While this guidance is not intended as a substitute for formal education in research practices, it should inform the oversight and conduct of research throughout the personnel security community.

Appendix A

The research guidelines described in this document were written for a lay-audience unfamiliar with the nuances of scientific research design and procedures. This appendix was included for the reader interested in more detailed information on technical research considerations. It parallels several sections previously referred to in this document and includes additional information on specific considerations or strategies pertaining to research design, data collection methods, and statistical analysis.

While providing more detailed information, the appendix does not sufficiently capture every consideration involved in personnel security research. Researchers or policymakers should reference readily available resources in behavioral science research methods and design for more information. A brief list of potential resources is provided at the end of this section.[8]

Research Design

Data Coding

Data coding may be improved through certain quality control measures. Specifically, it may be necessary, particularly when coders are not thoroughly familiar with the data or with a coding tool, to pretest and post-test a defined coding scenario. For instance, let us assume that coders are reviewing security files for specific information and that we have constructed a coding sheet that instructs them as to what type of information to extract. Before beginning actual data collection, it is wise to have participants code a small number of cases (e.g., 10) and provide the information to the research team. The research team will learn from that exercise and can revise the coding protocol and/or its associated instructions accordingly. Similarly, after completing a small number (e.g., 10-20) of "live" case reviews, the coding protocol should be revisited again so that adjustments can be made before a large amount of data is collected. It is critical at this point of the exercise to include multiple coders (from different agencies, if applicable) to maximize the benefits of the review.

Sample Size

Confidence Level and Confidence Interval. For most samples, confidence intervals and levels can be quickly calculated using existing software programs, and multiple websites offer easy to use sample size and confidence calculators. The following website is an example of a potential resource: http://www.gifted.uconn.edu/siegle/research/Samples/samplecalculator.htm.

Statistical Power. As with confidence interval calculations, power analyses can be quickly calculated using existing software programs or easy to use online calculators, such as the power analysis program, G*Power, at http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/.

---

[8] The resources at the end of this section, as well as the references to links to online applications or calculators, should not be construed as an endorsement by the Federal Government; they are merely examples of the tools and resources available to researchers conducting behavioral science research.

Data Collection Methods

Qualitative Data Collection

Observation. Observation is a type of qualitative data collection that involves watching a behavior, action, or process occur in its natural environment and creating detailed records or descriptions of what is observed. An example in the personnel security arena might be observing personnel security interviews to determine how subjects respond to different types of interviewer personalities and recording the details about each subject's body language or responses. Observation is used when a researcher is trying to learn as much as possible about a topic.

Interview. Unstructured interviews are used when a researcher has some level of understanding about a topic but is attempting to gain additional information that may or may not corroborate what is already known. Interviews generally start with a standard set of questions, but are not limited in the questions that may be asked or in the answers that may be provided.

Focus Groups. Focus groups are used when a group of individuals has shared knowledge about a topic that is of interest to the researcher. Like unstructured interviews, focus groups may be open-ended to allow conversations to flow naturally while following a predetermined format. Focus groups may also be used when a researcher is curious to know the level of consensus among a certain group of subject matter experts.

Literature Review. A literature review is a critical evaluation of previous research concerning a particular area of interest. It includes a review of scholarly articles, books, and other relevant sources, and then provides a summary of significant literature published on that topic. The goal of a literature review is to bring the reader up to date with current literature on a topic providing a description, summary, and critical evaluation of available information.

Quantitative Data Collection

Observation. Observation using quantitative methods is different from qualitative observation. Here, observation is based on recording specific data according to a well-defined data collection instrument (e.g., coding process), in numerical form, and in exactly the same manner across all subjects. Observation may be used when the researcher is interested in the number of times a variable occurs in an observable situation. Examples may be the number of recruits who visit recruiting stations in a specific time frame, or the number of times an investigation is closed before all investigative elements are complete.

Survey. Surveys or questionnaires are used when a researcher is interested in knowing a specific population's beliefs, attitudes, or behaviors based on predefined scales of measurement. Surveys present respondents with closed-ended questions and ask them to choose among pre-defined outcomes (e.g., To what extent do you agree…? How important is…?). Surveys may also include open-ended questions about objective measures (e.g., How many times do you use Website Z in a week?).

Meta-Analysis. Meta-analysis combines the results of several studies that address a set of related research questions. The researcher combines the results for a given finding from two or more studies and then employs statistical analyses to compare across them. The aim of meta-analysis is to more powerfully estimate the statistical relationship between variables than what could otherwise be derived in a single study.

Data Extraction. Data extraction refers to the practice of using existing data stored in a company or agency's existing data system. Data extraction is used when the data necessary to answer a research question are known to be previously recorded, stored, and available for evaluation. Examples of data extraction may be the cost data pertaining to various elements of a security investigation, time to completion for different investigative elements, or the gender and ages of clearance holders.

Experimentation. An experiment is a controlled study in which the researcher attempts to understand cause-and-effect relationships. The study is "controlled" in the sense that the researcher controls how participants are assigned to groups and which intervention(s) are applied to each group.

Experimentation relies on two different types of variables: (1) independent variables, or the condition or manipulation that is applied; and (2) dependent variables, or the factors that are expected to be influenced by the independent variables. To further control the study, researchers may randomly assign individuals to groups. The researcher can eliminate the influence of other variables by using this "randomization" process and by ensuring that the variable in question is just as likely to affect one group as another. This way any difference observed between groups can be reasonably attributed to the influence of the variable being tested.

As an example of experimentation, consider an analyst interested in knowing if the use of automated systems affects the quality of security investigations as rated by adjudicators. The independent variable is the application of automated systems; one group would be investigated using automated systems and one group without. A sample of investigations could be randomly assigned to one of the two groups. Then, after running the investigation, each investigation would be ranked by adjudicators in terms of quality (the dependent variable). In the analysis phase, the researcher compares group quality scores and attempts to determine whether the use of automated systems affects quality ratings for each group.

Statistical Analysis Methods

Whether conducting analyses of research or reading a research report, it is important for both researchers and consumers of research to understand the meaning of particular test statistics, degrees of freedom, effect sizes, p-values, main effects and interactions, and common post-hoc tests. For further information on these, see the resources provided at the end of this report.

Inferential Statistics

Independent Samples t-Tests. Independent samples t-tests are used to evaluate two sets of data coming from two completely separate samples to determine, for instance, if there is a real difference between two populations or two conditions. Two examples of research questions that might be answered by independent samples t-tests are:

- Is the rate of marijuana use in applicants under 25-years-old different from applicants over 25-years-old?
- For reinvestigations, is the number of issues identified for subjects undergoing routine (every five or 10 years) reinvestigation different from the number of issues identified for applicants subject to random interval reinvestigation?

Independent samples t-tests may also be used to test the directionality of effects where an analyst wants to know what type of difference exists in addition to whether a difference exists. For example, we may expect that the number of issues identified for applicants subject to random reinvestigation will be lower than for those subject to routine reinvestigation.

Assumptions of the independent samples t-test are as follows: (1) observations within each sample must be independent, i.e., not related to one another in any way; (2) the two populations from which the samples are drawn must be normally distributed; and (3) the two populations from which the samples are drawn must have *homogeneity of variance*— meaning that the variability of the two populations is comparable.

Related Samples t-Tests. The goal of the related samples t-test is to see if there is a meaningful difference between measurements at two different times (for instance, before and after a manipulation) for the same person.

The first type of related samples t-test is the ***repeated measures test***. The repeated measures test relies on two sets of data from the same sample, collected at two different points in time, usually prior to and after a manipulation. For example, we may want to know how changes to clearance reporting requirements affect a service member's willingness to seek mental health counseling and, specifically, the number of clearance applicants who report seeking mental health counseling when it was required reporting on the SF-86 (Time 1) and the number of those who seek mental health counseling after it is no longer required to report some types of counseling on the SF-86 (Time 2). A single sample of individuals is measured at Time 1 and the same sample at Time 2.

These tests are best used in the following circumstances: (1) when relatively few subjects are available, because repeated measures requires fewer participants; (2) when studying how an individual changes over time as a result of a experimental condition; and (3) when you want to eliminate problems caused by individual differences or characteristics that vary from person to person.

Repeated measures tests do have some disadvantages. First, *carryover effects* may occur when a subject's response at Time 2 has been influenced by the response at Time 1. In our example, subjects who "grew up" in a culture where seeking help could lead to adverse consequences may be less reluctant to seek help throughout their careers. Additionally, a *progressive error* may occur when an individual's performance changes consistently over time. In our example, time in service and greater knowledge of procedure may affect one's likelihood to seek mental health counseling; the change in attitude, if any, may not be because of the actual change in policy.

A second type of related samples t-tests is the ***matched subjects test***. In the matched subjects test, two sets of data are drawn from separate yet practically identical samples, where each subject in Sample 1 (who is asked to report on willingness to seek mental health counseling prior to policy change) has a match in Sample 2 (who is asked to report on willingness to seek mental health counseling after the policy change). In our example, respondents in Sample 1 might be matched to respondents in Sample 2 based on age, rank, gender, and time in service.

The advantages of matched subjects' tests are that the time required to complete data collection may be greatly reduced by using historical data or by collecting data from both set of subjects at one time, and carryover effects and progressive error may be reduced.

The assumptions of both types of related samples t-test are as follows: (1) observations within each condition must be independent of one another; and (2) the population from which the sample is drawn must be normally distributed.

Analyses of Variance (ANOVA). ANOVA is a hypothesis testing procedure used to evaluate the mean differences between two or more conditions. The primary goal of ANOVA is to determine the different effects two or more conditions or manipulations might have on the variable of interest. Overall, we want to determine if the differences observed between samples provide sufficient evidence to tell if the manipulations have significantly different effects.

The main advantage of the ANOVA over *t* tests is that the ANOVA can be used when there are more than two conditions or manipulations to compare. In fact, multiple factor ANOVA may be used to examine more complex, realistic situations in which more than one factor (independent variable) is acting on the variable we are measuring (the dependent variable).

For example, ANOVA would be the appropriate test if we are interested in knowing how the potential repercussions for falsification during the clearance process, combined with how well recruiters explain those repercussions, affect the number of instances of falsification during security processing. The following table represents six possible combinations of conditions in this case:

	Repercussion for falsification:		
	No repercussion	Civil Penalty	Criminal Penalty
Explained by Recruiter	A	B	C
Not Explained by Recruiter	D	E	F

Results of the ANOVA would help us determine under what conditions applicants are most and least likely to provide false information during clearance processing.

Assumptions for the (multi-factor) ANOVA are as follows: (1) The observations within each sample are independent; (2) the populations from which each sample is selected must be normal; and (3) the populations from which each sample is selected must have equal variances.

Correlation. Correlation is a statistical technique used to measure and describe the difference between two variables. Correlations are presented numerically, with values ranging from 0.0 (no correlation) to 1.0 (perfect correlation). Results also indicate the direction of the relationship between two factors. A positive relationship means that as one factor increases in value, the other factor increases in value; a negative relationship means that as one factor increases in value, the other decreases in value.

Correlations have several uses. First, they may be used to predict how one variable will change based on changes in another variable. For example, correlation between age and the number of adverse issues on a credit report might indicate whether credit issues tend to increase or decrease with age. Next, they may be used to validate new procedures or to determine how well a new measure correlates with a known, substantiated measure. For example, if the number of criminal conduct issues identified through in-person law record checks is strongly and positively correlated with the number of automated, web-based law record checks, we have an indication that the automated checks perform similarly to the in-person checks. Finally, correlation may be used for theory verification, or to test predicted relationships between two variables, such as whether the number years of investigative experience is related to time required to complete a report of investigation.

Care should be taken to remember that correlation does not imply causation; correlation only indicates whether two things are related or not. Just because two variables show a strong correlation does not mean that one factor causes the other.

The Chi-Square Test. Chi-square tests are considered non-parametric statistical tests. Unlike other analyses described, these tests make no assumptions about the population of interest and may be used in research situations where variables are not represented in terms of numeric data (e.g., gender, investigation type, etc.). Two types of chi-square tests are described below.

Chi-Square Goodness-of-Fit Test. The goal of the chi-square goodness-of-fit test is to see how well-collected data fit expected proportions. This test may be used when we want to know about proportions or relative frequencies in a distribution. For example, how does the number of NACLC applicants compare to the number of TS/SCI applicants in the US Air Force? Do people tend to prefer paper-based or online form completion?

Individuals are classified into categories, and the sample data is used to test pre-conceived ideas about the shape or proportions of a population distribution to determine how well the proportions match those ideas.

Chi-Square Test for Independence. Chi-square may also be used to test whether there is a relationship between two variables (e.g., Is employing agency related to foreign travel?). The goal is to evaluate whether the two variables are consistently and predictably related, particularly when data cannot be classified numerically (e.g., employing agency). Results from our example may answer whether or not knowing one's service affiliation or employing agency could help determine how often he or she travels internationally.

The assumptions of chi-square tests are as follows: (1) The sample must be randomly selected from the population; and (2) the sample size must be large enough so that the expected count in each cell is greater than or equal to five.

Additional Resources

American Psychological Association. (2010). Publication manual of the American Psychological Association (6th ed.). Washington, DC: Author.

Cohen, J. (Ed.). (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Gravetter, F.J., & Forzano, L.B. (Eds.). (2011). *Research methods for the behavioral sciences* (4th ed.). Thousand Oaks, CA: Wadsworth, Cengage Learning.

Nolan, S. & Heinzen, T. (Eds.).(2011). *Statistics for the behavioral sciences* (2nd ed.). NY: Worth Publishers, Inc.