INSTITUTE FOR DEFENSE ANALYSES

Meta-Synthesis Framework to Understand Results across Defense Experiments, Studies, and Wargames

Yuna H. Wong, project lead
Darrin Denny, deputy project lead
Akar Bharadvaj
Burgess Laird
Carolyn A. Leonard
Neil Mithal
Robert Polk
Robert Stimets
Cameron Swann
Ana Venegas

June 2022

Approved for public release; distribution is unlimited.

IDA Paper P-33123
Log: H 22-000237

INSTITUTE FOR DEFENSE ANALYSES
730 East Glebe Road
Alexandria, Virginia 22305

About This Publication

This work was conducted by the Institute for Defense Analyses under contract HQ0034-14-D-0001, Project CA-8-4931, "Analytic support to J7 to Implement a Meta-Analysis Framework to assess elements from the Joint Warfighting Concept" for the Joint Staff J7. The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

Acknowledgements

The study team would like to thank Joseph Adam, Sue Numrich, Joel Resnick, Heather Wojton, and Kevin Woods for acting as reviewers on earlier drafts of this report. The team would further like to thank Eleonore Douglas, Tom Greenwood, Hank Miranda, Patrick Savage, Alec Wahlman, Eric Walters, and Gregory Zehner for additional contributions to the report.

For More Information

Yuna Wong, project leader, Joint Advanced Warfighting Division
ywong@ida.org | 703-845-6825

Darrin Denny, sub-task leader, Joint Advanced Warfighting Division
ddenny@ida.org

Daniel Y. Chiu, director, Joint Advanced Warfighting Division
dchiu@ida.org | 703-845-2439

INSTITUTE FOR DEFENSE ANALYSES

IDA Paper P-33123

Meta-Synthesis Framework to Understand Results across Defense Experiments, Studies, and Wargames

Yuna H. Wong, project lead
Darrin Denny, deputy project lead
Akar Bharadvaj
Burgess Laird
Carolyn A. Leonard
Neil Mithal
Robert Polk
Robert Stimets
Cameron Swann
Ana Venegas

Executive Summary

Purpose and Overview

The purpose of this study was to develop a framework to better enable a synthesis of results across experiments, studies, and wargames. The goal of the framework is to support judgments on when there is sufficient evidence to draw conclusions about concepts, approaches, and proposed solutions in joint domains.

To develop a meta-synthesis framework, IDA—

- Identified analytic needs,
- Reviewed a broad range of existing frameworks,
- Reviewed approaches to determining confidence levels,
- Adapted elements from various frameworks and approaches to confidence,
- Developed a framework, and
- Tested it.

Meta-Synthesis Framework

The five phases to the meta-synthesis framework are:

I. **Frame the question** – Includes actions such as documenting the original request for information, documenting changes to the question, defining terms, scoping research and analysis to be done, developing points of contact, identifying foundational documents, and assessing the expertise of the analyst using the framework. The analyst may choose a slightly shorter version of the framework that skips certain assessments.

II. **Select the data** – Involves identifying inclusion and exclusion criteria for the events that will be reviewed, identifying data sources, prioritizing report types, identifying search terms, conducing searches, and documenting search results.

III. **Assess the data** – Analysts assess reports from the experiments, studies, and wargames that come out of Phase II of the framework. This includes reviewing assumptions, scenarios, experimental conditions, major results, and other aspects.

IV. **Synthesize the data** – Analysts synthesize and summarize findings and uncertainties across event types.

V. **Communicate the results** – The last phase involves communicating results by documenting the original request for information, documenting the framework version used, characterizing the work that was done, documenting relevant results and conclusions (and the level of confidence), and offering supporting documentation.

IDA developed a step-by-step Analyst Notebook, presented in Appendix C, for analysts to use to document each step.

Existing Analytic Frameworks and Approaches to Confidence

IDA reviewed existing analytic frameworks that ranged across numerous fields to develop the meta-synthesis framework: systematic reviews and frameworks from medicine, statistics, data science, natural sciences, social sciences, organizational learning, intelligence analysis, and defense analysis. Casting a wide net enabled drawing on and assessing a variety of potentially useful ideas to could serve the sponsor's needs.

The sponsor also asked IDA to review approaches to measuring levels of confidence. The IDA team selected and reviewed the literature around statistical confidence and the levels of confidence developed by the intelligence community.

Knowledge Management and Exercising JExNET

IDA also had the opportunity to review JExNET and then to attempt to use JExNET contents to support the J7. The team found significant gaps in the information uploaded into JExNET. For example, less than 12% of events in JExNET were accompanied by a report, and fewer than 75% had a listed point of contact. It was also unclear what percentage of Department of Defense (DoD) events make it into JExNET at all—the research team identified well-known events lacking an entry or documentation.

IDA also delved into the database's content to support a J7 review of whether concept required capabilities (CRCs) were addressed in JExNET entries. The CRC descriptions and JExNET entries contents are classified, and IDA provided a classified briefing to the sponsor on the results of using JExNET to review CRCs. As it stands, JExNET does not contain enough content to help the joint community even with the relatively "easy" question of whether database entries may apply to individual concept required capabilities.

Conclusions & Recommendations

Conclusions

- There are significant challenges in synthesizing learning across DoD. Events are conducted for different purposes, by different organizations, with many unstated assumptions A meta-synthesis framework or other tool is essential for organizational learning to occur.
- Trade-offs between rigor and speed are unavoidable. The framework and accompanying Analyst Notebook are designed to be systematic and rigorous; however, rigor takes time and time pressures necessitate weakening rigor.

- Any database, archive or knowledge management system will only be as useful as the completeness of it's the information it contains. The amount of information *not* in JExNET significantly limits its ability to serve as a significant information source for future J7 queries or meta-syntheses.
- Automation will not solve problems that require judgment and interpretation. The complexity of military judgment necessary to synthesize disparate information in order to support joint understanding is highly unlikely to be a fit for machine learning.
- The core challenges that J7 faces are not knowledge management or assessing evidence to support a position. Instead, the challenges are the more difficult ones of supporting institutional memory (what has happened), organizational learning (what do we now know), and organizational sensemaking (what new patterns of behavior should the organization adopt in the face of new contexts). A meta-synthesis framework would only address the tip of such an iceberg.

Recommendations

The J7 should consider—

- Continuing to populate expected archives of information. This is foundational to any future meta-synthesis or organizational learning. Such an effort may require significant resources and cannot be assumed.
- Continuing to adapt the meta-synthesis framework presented in this report. The best version of the framework will be the one most fitted to J7's staff, organizational processes, and other idiosyncrasies. For best results, emphasize repeated tests and adaptations around a user, changing design to incorporate more and more of a user's ongoing feedback.
- How to balance rigor versus speed. The framework in this report was built around allowing for a rigorous meta-synthesis. However, J7 staff may find it too time- and labor-intensive to be practical for J7's timelines, and hence may want to consider further tradeoffs in rigor for speed.
- Examining J7's structure and resourcing for organizational learning. While it was beyond the scope of this project, it is in J7's interest to examine whether it is organized and equipped to support institutional memory, organizational learning, and organizational sensemaking.

Contents

1. Introduction

A. Purpose

The purpose of this study was to develop a framework for the joint experimentation and analytic community that would facilitate synthesizing results across experiments, studies, and wargames. The goal of the framework is to support judgments on when there is sufficient evidence to draw conclusions about concepts, approaches, and proposed solutions in joint domains.

The study's sponsor, the Joint Staff J7, asked the Institute for Defense Analyses (IDA) to select, adapt, or develop an analytic framework that could be applied to key elements from joint concepts and their development efforts, and then to apply it. J7 charged IDA to draw from a wide variety of fields to examine potentially useful frameworks and approaches to understanding levels of confidence. To this end, IDA identified the following objectives to develop an appropriate meta-synthesis framework.

1. Identify Joint Staff J7 analytic needs.
2. Review existing frameworks for their usefulness to Joint Staff J7's analytic needs.
3. Identify existing approaches to assigning confidence to how relevant a framework might be to Joint Staff J7's analytic needs;
4. Develop or adapt a meta-synthesis framework that incorporates relevant components of existing frameworks and approaches to assigning confidence.
5. Pilot the meta-synthesis framework on real-world experiment, study, and wargames reports and improve the framework.

B. Background

In any given year, the U.S. Department of Defense (DoD) engages in a tremendous number of activities—wargames, experiments, analyses, exercises, and other events—in an attempt to foster and gain insights, results, and recommendations about a vast set of topics. The sheer scale of information garnered during these efforts complicates and essentially hobbles drawing comparisons and summary judgments from them.

Anyone wanting to better understand whether a certain solution or approach is a good idea should look first to DoD's relevant and existing body of work; however, there are challenges. One is discovery. Again, the scale of activity, the large number of organizations involved, and the rotation of individuals across the enterprise present enormous difficulties to learning across the entire range of DoD's work. Two other challenges include the difficulty of comparing or summarizing across so many results, and whether something qualifies as either "joint" or a true look at "contested logistics."

DoD's experimentation, wargaming, and analysis community does not have a common meta-analytic or meta-synthesis framework to allow it to compare and aggregate the results of wargames, warfighting experiments, analyses, and studies that could inform Joint Force design recommendations.[1] Developing such a framework could help the community better synthesize the results across disparate types of events—though often with considerable variation in assumptions, wildly differing aims and start points or time frames, and divergent tools.

Significant challenges for people whose responsibility it is to conduct such synthesizing analysis at the joint level can include—

- Limited experience in either the analytic methods or the subjects they are asked to assess.
- High turnover of military personnel assigned to conduct analyses, which may negatively impact institutional memory.
- No single data source for analysts to review. Although JExNET can now pull from multiple databases, reviewable data are spread across the Department, and much of it remains difficult to access.
- The Services and other agencies play a leading role in shaping the development of future capabilities. Capability demonstrations and development are often conducted in separate Service lines of effort rather than as part of a Joint campaign. This can hamper analysts' understanding of the Joint context for individual Service-led events.
- The time horizons of these requests are often short resulting in an inability to fully analyze a specific capability and make quality recommendations.

Given the disparate nature of the data used to answer questions related to the status of future capabilities or technologies intended to improve the future Joint Force, there is no systematic protocol for how an analyst should approach the problem space and no way for leaders to engage with and assess the answer their analysts gave them. The meta-synthesis framework developed for this project gives analysts a logical, systematic, and transparent approach to analyses. It is intended to be primarily relevant to the analyst researching the development of specific capabilities or technologies, although it is potentially relevant to broader applications.

C. Approach

The main task—to develop an analytic framework for Joint Staff J7 analyses—was divided into the following sub-tasks.

1. Identify existing and future analytic needs.
2. Conduct a literature review cross a range of existing analytic frameworks.

[1] While the original project description called for developing a *meta-analysis* framework, *meta-synthesis* is a broader approach that better aligns with J7's objectives for a framework. *Meta-analysis* is more often associated with a specific approach in medical research.

3. Review existing approaches to determining confidence levels.
4. Assess applicability of existing frameworks, including individual components, and approaches to confidence levels to J7's analytic needs.
5. Adapt or develop a meta-synthesis framework.
6. Test the meta-synthesis framework.

The study's overall methods were literature review for assessing analytic frameworks, and the design thinking process for developing the meta-synthesis framework. This involved iterating drafts of the framework with the sponsor and testers at IDA who had not been involved in developing the framework. Results of some of the tests are also included in Appendix C.

1. Identify Analytic Needs

In order to identify the Joint Staff J7's existing and future analytic needs, IDA held in-person discussions with both J7 leadership and action officers. IDA also provided J7 with an *analyst questionnaire* to gain additional detail on J7's analyst needs (see Appendix A). The questionnaire attempts to produce insights into the demands on J7 analysts for information and the data sources available to them. Two de-identified analyst responses to the questionnaire are also included in Appendix A to provide a sense of the context in which J7 analysts may be using the meta-synthesis framework developed over the course of this project.

Another way that IDA identified analytic needs was by reviewing existing reports in the Joint Experimentation Network (JExNET) to try to help J7 determine to what extent the information in JExNET applied critical required capabilities and mission-level concepts.

One part of the of the research dealt with data sources. IDA identified shortfalls in event data submitted to JExNET and event reports when available, for the types of analyses J7 leadership wants to accomplish. (See Chapter 5.)

2. Review Existing Analytic Frameworks

IDA reviewed existing analytic frameworks ranging across a number of fields. The review included frameworks from medicine, statistics, data science, natural sciences, social sciences, organizational learning, intelligence analysis, and defense analysis. Although there are a wide variety of frameworks that address different aspects of synthesis or validity, no one framework is suitable for the disparate results from DoD activities.

The most useful frameworks included:

- The overall framing of systematic reviews, including from medicine;
- The way to focus on the initial question and criteria from meta-ethnography;
- Types of validity from social science;
- The cycle of research linking defense studies, wargames, and exercises;
- Verification, validation, and authorization from defense modeling and simulation.

3. Review Existing Approaches to Confidence Levels

IDA reviewed existing approaches to measuring uncertainty, specifically tests of statistical significance and confidence levels employed in intelligence analysis.

4. Framework Development

IDA then held a workshop to identify relevant best practices that could be applied to an analytic framework for J7. Each team member presented their literature review to the entire team. Given the Joint Staff's desire for a framework based on the medical research community's approach to meta-analyses, IDA then used PRISMA as a starting point.[2] When the nature of DoD data made the PRISMA paradigm untenable, IDA researchers nominated alternative framework components to create a framework suited for Joint Staff analytic needs. The results of this process were phased worksheets compiled into an "Analyst Notebook" for practical application of the meta-synthesis framework.

5. Internal Testing

While conducting a pilot program was called out in the project description, Joint Staff J7 later requested that IDA not pilot the framework given constraints on J7's time. Instead, IDA tested the framework internally. IDA analysts acted as proxies for Joint Staff J7 analysts as if they were responding to an internally developed request for information (RFI). The exercise produced feedback that the team used to refine the framework and a complete set of exemplar worksheets. The completed worksheets and companion narratives are also provided for J7 analysts as an example of applying the framework.

D. Report Organization

Chapter 1 of this report contains the study's purpose, background, and approach. Chapter 2 describes the meta-synthesis framework (with the supporting Analyst Notebook in Appendix C). Chapter 3 contains the results of the team's review of existing frameworks, and Chapter 4 presents information on existing approaches to assigning confidence. Chapter 5 discusses knowledge management implications that surfaced during the study, particularly around JExNET. Chapter 6 reviews IDA's experience with exercising the sponsor's existing approach to collecting information about concept required capabilities and mission-level concepts. Chapter 7 presents conclusions and recommendations.

Several appendices accompany and enrich this report—

- Appendix A contains the analyst questionnaire.

---

[2] Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). PRISMA, "PRISMA: Transparent Reporting of Systematic Reviews and Meta-Anlayses," PRISMA, 2021, undated.

- Appendix B contains a briefing to the sponsor at an early project milestone on frameworks the team had reviewed.
- Appendix C is the Analyst Notebook with the worksheets for implementing the meta-synthesis framework. It also contains the Status Report.
- Appendix D contains a Senior Leader Guide to the meta-synthesis framework.
- The remaining appendices contain QUOROM[3]/PRISMA best practices (Appendix E), References (Appendix F), Glossary (Appendix G), and Abbreviations (Appendix H).

---

3 Quality of Reporting of Meta-Analyses (QUOROM). David Moher et al., "Improving the Quality of Reports of Meta-Analyses of Randomized Controlled Trials: The QUOROM Statement," *Lancet* 354, no. 9193 (27 November 1999).

2. Meta-Synthesis Framework Overview

The meta-synthesis framework is intended to

- Be effective at all levels and across organizations,
- Be effective over different research timeframes,
- Offer leaders an assessment on the validity of answers,
- Provide aggregate confidence levels for conclusions,
- Be adaptable for a range of questions,
- Influence understanding to better optimize future campaigns of experimentation, and
- Work at different classification levels.

To help analysts apply this framework, IDA developed an Analyst Notebook with individual worksheets for each framework phase (Appendix C). The last section of this chapter discusses the use of the Analyst Notebook, drawing on IDA analysts' experience of testing the framework.

A. The Framework

The five phases to the meta-synthesis framework are

I. Frame the question
II. Select the data
III. Assess the data
IV. Synthesize the data
V. Communicate the results

The phases are presented as a flow chart in Figure 1 to depict the progression between phases.

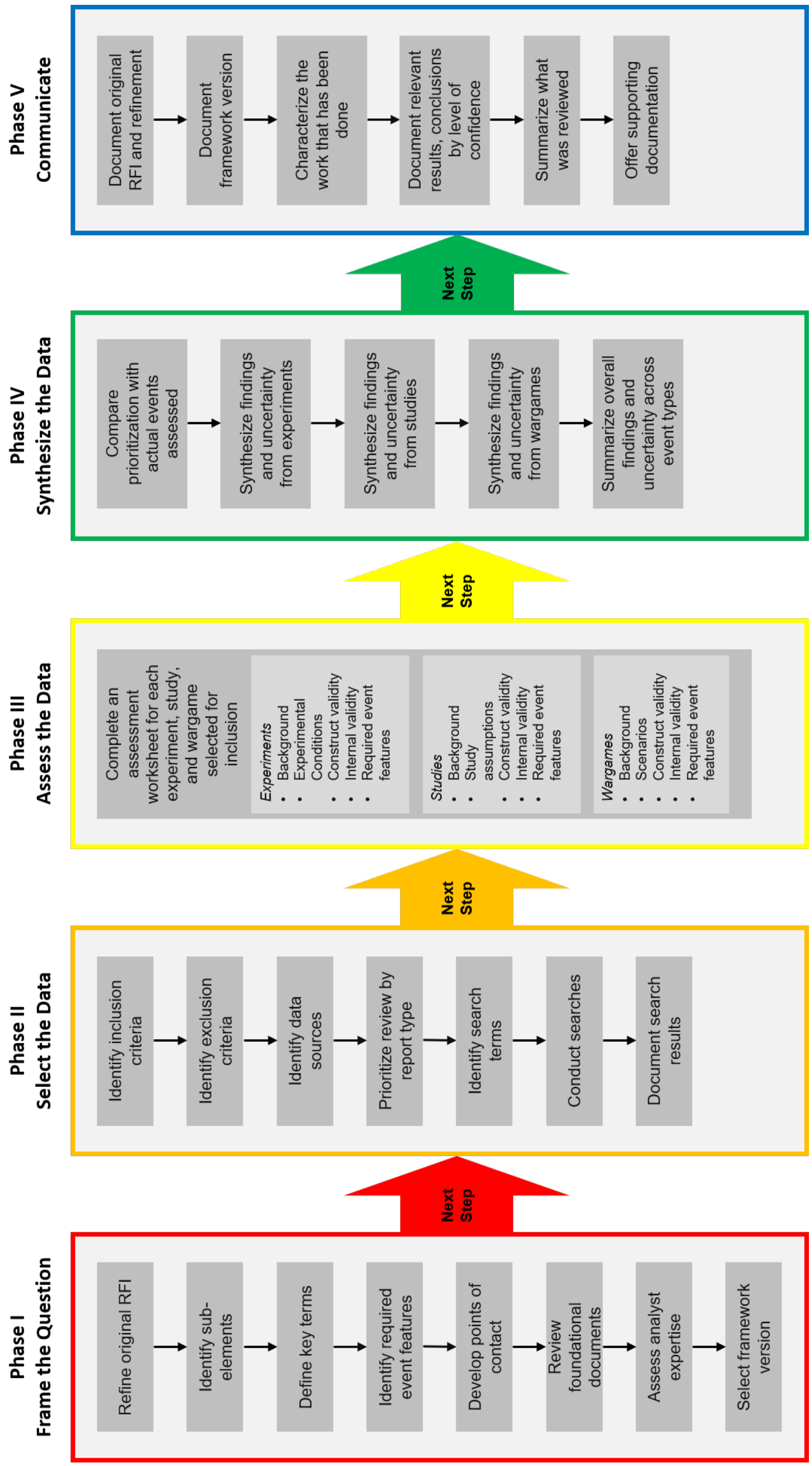

Figure 1. Meta-Synthesis Framework Flow Chart

1. Phase I – Frame the Question

Phase I addresses the exploration stage of traditional research methods and focuses on the research question and the literature review. In contrast to a hypothesis statement, which attempts to predict the relationship between two variables, a research question is a concise, focused, and debatable question that defines a research project's scope. Research questions are open-ended and best suited for exploratory research into novel areas where limited research already exists.[4] When a senior leader requests information, the request may not be in the form of a research question. Thus, the first step in this phase requires the analyst to refine the original request for information (RFI). The initial refinement could break a complex question into a set or series of questions or simply restate the question in simpler language.

Regardless of whether an RFI requires refining, all RFIs should be broken down into sub-elements that can be thought of as supporting lines of effort to conduct in order to answer the main question, for example: (1) characterizing work that has been done, (2) identifying what is well understood vice what is not, (3) determining the level of confidence in understanding, and (4) identifying aspects to investigate further.

Further scoping research and analysis occurs by defining ambiguous terms, thus presenting the senior leader and future users with a better understanding of concepts or factors being discussed. For example, if the RFI is asking about a capability to address a future adversary, how far in the future? Even if the requestor does not articulate a timeframe, by documenting the timeframe the analyst uses, others will know whether the answer applies to the question. Defining key terms ensures that future users of the research will know the analyst's and leader's perspective when the research was conducted, regardless of their understanding of the terms and concepts. Similarly, by documenting used definitions, analysts ensure the senior leader understands the scope and context of the research that underpinned the analyses.

The analyst should determine which event features are essential if the overall question is to be answered with certainty. *Required event features* include types of forces involved, warfighting functions required, geographic or functional commands, and other attributes. For example, an RFI that asks whether DoD has sufficient logistics capability in the event of a conflict on the Korean peninsula would make logistics a required function, and U.S. Indo-Pacific Command (INDOPACOM) a necessary geographic command. In other words, any relevant event would cover logistics and would use a scenario set in INDOPACOM. These features should be considered within the context of a single service, Joint, Combined, or adversary's force.

Developing points of contact (POC) within relevant offices is meant to help the analyst leverage the POC's knowledge about 1) the development of the concept, capability requirements, or technology under review; 2) individual experiments, studies, wargames; 3) campaigns

---

[4] Patricia Farrugia et al., "Practical Tips for Surgical Research: Research Questions, Hypotheses, and Objectives," *Canadian Journal of Surgery* 53, no. 4 (August 2010); NMBU, "Scientific Writing Portal: Creating Research Questions," Norwegian University of Life Sciences (NMBU), 2022.

of learning; and 4) functions, geography, and other contexts should be included. POCs may be used to identify documents and relevant events that the POCs consider foundational to understanding the topic. Such foundational documents could include relevant Joint concept documents, key guidance, or reports and results that figure prominently in informing DoD's understanding of the topic.

Foundational documents should be identified and reviewed as part of a light literature review in order to 1) survey the state of knowledge in an area of inquiry, 2) identify key authors, articles, theories, findings, and methodologies in an area of inquiry, and 3) identify gaps in existing knowledge.[5] The Phase I worksheet focuses on the first aim—it has the added benefit of establishing a baseline understanding regardless of an analyst's expertise.

Each analyst responding to the RFI should assess his or her background and expertise in each required feature that was identified earlier in the framework. Senior leaders can use this self-assessment when evaluating the recommendations and managing staffing. Finally, the analyst selects whether to use a complete or abbreviated framework.

2. Phase II – Select the Data

Phase II focuses on the data selection process and data collection methods employed. The former relates to determining appropriate data types, data sources, and data collection instruments that enable the analyst to adequately answer a research question. The latter relates to documenting the means of collecting, storing, and processing data.

As part of selecting data, analysts identify inclusion and exclusion criteria. Together they define the boundaries of a systematic review. Inclusion criteria define the specific conditions or characteristics that make a subject, or in the case of these worksheets, an event, appropriate to include in a research study.[6] They are similar to but distinct from required event features. Whereas required event features, if present in an event, increase an analyst's level of confidence in the findings and their applicability to the overall question; inclusion criteria are event criteria that must be present for an event to be included in the meta-synthesis. Inclusion criteria can also be used as a starting point when developing search terms. The left and right boundaries of inclusion are refined by exclusion criteria.

Exclusion criteria define the conditions or characteristics that make a subject or event inappropriate for including in a research study.[7] They are determined after setting the research

---

5 University of Northern Florida, "Library Guides: Conducting a Literature Review—Benefits of Conducting a Literature Review," University of Northern Florida, updated 13 July 2021, 2022.

6 Neil J. Salkind, "Inclusion Criteria," in *Encyclopedia of Research Design* (Thousand Oaks, CA: SAGE Publishing, 2010). Timothy Meline, "Selecting Studies for Systematic Review: Inclusion and Exclusion Criteria," *Contemporary Issues in Communication Science and Disorders* 33 (Spring 2006).

7 Neil J. Salkind, "Exclusion Criteria," in *Encyclopedia of Research Design* (Thousand Oaks, CA: SAGE Publishing, 2010). Meline, "Selecting Studies for Systematic Review."

question and before research begins.[8] Good validity and reliability of exclusion criteria help minimize random error and selection bias, thus improving the likelihood of finding associations and reducing the required sample size.[9] It may be necessary to define selection criteria used for particular concepts in the research question.

Once selection criteria are defined, the analyst should identify potential data sources. Which databases contain potentially relevant data? The analyst should consider databases across all classification levels. Which offices might have sponsored or know about potentially relevant events that have not shown up in the databases selected previously?

Ideally, an analyst will review all events resulting from applying inclusion and exclusion criteria; if time does not permit this, the analyst should review a random sample of the identified set that would provide an unbiased sample. Another alternative is to use purposive sampling, a form of non-probability sampling that relies on the analyst's judgment when selecting a sample set from a larger population. The analyst should also factor in how he or she chose to prioritize the review by report type. If the analyst chooses a *purposive sampling schema*, the logic behind the sampling should be documented.

The analyst should next develop a list of potential search terms. As the analyst begins searching on and identifying additional terms, he or she should document additional search terms. Once complete, the analyst should document the results of all searches and inquiries and then deduplicate the list of events, and apply the selection criteria to the full deduplicated list to identify which events to include in the meta-analysis. The final step is to record all included events and categorize them in terms of experiment, study, or wargame.

3. Phase III – Assess the Data

Phase III examines data suitability by assessing measures of validity—specifically in the form of construct, internal, and an adapted form of external validity. This phase also identifies gaps in the data. Assessing the data helps the analyst understand the type and quality of data that will be the inputs to the meta-synthesis. Because each event type has unique features, there are event type-specific worksheets to help the assessment. The analyst should complete an event-specific worksheet for each event in the meta-synthesis.

The three worksheets share several features. Specifically, they require the analyst to document general background information, constructs of interest employed in the event (such as "joint fires" or "contested logistics"), inputs (such as outputs from other events) to the event, and conclusions and takeaways. All three event-specific worksheets require the analyst to assess

[8] University of Melbourne, "Library Guides: Systematic Reviews—Inclusion and Exclusion Criteria," University of Melbourne, updated 18 October 2021, 2021, accessed 31 January, 2022.

[9] Salkind, "Exclusion Criteria."

each event against the *required event features* identified in Phase I and to assess overall applicability of the event to the overall question.

Each event-specific worksheet looks at three additional areas that are assessed in different ways depending on the type of event: (1) event parameters, (2) construct validity, and (3) internal validity. Event parameters—experimental conditions, study assumptions, and wargames scenario—are a proxy for external validity and provide insight into the level of continuity across events and ensures the analyst compares "apples to apples."[10] Given the nature of DoD events, analysts will not likely be comparing "apples to apples." Even if they are not, it is important to understand the differences between events, which gives the decision-maker additional inputs when assessing the relevance of synthesized findings. Understanding how data were measured, collected, and analyzed allows the analyst to determine their confidence in event outputs. This is an important step because not all data are equal and the assessment of quality can help place contradictory data in context.

Construct validity examines how well the theoretical construct of interest in captured in the event or research. For example, how well is *contested logistics* incorporated into the experiment, wargame, or study? Additional information on construct validity is available in Chapter 4 of this report.

Internal validity examines whether an observed change in a dependent variable is indeed caused by a corresponding change in an independent variable. Causality requires three conditions: 1) covariation of cause and effect, 2) temporal precedence, and 3) no plausible alternative explanation or spurious correlation.[11] For experiments and studies the following are used as proxies for construct validity: an assessment of participant expertise; methods, procedures, and protocols used; and the data analysis. In wargames, the analog to internal validity may be assessed through participant expertise and wargame adjudication.

If the analyst is pressed for time, he or she can skip assessing construct and internal validity; however, the tradeoff may be to have less confidence in individual report results.

4. Phase IV – Synthesize the Data

Phase IV guides the analyst through aggregating data across events, which involves synthesizing data, assigning confidence, and identifying sources of uncertainty. Meta-synthesis is a way of looking across data to make observations, draw conclusions, and identify recommendations. The analyst first summarizes which events were identified to include against how many

---

[10] External validity examines whether the observed associations can be generalized from the observed sub-set to the population as a whole or to other groups or contexts. In terms of DoD events, it examines whether the observed associations can be applied in other contexts.

[11] Bhattacherjee, *Social Science Research*; Cecilia Maria Patino and Juliana Carvalho Ferreira, "Internal and External Validity: Can You Apply Research Study Results to Your Patients?," *Journal of Brasilian Pneumology* 44, no. 3 (May-June 2018); American Psychology Association, "Internal Validity," in *APA Dictionary of Psychology* (Washington, DC: American Psychological Association, 2022).

were eventually included as inputs to the analysis. The tallies should be done by event type (i.e., experiment, study, wargame). This information helps the requestor understand how complete the analysis was. Next the analyst will compare the data included with the review priorities that were initially set. The analyst will then synthesize the data across experiments, studies, and wargames independently but similarly.

1. Tally the 0-, 1-, and 2-point scores awarded in the assessment of event parameters, construct validity, internal validity, and overall relevance
2. Catalog and discuss the required event features that were/were not well represented
3. Catalog and discuss the relevant and credible results and conclusions
4. Catalog and discuss conflicting or differing results and conclusions
5. Note key sources of uncertainty and areas for additional research

Cataloging and discussing the required event features that were and were not well represented across the events suggests how mature and complete a DoD campaign of learning is. The analyst will also catalog and discuss relevant and credible results and conclusions. Given the nature of DoD campaigns of learning, conflicting or differing results are to be expected, though it is imperative that the analyst attempt to explain why these differences have occurred. Understanding this offers the requestor a measure of certainty when determining how applicable the observations, conclusions, and recommendations are to the question.

Uncertainty is anything that occurs in an event that could produce uncertainty in results. Sources of uncertainty can be introduced at any step in an event and can be classified in one of two categories: random (sources that cannot be predicted), or systematic (uncertainty related to procedure or instrumentation). Noting the key sources of uncertainty and areas for additional research is another means of assessing the maturity and completeness of a DoD campaign of learning.

Finally, the analyst will conduct an overall synthesis across all event types, informed by the event-specific synthesis conducted earlier in the worksheet. Again, the analyst will be asked to catalog and discuss required event features that were/were not well represented across the campaign of learning. They will then assign one of four levels of confidence in the observations made and conclusions drawn based on their synthesis of findings.

- **High confidence**: Major conclusions are supported by multiple independent experiments, studies, and wargames conducted on key aspects of the question under a variety of conditions. Body of work reflects campaign of learning as questions raised by one event are further explored in subsequent events. Difference in results are explainable.

- **Medium confidence**: Major conclusions have some support in individual experiments, studies, or wargames. Some key aspects of the question remain unexamined or were examined only under a narrow set of conditions and assumptions. Some differences in results remain unexplained.

- **Low confidence**: Major conclusions across events are mixed or contradictory. It is unclear whether key aspects of the question have been sufficiently addressed to justify conclusions. Only a few events can be said to have addressed key aspects of the question; or events may have addressed key aspects, but different events addressed different aspects, making comparisons of results and conclusions difficult.
- **Not enough information**: Key aspects of the question have not yet been sufficiently addressed in experiments, studies, or wargames to draw conclusions.

As part of the overall synthesis, the analyst will also characterize work to date; identify what, if any, aspects still need to be investigated; and provide additional input or feedback to the requestor. Qualitative analysis depends heavily on a researcher's analytic and interpretive skills and on personal knowledge to contextualize data. The emphasis is on understanding a phenomenon rather than predicting or explaining it.[12]

5. Phase V – Communicate the Results

Phase V facilitates transparency into the analyst's research and analytic process, enabling senior leaders to review the analyst's conclusions and recommendations. Research transparency relies on the accurate, complete, and accessible documentation of, among other things, the research approach, methods, data sources, data, and analysis.[13] If the analyst has properly documented the research and findings from the previous four phases, the Phase V should be trivial. Each table in the Phase V worksheet refers back to a specific table in one of the previous four worksheets so that it is easy to copy the data from each identified table to the Phase V worksheet. Completing this worksheet allows the analyst to communicate not only the results of the meta-synthesis, but also the information necessary to assess the analysis validity of the observations, conclusions, and recommendations.

B. Status Report

IDA developed one additional worksheet to track the progress of all RFIs within a given time period (e.g., within a given fiscal year). This gives DoD leadership a snapshot of workload on and throughput of their analysts. This is included in Appendix C of this report.

C. Using the Analyst Notebook

This chapter reviewed the framework. Detailed instructions on using the framework is available in Appendix C in the form of an Analyst Notebook. The purpose of this notebook is to guide analysts through the framework phases via concrete worksheets to document their work and render the process transparent and reviewable.

---

[12] Bhattacherjee, *Social Science Research*, 113.

[13] Andrew Moravscik, *Transprency in Qualitative Research* (Thousand Oaks, CA: SAGE Publications, 2019).

The Analyst Notebook is detailed and allows the analyst to examine a variety of inputs from experiments to analytic studies using modeling and simulation, which was at the sponsor's request. With the initial learning curve and the level of rigor aimed for in the Analyst Notebook, an analyst will want to schedule a considerable amount of time to learn and then apply the worksheets.

The notebook is also designed to allow at least one basic tradeoff between thoroughness and relative speed by making certain features optional. To this end, there are two versions of the framework: 1) the full version, and 2) a shortened version that skips construct validity and internal validity. The full version of the framework is meant to address the sponsor's request to focus on rigor and offer analysts the ability to "grade homework," i.e., examine and judge the quality of outputs from individual events. By looking construct validity, the analyst should be able to make a judgment on whether an event genuinely captured an idea such as "contested logistics." By looking at internal validity, the analyst is making a judgment on whether a report's conclusions logically flow from the inputs and methodology. The shortened version lets the analyst, in the interests of time, assume that organizations conducting experiments, studies, and wargames genuinely captured the concepts they said they were dealing with, and know how to conduct events and write appropriate reports. In the event that an analyst suspects that Service events are not testing a joint concept to the extent that the Services claim, the analyst should scrutinize construct validity in the events.

Over time, the analyst may consider truncating parts of the framework even further, focusing on the elements the analyst finds most useful, to allow for greater speed and targeted effort. The framework and notebook are intended to offer ideas to analysts who may not have the time or inclination to use the full version of the product developed in this report, but who may also want ideas on how to synthesize different types of information from multiple sources.

The sponsor's initial guidance to IDA also raised the possibility of using the resulting product as a common framework for analysts across DoD to communicate. Whether the framework or a variation of the framework presented in this chapter becomes a common tool will depend on uptake among the analyst community. This discussion item, however, points to the need for some common means of communicating results in order to support institutional memory, organizational learning, and organizational sensemaking across the body of DoD's activities.

3. Existing Analytic Frameworks

To develop the framework in Chapter 2, IDA reviewed existing analytic frameworks ranging across a number of fields. The review included frameworks from medicine, statistics, data science, natural sciences, social sciences, organizational learning, intelligence analysis, and defense analysis. The purpose was to cast a wide net from which potentially useful ideas could be drawn and assessed against the sponsor's needs. The interdisciplinary and wide-ranging nature of the J7's analytic requirements made an interdisciplinary scope appropriate and necessary. Given initial sponsor interest in some of these fields, it also made sense to begin with the above list.

A. Scope of Framework Review

The research team reviewed existing analytic frameworks from the fields of medicine; statistics, data science and operations research; the physical and social sciences; organizational learning; and defense research. This was based on a draft proposal to the sponsor and intended to sample a cross section of diverse fields. Other fields that IDA ultimately did not review due to time constraints and less direct applicability included models of software development, design thinking, other fields in engineering, management science, and additional branches of statistics.

The team also considered the question of the work that the national labs, other branches of the federal government, and allied governments are conducting and whether we could learn from them. Given the highly technical nature of the work typical in the national labs, but the comparatively less quantitative nature of many defense studies and reports, it did not appear that a meta-synthesis framework for DoD could meaningfully incorporate approaches to confidence developed, for example, for the Department of Energy.

The federal agencies involved in medicine and public health—such as Centers for Disease Control, Food and Drug Administration, and the National Institutes of Health—also use approaches to determining confidence. Again, however, the relatively well-defined and directly observable outcomes in clinical medicine posed considerable challenges to applying frameworks to activities like wargaming future concepts against future adversaries in hypothetical operational contexts. Other practices, such as convening independent advisory panels to advise the government on individual decisions, are scientifically valuable but less practical for action officers and decision makers within J7.

The United States also has a number of allies whose frameworks and approaches are available for examination; British defense frameworks to experimentation were included in the literature review. However, the smaller scale of defense activities overall among allies such as

Canada and Australia means that defense analysts in these countries feel less need to aggregate and compare activities across their defense ministries. Previous discussions and collaborations between the IDA team and British, Canadian, and Australian operations research and defense analysts has pointed to a significantly smaller scale of reports and sometimes improved long-term institutional memory because the same small group of analysts appear better able to remain aware of everything their organizations are doing.[14]

Previous IDA team experience with the NATO defense analysis community through the NATO Science and Technology Organization, NATO Allied Command Transformation, and other professional venues had revealed that other NATO allies often deal with even smaller defense enterprises and face fewer barriers as organizations to remaining aware of their experiments, studies, and wargames.[15]

For each framework reviewed, IDA identified its primary purpose, who uses it and how, strengths and weaknesses, and data requirements. The research team also identified relevant practices that could be applied to a meta-synthesis framework for J7.

B. Implications of Framework Review

As with any review that attempts to draw from multiple fields, time, resources, and existing familiarity naturally limit the range and direction of a framework search and literature review. Although there are a wide variety of frameworks that address different aspects of synthesis or validity, no one framework is suitable for the disparate range of results from DoD activities. The frameworks reviewed for this task (see below) tended to be domain-specific and designed to compare apples to apples, not apples to rocks to birds across a range of domains. Because of this, the IDA team drew from multiple frameworks for the meta-synthesis framework developed for this task and presented in Chapter 2.

The most useful frameworks included:

- The overall framing of systematic reviews, including from medicine;
- The way to focus on the initial question and criteria from meta-ethnography;
- Types of validity from social science;
- The cycle of research linking defense studies, wargames, and exercises;
- VV&A from defense M&S.

Each is discussed below among the many other frameworks evaluated for this task.

---

14 Frequent reassignments and the "up or out" system in the U.S. military also hampers institutional memory.

15 This includes participating in NATO Science and Technology Organization Systems Analysis and Studies panels, Military Operations Research Society events attended by partners and allies, and international wargaming conferences such as Connections UK, Connections North (Canada), and Connections US.

C. Medicine

Evidence-based medicine is a concept that grew out of a series of articles, published in the 1980s, describing the importance of critically assessing clinical evidence. The term was coined in the 1990s with a stated goal of being "aware of the evidence on which one's practice is based, the soundness of the evidence, and the strength of inference the evidence permits."[16] An emphasis on providing efficient, cost-effective, and evidence-based patient care has since been written into law. However, the body of research on any given topic in the medical field will rarely comprise just one study; clinical decisions cannot be based on any one study.

Systematic reviews and meta-analyses are two related frameworks used in the medical field to extract accurate, high-quality data from the overabundance of data in order to offer clinicians the most robust treatment options.[17] Both frameworks combine and analyze data from different studies conducted on similar research topics. In medical research, they are considered as twin frameworks differentiated by data and analysis type: systematic reviews are qualitative analyses of structured and unstructured data, whereas meta-analyses are quantitative analyses of structured data. However, this differentiation ignores the sequential relationship of the two frameworks and the importance of conducting a systematic review before a meta-analysis. While a meta-analysis requires a systematic review, the reverse is not true.

1. Systematic Review

A systematic review is any type of research review that searches for, assesses, synthesizes, and interprets research evidence of multiple studies in a manner that adheres closely to a set of recognized, standard methods for conducting the review. With few exceptions, systematic reviews aim to comprehensively and exhaustively identify the studies presenting research evidence relevant to the issue or issues being examined. The unit of interest in a systematic review is an individual study, not a study's data points.

Analytically, systematic reviews seek to provide unbiased syntheses of studies using rigorous and transparent methods, identifying what is known about the research on the topic and what remains unknown, addressing uncertainty around findings, and recommending future research. A systematic review collates the empirical evidence from two or more studies that fit a pre-specified set of inclusion/exclusion criteria in order to answer a specific research question. Systematic reviews are designed to summarize the findings of multiple studies, reducing the time delay in between research discovery and implementing findings and improving the generalizability and consistency of results. However, given the heterogeneity of study methodologies

---

[16] Gordon Guyatt, John Cairns, and David Churchill, "Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine," *Journal of the American Medical Association* 268, no. 17 (4 November 1992).

[17] The systematic review and meta-analysis are used across many disciplines. We introduce and define them here in the Medicine section because of the primacy and high standard the discipline assigns them, and do not suggest they are not used in other areas.

and reported data, it can be difficult to combine findings from different studies. The best systematic reviews transparently report criteria used to identify, screen, and include/exclude studies in analyses.[18] They evaluate the quality of evidence in terms of level of evidence presented in a study as well as the soundness of the methodology (see Figure 2).[19] These habits do not prevent bias, but allow researchers to evaluate for themselves the systematic review's quality.[20]

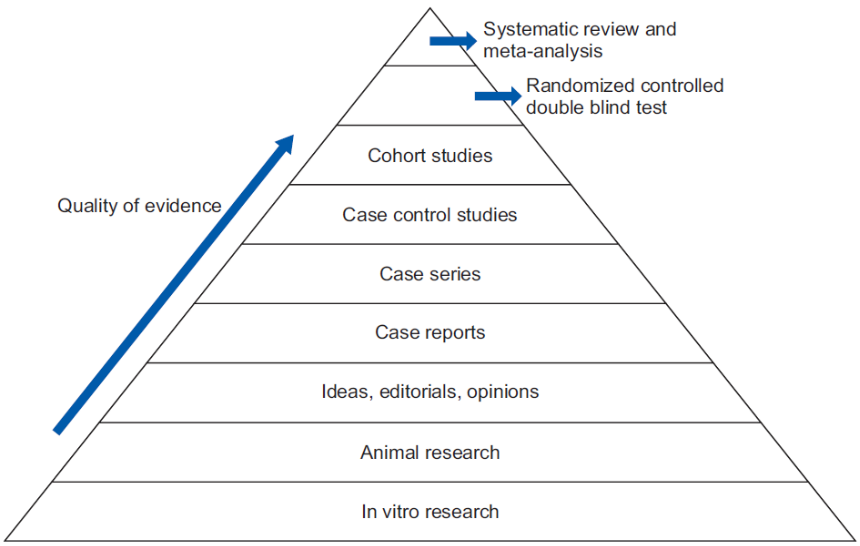

Figure 2. Levels of Evidence specific to Natural Science and Medical Research[21]

Because there are different types of research and research findings that address a range of questions, there are diverse types of systematic reviews that require different types of synthesis.[22] Generally though, given their nature as summaries, systematic reviews are less costly and quicker to run than the studies they summarize. However, there can be a loss of precision or nuance when data are summarized. While systematic reviews can reduce many different types of bias in design, analysis, and reporting, producing reliable, accurate conclusions, they are prone to publication bias.

---

18 Alessandro Liberati et al., "The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration," *PLOS Medicine* 6, no. 7 (21 July 2009).

19 Gabriel Rada, "Evidence Based Medicine (Emb) Toolkit: Discuss Emb: What Is the Best Evidence and How to Find It," BMJ Best Practice, 2022, accessed 25 April, 2022.

20 Axel Finckh and Martin R. Tramer, "Primer: Strengths and Weaknesses of Meta-Analysis," *Nature Clinical Practice Rheumatology* 4, no. 3 (March 2008); Ahn and Kang, "Introduction to Systematic Review and Meta-Analysis."

21 Ahn and Kang, "Introduction to Systematic Review and Meta-Analysis," 104.

22 Duke University, "Duke University Medical Center Library & Archives: Systematic Reviews," Duke University, updated 18 January 2022, 2022.

2. Meta-Analysis

Meta-analysis is a research technique for quantitatively synthesizing the results of multiple empirical studies through statistical analysis. Meta-analysis has been used extensively since the 1970s in biological and medical research, as well as within the behavioral and social sciences.

Relative to other systematic review methods, meta-analysis introduce less subjectivity while enabling the synthesis of studies with disparate research methods and findings; it is useful when empirical findings from different studies diverge. Meta-analysis is argued as having the following benefits.

- Uses quantitative data
- Eliminates study selection bias
- Findings are transformed to commensurable expressions of effect magnitude
- Data aggregated from multiple randomized, controlled studies enlarges the *n* of the study and, in turn, increases the accuracy of and confidence in results.
- Study characteristics that may mediate findings are defined, measured, and their covariation studied.
- Seeks general conclusions that do not obscure an important interactive finding

By combining data from independent studies, meta-analyses objectively appraise a range of studies, offering the clinician a means of keeping up to date in a rapidly evolving field while improving statistical precision over the independent studies. They are, however, subject to publication, reporting, and eligibility bias. Given the statistical nature of meta-analyses, study heterogeneity can be measured and reported. Finally, meta-analyses are ideal for identifying and understanding rare events and secondary disease outcomes.[23]

3. Conducting a Systematic Review or Meta-Analysis

Generally speaking, there are six agreed-upon steps in a systematic review and meta-analysis: (1) define the research question, (2) define selection criteria, (3) conduct search, (4) assess the quality of studies, (5) analyze the data, and (6) publish the results.[24]

The first step involves describing the problem or defining the question. This step scopes the review and analysis by defining Population, Intervention, Comparison, Outcomes (PICO) parameters. The first parameter—population—describes what population or group of patients

---

23 Finckh and Tramer, “Strengths and Weaknesses of Meta-Analysis,” 146-150.

24 Ahn and Kang, “Introduction to Systematic Review and Meta-Analysis.”; The Darwin Trust of Edinburgh, “The Scientific Method (Archived),” The Darwin Trust of Edinburgh, 2019; Gopalakrishna and Ganeshkumar, “Systematic Reviews and Meta-Analysis.”; Markus MacGill and Deborah Weatherspoon, “What Is a Systematic Review in Research?,” *Medical News Today* (25 February 2019).

are considered in-scope (e.g., lung cancer patients). Intervention, the second parameter, describes the main intervention, prognostic factor, or exposure being considered (e.g., tobacco use). Comparison, the third parameter, describes the main alternative to compare with the chosen intervention, prognostic factor, or exposure (e.g., non-tobacco use). The final parameter—outcomes—describes which affects, measures, or improvements will be considered (e.g., nodule quantity and size).[25]

The second step is to define inclusion and exclusion criteria. Such criteria may cover the type of study design (e.g., double-blind, randomized, controlled studies), patient characteristics, publication status, language, and the period of research. The third step encompasses literature searches and identifying studies that meet the selection criteria; also included in this step is documenting search terms, databases searched, types of searches conducted, and numbers of results returned and screened (for more on this, see the discussion of QUOROM/ PRISMA best practices in Appendix E).[26]

The fourth step involves assessing the quality of evidence based on the level of evidence (see Figure 2 for a depiction of the different levels of evidence and their relative strength), soundness of methodology, and risk of bias, which enables excluding studies with weak or questionable evidence due to a lack of transparency. One methodology for assessing the quality of evidence is the GRADE (Grading of Recommendations, Assessment, Development and Evaluation) system, which evaluates on the basis of study limitations, inaccuracies, incompleteness of outcome data, indirectness of evidence, and risk of bias.

Several types of bias must be considered: selection, performance, detection, attrition, and reporting. The quality assessment step is critical because when underlying evidence is poor, the findings can be inaccurate or incorrect. Because end-users will ultimately assess the quality of the systematic review or meta-analysis themselves, it is important to document and report assessments of included and excluded studies. This documentation is only part of the overall documentation during the fifth step—data extraction—which requires developing coding instruments (e.g., coding sheets and codebook). Coding instruments are designed for each specific review or analysis and collect four categories of data: methodology and substantive features of the underlying studies, study quality, intervention descriptors, and outcome measures.[27]

The penultimate step in a systematic review or meta-analysis is data analysis. The key differences between the two frameworks lie in the types of data included, and thus the type of analyses conducted. Whereas systematic reviews can accommodate a range of data types, the

---

[25] Kathy A. Jensen, *Seven Steps to the Perfect Pico Search: Evidence-Based Nursing Practice*, EBSCO Health (Ipswich, MA: EBSCO Health, 2018), 4.

[26] Ahn and Kang, "Introduction to Systematic Review and Meta-Analysis," 105.

[27] Ahn and Kang, "Introduction to Systematic Review and Meta-Analysis," 105; Brown, Upchurch, and Acton, "A Framework for Developing a Coding Schem for Meta-Analysis," 207.

ideal data for a meta-analysis are those produced in a randomized, controlled study.[28] As such, systematic reviews are limited to qualitative synthesis of data in the form of descriptive statistics. In contrast, meta-analyses are a quantitative synthesis of data aggregated from multiple randomized, controlled studies that enlarges the *n* of the study and, in turn, increases the accuracy of and confidence in results. The final step requires researchers to present and/or publish their results.[29]

D. Statistics, Data Science, and Operations Research

1. Design of Experiments

Experiments in general involve one or more components, called "factors," that are tested for their hypothesized effects on a system. When an analyst hypothesizes whether a particular factor is present or whether the treatment level of the factor's presence helps to determine the measurable outcome in the system, the typical process is to compare the results of experimental trials using various treatment levels of the factor, such as with/without or high/medium/low doses. However, when multiple factors, each with various treatment levels, are included in an experiment, choosing all of the potential treatment combinations for each factor (known as a full factorial design) can quickly expand to an unwieldy set of trials, and can have multiple design issues: they can be so large as to be impractical—or impossible—to perform, and they can fail to produce a satisfactory level of confidence in estimates.[30]

Design of experiments refers to a collection of mathematical techniques designed to help efficiently reduce the variance in statistical estimation of experimental parameters. Here, efficiency is related to the cost in time, effort, and money of performing experiments with the purpose of evaluating the contribution of (typically) multiple factors to the effectiveness of a system. Efficiency is particularly important when the number of factors is even more than a few, or the cost of each trial of an experiment is prohibitive. A well-designed experiment can produce an optimally-high degree of confidence in estimate in the value of multiple factors given the constraints of time or money or availability of the factors to be tested.

a. Basics of Experimentation

Consider a system with three potential inputs (i.e., factors) whose presence or absence is hypothesized to affect the output of the system. An analyst might set up an experiment where all the possible combinations of factor levels (in this case, eight of them) are tested (see Table 1).

---

[28] Ahn and Kang, "Introduction to Systematic Review and Meta-Analysis," 103.

[29] Ahn and Kang, "Introduction to Systematic Review and Meta-Analysis," 105-110.

[30] Sections B.1–B.2a/b are guided by D.R. Cox and Nancy Reid, *The Theory of the Design of Expeirments*, Crc Monographs on Statistics and Applied Probability, (London, UK: Chapman & Hall, 2000); Keith M. Bower, "Quality Resource, Glossary: Design of Experiments (Doe)," American Society for Quality, 2022.

Table 1. Example of Testing All Possible Combinations of Factors

	Factor A	Factor B	Factor C
Trial 1	Not Present	Not Present	Not Present
Trial 2	Not Present	Not Present	Present
Trial 3	Not Present	Present	Not Present
Trial 4	Present	Not Present	Not Present
Trial 5	Not Present	Present	Present
Trial 6	Present	Not Present	Present
Trial 7	Present	Present	Not Present
Trial 8	Present	Present	Present

This design may be practical and, under certain assumptions about the nature of the variance of the factors, may in fact be the most optimal. However, running eight trials in a warfighting experiment, for example, may not be practical or even possible. The analyst may decide to construct a new design, using half the number of trials as in Table 2.

Table 2. Example of Orthogonal Testing

	Factor A	Factor B	Factor C
Trial 1	Present	Not Present	Not Present
Trial 2	Present	Present	Present
Trial 3	Not Present	Not Present	Present
Trial 4	Not Present	Present	Not Present

This particular design is called orthogonal because each factor has an equal number of all treatment levels (two *Present* and two *Not Present* in this case), and each pair of factors has an equal number of treatment level combinations (one each, in this case). The advantages of this design are that estimates for the main effects of factors can be made independently and the number of trials is reduced. The chief disadvantage is that the main effects estimates will have a slightly larger variance due to small sample size.

Design of experiments can help optimize trials when factors are not assumed to have the same variance, or when some factor levels require more time and or expense to use and when the analyst has a limited budget. Optimal designs will, in these cases, attempt to increase the number of trials that use the higher-variance factors, and to decrease the number of "expensive" factors.

b. Design of Experimentation in Wargaming

One would generally not think of applying experimental design techniques to meta-analyses, though the principles for designing experiments themselves can be useful when selecting data for comparing a compilation of trials. The process for collecting, selecting, and analyzing such data can be time consuming. If an analyst is fortunate enough to have a collection of experiment reports when the nature of the experiments is relatively similar, and when there are multiple factors examined across the collection, there is a potential for using experimental design techniques.

Suppose a series of wargames or experiments is conducted to establish the potential benefits of three conceptual capabilities as proofs on concept, both as individual events and in tandem. An analyst could combine the findings of the several experiments to find more robust results.

Table 3records the presence of each capability. This particular design, while not orthogonal, might nevertheless be optimal or close to optimal if Capability C has a lower known variance, and Capability D, a greater one.

Table 3. Alternative to Testing All Possible Combinations of Factors or to Orthogonal Testing

	Capability A	Capability B	Capability C	Capability D
Experiment A	o	✓	o	✓
Experiment B	✓	o	o	o
Wargame C	o	✓	o	✓
Wargame D	✓	o	✓	o
Exercise E	✓	o	o	✓
Exercise F	O	✓	✓	o
Exercise G	o	o	O	✓

The main requirement for using design of experiment techniques for a meta-analysis is that there is some quantifiable measure of effectiveness for the system being examined. This is, of course, generally required for any meta-analysis with a quantifiable objective.

One reminder about using design of experiments on wargame and experimentation results is that comparisons across multiple events should consider learning between individual events. It is possible to account for learning within the experimental design itself. Particularly when wargames are repeated, players have a learning curve and individual wargames should not be treated as analogous to individual runs of a statistical simulation.

2. Ensemble Learning

Ensemble learning, a statistical and machine learning approach that uses multiple learning algorithms to obtain better predictive performance, helps overcome problems associated with the dimensionality and class imbalance of data. It is thus well-suited to classifying high-dimensional data, though limited for most wargaming analyses.[31] There are two broad classes of ensemble learning: homogeneous and heterogeneous. The former uses a collection of classifiers of the same type built upon different subsets of data. Individual machine learning models (e.g., decision trees) tend to have high variance. Random forest models that combine multiple decision trees improve results by reducing variance. Applying similar random forest models to different data sets provides different classifiers. Heterogeneous ensemble learning, in contrast, uses a set of classifiers of different types built on the same data.

3. Model & Simulation Validation

Model and simulation validation is a broad concept that includes the construct, internal validity, and external validity. The nomenclature for validity depends on the discipline using models and simulations. Regardless of the nomenclature used, validation involves demonstrating that a model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application. Validation can depend heavily on a model's selection and evaluation, and its comparison to other models. The validation process begins with an objective about what the model is intended to do and by examining the outputs of the model, determines whether the model meets that objective. Even if a model is not "correct", it can still meet a useful purpose.[32] For example, if a model is intended to generate events for a training exercise to allow participants to confront different scenarios, it does not matter if the model is theoretically or statistically correct, because the intended output is simply a range of different conditions.

Validation demonstrates that models, statistics, and wargames accurately represent the real-world relationships as they are understood. It can be applied to models that use both qualitative and quantitative inputs, and it helps ensure that data inputs are good enough to meet the analytic objectives.[33]

---

[31] Mohammad Nazmul Haque and Pablo Moscato, "From Ensemble Learning to Meta-Analytics: A Review on Trends in Business Applications," in *Business and Consumer Analytics: New Ideas*, ed. Pablo Mostcato and Natalie de Vries (New York, NY: Springer Publishing, 2019); Cha Zhang and Yunqian Ma, *Ensemble Machine Learning: Methods and Applications* (New York, NY: Springer Publishing, 2012).

[32] Department of Defense, *Test & Evaluation Management Guide*, 6th ed. (Washington, DC: Department of Defense, August, 2016); Systems Engineering Group, "Systems Engineering Guide: Verification and Validation of Simulation Models," MITRE, 2021.

[33] Department of Defense, *Test & Evaluation Management Guide*; Systems Engineering Group, "Systems Engineering Guide."; Heather Wojton et al., *Handbook on Statistical Design & Analysis Techniques for Modeling & Simulation Validation*, Institute for Defense Analyses (Alexandria, VA: Institute for Defense Analyses, February 2019).

E. Natural Sciences

The natural sciences encompass naturally occurring objects and phenomena and can be further sub-classified as physical science, earth science, and life science.[34] Below is a discussion of aspects of three approaches to natural science research.

1. Scientific Method

The scientific method is not a single, agreed-upon set of steps to answer a scientific question.[35] Rather, it is a logical problem-solving approach that can be applied and adapted to answer many scientific and non-scientific questions. In essence, it offers a means of discovering cause and effect relationships. Despite the lack of complete agreement on individual steps, there are seven recognizable steps that comprise the scientific method.[36]

1. Make an observation
2. Ask a question
3. Formulate a hypothesis
4. Predict based on the hypothesis
5. Test the prediction/hypothesis
6. Analyze data and draw conclusions
7. Reproduce or iterate
8. Publish/communicate

Initially, a researcher must observe—recognize or note—a phenomenon, fact, or occurrence.[37] By posing a question about that observation—why something happens or how something came to be—a researcher is driven to form a hypothesis in an attempt to answer the question.[38] The hypothesis—a best-guess explanation, commonly in the form of a conditional statement—does not necessarily need to be correct, but does need to be testable.[39] Based on the hypothesis and given a set of testing conditions, a researcher predicts a result and devises a test

---

[34] Bhattacherjee, *Social Science Research*, 1-2.

[35] Henry M. Cowles, *The Scientific Method: An Evolution of Thinking from Darwin to Dewey* (Cambridge, MA: Harvard University Press, 2020), 1; Max Black, *Critical Thinking: An Introduction to Logic and Scientific Method* (Aukland, NZ: Muriwai Books, 2018), 303.

[36] Biology Dictionary refers to steps 1-3 and 5-7 (iterate). Larry Li, "Scientific Method," Biology Dictionary, updated 6 November 2020, 2020. The Darwin Trust of Edinburgh refers to steps 1, 4, 5, and 7 (iterate). The Darwin Trust of Edinburgh, "The Scientific Method (Archived)." Encyclopedia Britannica refers to steps 1, 2, and 5. Encyclopedia Britannica, "Scientific Method," in *Encyclopedia Britannica* (Chicago, IL: Encyclopedia Britannica, Inc., 2021). Live Science refers to steps 1-3 and 5-7 (reproduce). Alina Bradford, "Live Science/What Is Science," Future US, Inc., 2017.

[37] Li, "Scientific Method."; The Darwin Trust of Edinburgh, "The Scientific Method (Archived)."; Encyclopedia Britannica, "Scientific Method."; Bradford, "Live Science/What Is Science."

[38] Li, "Scientific Method."; Encyclopedia Britannica, "Scientific Method."; Bradford, "Live Science/What Is Science."

[39] Li, "Scientific Method."; Bradford, "Live Science / What Is Science."

to prove or disprove the hypothesis.[40] Only through analyzing data can one draw a conclusion whether the hypothesis predicted a result or requires refining.[41] If the former, a researcher would replicate the test to increase the scientific community's confidence in the hypothesis and publish her/his findings. If the latter, a researcher would refine or replace the hypothesis and begin the process again until ready to publish (See Figure 3 for a schematic of the process).

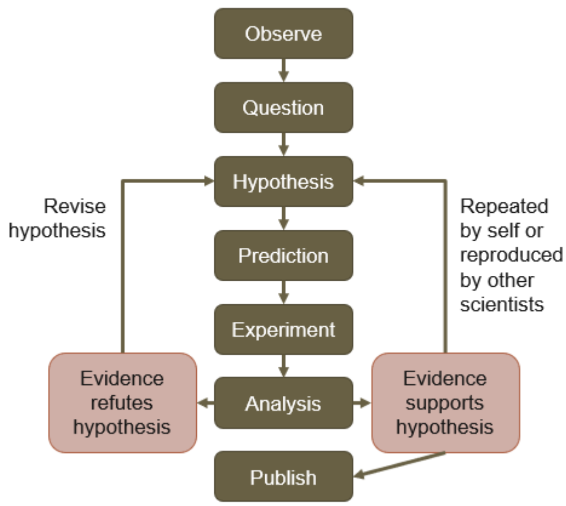

Figure 3. Schematic of the Scientific Method

There are two main types of testing applied to the scientific method: controlled experiments and non-experimental hypothesis testing, sometimes referred to as quasi-experimentation. Both aim to prove or disprove a hypothesis through scientifically rigorous testing; however, they cannot be employed interchangeably.

a. Controlled Experiments

Controlled experiments have three hallmarks: they employ control and experimental groups in order to study a single variable, they measure the relationship between dependent and independent variables, and they increase confidence in the results through sufficient sample size and repetition.[42] All with the purpose of determining a cause-and-effect relationship.

In an ideal experiment, control and experimental groups are identical in every respect except one. A researcher would vary a single condition—the independent variable—between the two groups, then measure the effect of the independent variable on one or more dependent

---

[40] Li, "Scientific Method."; The Darwin Trust of Edinburgh, "The Scientific Method (Archived)."; Encyclopedia Britannica, "Scientific Method."; Bradford, "Live Science/What Is Science."

[41] Li, "Scientific Method."; Bradford, "Live Science/What Is Science."

[42] Li, "Scientific Method."; The Darwin Trust of Edinburgh, "The Scientific Method (Archived)."; Encyclopedia Britannica, "Scientific Method."; Bradford, "Live Science/What Is Science."

variables—outcomes that potentially change based on changes in the independent variable. To increase confidence in experimental data, researchers use control and experimental groups of statistically sufficient size. Additionally, they run experiments under identical conditions multiple times. This replication further increases confidence in experimental results.[43]

b. Non-Controlled Hypothesis Testing

In contrast to controlled experiments, non-experimental hypothesis testing measures variables as they occur naturally without further manipulation. While there are independent and dependent variables, the researcher does not manipulate the independent variable before collecting data on the dependent variable. This type of testing is employed when direct experimentation is either impractical (e.g., testing hypotheses that would require the passage of tens, hundreds, thousands, millions, or billions of years) or unethical (e.g., human experimentation).[44] There are two main types of non-experimental designs: cross-sectional and correlational research.

Cross-Sectional Research

Cross-sectional research is observational, capturing a snapshot in time.[46] Subjects are selected based on the study's inclusion/exclusion criteria.[47] Ideally, subjects are similar in all variables except the one under review; however, because subjects are in pre-existing groups and not randomly assigned, groups can be dissimilar and non-homogenous. Because not all variables are controlled for, the strength of statistical analyses are limited. Given the limitations of statistical analyses and the fact that outcomes and exposure are measured at the same time, this type of research is considered descriptive, rather than causal or relational. It is ideally suited for describing common characteristics in a group and can form the baseline for correlational research.[48]

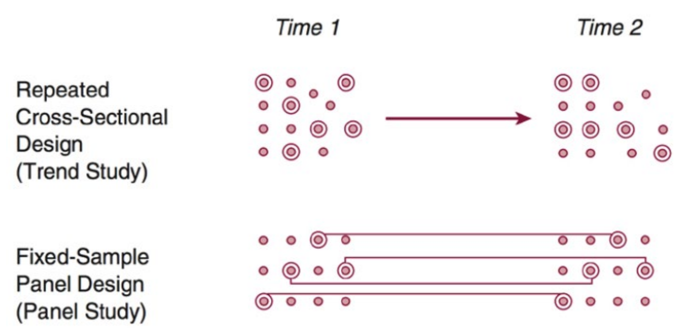

Figure 4. Repeated Cross-Sectional versus Longitudinal Panel Studies[45]

The benefits of cross-sectional research are three-fold. Because all

[43] Khan Academy, "Controlled Experiments."

[44] Khan Academy, "Controlled Experiments."

[45] Lipps, "Panel Surveys."

[46] Kendra Cherry and Dr. Steven Gans, "How Does the Cross-Sectional Research Method Work?," About, Inc., updated 10 October 2019.

[47] Maninder Singh Setia, "Methodology Series, Module 3: Cross-Sectional Studies," *Indian Journal of Dermatology* 61, no. 3 (May-June 2016).

[48] Setia, "Cross-Sectional Studies."; Cherry and Gans, "How Does the Cross-Sectional Research Method Work?."

variables are collected only once, they are relatively quick and inexpensive. Additionally, researchers can study multiple outcomes and exposures simultaneously. Finally, this type of study is ideal for descriptive analyses (i.e., "descriptive statistics"), which can then be used to form hypotheses. However, cross-sectional studies are susceptible to low-response bias or misclassification due to recall bias.[49]

A repeated cross-sectional study is identical in structure to a cross-sectional study but repeats data collection with fresh subjects each time. This type of study can be used to identify trends within groups whose constituents are not constant, such as the wellbeing of employed versus unemployed people over time.[50]

Correlational Research

Correlational research is also observational, but measures the direction (i.e., positive, negative, zero), form (i.e., linear v. non-linear), and strength of the relationship between independent and dependent variables over time. A common type of correlational research is the longitudinal study, which observes variables over an extended period of time. Three common subsets of longitudinal studies include panel studies, cohort studies, and retrospective study.[52]

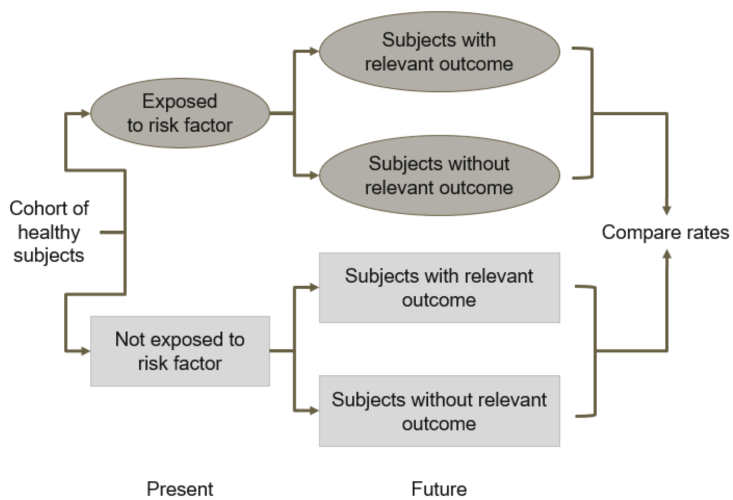

Figure 5. Notional Prospective Longitudinal Cohort Study[51]

Longitudinal panel studies, also referred to as time-series studies, collect repeated measurements from the same subjects at different points in time. Data are collected and tracked per individual. Most are designed for quantitative analyses, using structured data; but can also

[49] Setia, "Cross-Sectional Studies."; Health Knowledge, "Introduction to Study Design: Cross-Sectional Studies," Public Health Action Support Team CIC, 2020.

[50] Oliver Lipps, "Panel Surveys: Advantages and Disadvantages Compared with Repeated Cross-Sectional Surveys," *FORS Guide*, no. 14 (2021).

[51] Adapted from Mendel Suchmacher and Mauro Geller, "Study Type Determination," in *Practical Biostatistics*, ed. Mendel Suchmacher and Mauro Geller (Waltham, MA: Elsevier Academic Press, 2012).

[52] Francis Lau, "Methods for Correlational Studies," in *Handbook of Ehealth Evaluation: An Evidence-Based Approach*, ed. Francis Lau and Craig Kuziemsky (Victoria, BC: University of Victoria, 2016); Maninder Singh Setia, "Methodology Series, Module 1: Cohort Studies," *Indian Journal of Dermatology* 61, no. 1 (January-February 2016); Kendra Cherry and Amanda Tust, "The Pros and Cons of Longitudinal Research," About, Inc., updated 2 May 2020.

employ qualitative methods for collection and analysis of data.[53] A longitudinal panel study is similar to a repeated cross-sectional study, except where a repeated cross-sectional study would use new individual subjects each time, the subjects of a longitudinal panel study remain constant. Panel studies allow researchers to see trends at the individual vice group level. However, they are subject to attrition over time and panel conditioning (i.e., that answers to a survey are influenced by subjects being members of the panel).[54]

Longitudinal cohort studies are a forward-looking observational study that follows large groups of subjects over a period of time. Groups are alike in many respects but differ by certain characteristics (i.e., independent variable: female nurses who smoke v male nurses who smoke). A particular outcome (i.e., dependent variable: development of lung cancer) is compared between the groups at one or more points in the future.[55] Cohort studies can establish cause-effect, identify rare exposures, and determine disease incidence rates and relative risk. Additionally, researchers can determine whether more than one disease is associated with a single exposure. Finally, they minimize selection and information bias, but require large populations and a long time to complete, thus are costly. Additionally, they are prone to ethical issues.[56] Figure 5 illustrates a notional breakdown of a prospective (i.e., future-looking) longitudinal cohort study.

A *retrospective study*, also known as a case-control or record-linking study, is a backwards looking observational study that enrolls subjects who already have the relevant outcome (i.e., disease or condition). Researchers then enroll comparison subjects who are similar to disease subjects in most respects but do not have the disease or condition under study. Researchers compare the results of questionnaires and/or medical records between the groups to identify risk factors or other associations common among the individuals within the subjects of one

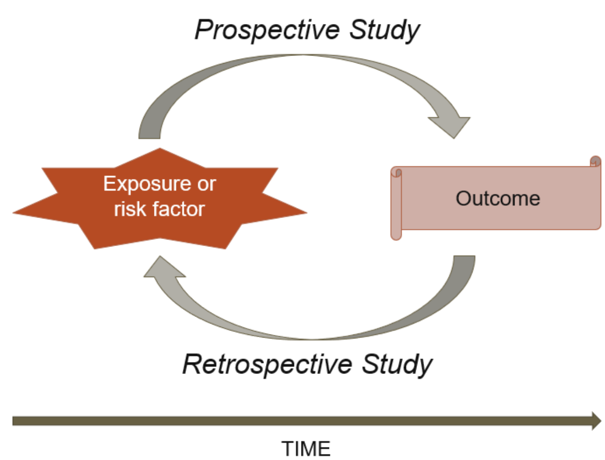

Figure 6. Retrospective versus Prospective Longitudinal Study[57]

---

[53] Heather Laurie, "Oxford Bibliographies, Panel Studies," Oxford University Press, updated 26 August 2013, 2021.

[54] Lipps, "Panel Surveys."

[55] Lau, "Methods for Correlational Studies."; National Cancer Instutite, "NCI Library > Longitudinal Cohort Study," National Cancer Institute, 2021.

[56] Lau, "Methods for Correlational Studies."; Dr. Twarita Chakraborty, "Analytical Study Design in Medical Research," Peseapro Scientific Services Ltd., updated 3 February 2014; Dr. Twarita Chakraborty, "Analytical Study Design in Medical Research: Cohort Study," Peseapro Scientific Services Ltd., updated 3 March 2014.

[57] Adapted from Setia, "Cohort Studies."

group but not the other.[58] Retrospective studies are relatively quick and inexpensive, can identify rare exposures, allow for multiple cohorts, and involve no ethical issues. However, the risk of sampling/selection bias is higher relative to prospective studies. Additionally, it may be difficult to identify disease versus control groups. Finally, available medical records may not contain all the requisite data for analyses.[59] See Figure 6 for an illustration of the relationship between retrospective and prospective studies.

F. Social and Behavioral Sciences

The social and behavioral sciences—psychology, sociology, economics, anthropology, and political science—share the logic of inquiry with the physical and life sciences. One key differentiator is the extent to which the objects of research in the social and behavioral sciences actively play a role in the process of inquiry.[60] However, the underlying logic of scientific inquiry and the goal of applying the scientific method to understand variables and resulting phenomena is also central to research methods is the social and behavioral sciences. This holds whether the underlying epistemic philosophy is *positivist* (focusing on causal inference and generalizable theory) or *constructivist* (arguing that multiple theories may hold for a given context).[61] Below is a discussion of aspects of and approaches to social and behavioral science research.

1. Measurement

Valid measurement occurs when scores meaningfully capture the design of the study concept.[62] Measurements should be vetted for error, reliability, and validity. All three inform the design of instrumentation administered in mixed-methods studies. Understanding and assessing the quality of measurements and instrumentation used in studies across a meta-analysis is an essential foundation.

The primary purpose of measurement as a framework is to ensure that variables are created and managed to effectively understand the relationships they have with each other.[63] Understanding said relationships between variables relies on the reliable and valid measurement constructs. By one framing, there are four levels of measurement: background concept, systematized concept,

---

58 National Cancer Instutite, "NCI Library > Retrospective Cohort Study," National Cancer Institute, 2021; Stephanie Glen, "Restrospective Study: Case-Control and Case-Series," Statistics How To, updated 22 August 2016.

59 Lau, "Methods for Correlational Studies."; Setia, "Cohort Studies."; Maninder Singh Setia, "Methodology Series, Module 2: Case-Control Studies," *Indian Journal of Dermatology* 61, no. 2 (March-April 2016).

60 Geoffrey Maruyama and Carey S. Ryan, *Research Methods in Social Relations*, 8th ed. (Hoboken, NJ: Wiley-Blackwell, 2014), 7.

61 Maruyama and Ryan, *Research Methods in Social Relations*, 5.

62 Robert Adcock and David Collier, "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research," *American Political Science Review* 95, no. 3 (17 January 2002).

63 Delbert C. Miller and Neil J. Salkind, *Handbook for Research Design and Social Movement*, 6th ed. (Thousand Oaks, CA: Sage Publications, January, 2002).

indicators, and scores for cases.[64] Within the background concept, a systematized concept is formulated through reasoning and exploring broader issues that could impact the concept. Once a concept is formulated, it is operationalized by developing indicators for scoring and classifying data that will be collected. When these indicators are applied to data collected within the confines of the concept, numerical or qualitative scores are produced that can be analyzed.

Measuring a concept requires independent/dependent variables, nonparametric variables (nominal and ordinal variables), and parametric variables (interval and ratio variables). Valid measurement occurs when scores meaningfully capture the ideas of the corresponding concept. Measurement is a universal framework; it is used to operationalize systematized concepts of quantitative, qualitative, and mix-methods nature. As such, there are no weaknesses to this framework because it serves as a foundation on which to build other frameworks. Measurement is used in all forms of research and should be integral to meta-analyses.

2. Triangulation

Triangulation is defined as the utilization of multiple methods or data sources to develop a comprehensive understanding of phenomena.[65] It is a procedure to search for a convergence of different types of sources to form themes or categories in a study.[66] Triangulation is a framework within social research to also test validity through a convergence of information from multiple sources. A common practice is for inquirers to provide corroborating evidence collected across observations, interviews, and documents. There are four types of triangulation: method triangulation, investigator triangulation, theory triangulation, and data source triangulation.

Method triangulation comprises multiple methods of collecting data about the same concept or phenomena. Investigator triangulation both confirms findings and offers a breadth of different perspectives to the object of study. Theory triangulation utilizes multiple theories to analyze and interpret data. Data source triangulation consists of a variety of data from different types of sources to gain multiple perspectives and validate data.

Combining qualitative and quantitative methods to explore a research concept can lead to three outcomes: (1) the results of the analyses may converge, (2) the results of the analyses might relate to different patterns or phenomena but still be complementary, and (3) the results of the analyses may diverge or contradict each other.[67]

---

64 Adcock and Collier, "Measurement Validity."

65 Nancy Carter et al., "The Use of Triangulation in Qualitative Research," *Oncology Nursing Forum* 41, no. 5 (September 2014).

66 John W. Creswell and Dana L. Miller, "Determining Validity in Qualitative Inquiry," *Theory into Practice* 39, no. 3 (24 June 2010).

67 Roberta Heale and Dorothy Forbes, "Understanding Triangulation in Research," *Evidence-Based Nursing* 16, no. 98 (August 2013).

There are potential weaknesses of triangulation as a framework for examining different types of analyses. First, the framework makes a significant assumption that data from the distinct sources or methods have similar or different weights in the research process. Second, convergence as a result of triangulation could be due to human or data error. Researchers should validate measurements and their data, and should have multiple strategies to ensure that data are dependable and credible.[68]

3. Meta-Ethnography

Meta-ethnography is an evidence synthesis methodology most frequently used in healthcare, social care, and educational research. [69] The goal of synthesis using this methodology is to discover higher order interpretations that may lead to new explanatory theories. The methodology requires studies using qualitative interpretive data, preferably rich, detailed accounts to facilitate the discovery of new themes or concepts across several studies. Meta-ethnography comprises seven phases. meta-ethnography likely requires a team of researchers, in order to confirm and validate the subjective interpretations of meanings and concepts found in the selected studies

1. *Getting Started*. Determining a research question that can be informed by qualitative research.
2. *Deciding what is relevant to the initial interest*— Identifying and selecting studies.[70] While quality appraisal is not transparent in all meta-ethnographies, some decisions on inclusion and exclusion criteria must be made at this phase.
3. *Reading studies*. Beginning to identify themes and concepts.
4. *Determining how studies are related*. Studies could be grouped by focus and synthesized within groups before synthesizing across groups, or by grouping common concepts and themes found within the studies (see Table 4 for an example).

---

68 Carter et al., "The Use of Triangulation in Qualitative Research."

69 George W. Noblit and R. Dwight Hare, *Meta-Ethnography: Synthesizing Qualitative Studies* (Newbury Park, California: Sage Publications, 1988). Emma F. France et al., "A Methodological Systematic Review of Meta-Ethnography Conduct to Articulate the Complex Analytical Phases," *BMC Research Methodology* 19, no. 1 (February 2019 2019).

70 France et al., "A Methodological Systematic Review of Meta-Ethnography Conduct to Articulate the Complex Analytical Phases," 2.

Table 4. Completed Grid: Lay Meanings of Medicines[71]

Methods & Concepts	Donovan & Blake	Morgan	Britten	Rogers *et al*
Sample	54 patients with suspected inflammatory arthropathy	60 White and Afro-Caribbean patients treated for hypertension for at least one year	30 patients, attenders, and non-attenders	34 patients with a diagnosis of schizophrenia or schizo-affective disorder
Data Collection	Home interviews with pre- and post-consultations; observation of consultations	Home interviews	Home interviews	Interviews
Setting	3 rheumatology units	15 general practices	2 general practices	Different points in the mental health system
Type(s) of Medicine	MSAIDs and second-line drugs	Antihypertensive drugs	Unselected	Neuroleptic medication
Adherence / Compliance	Patients do not perceive compliance to be an issue	Stable adherence and problematic adherence	Correct behavior and routine medicine taking	Patients mentioned benefits of taking medicines
Self-regulation	Levels of non-compliance	Leaving off drugs	Preference for not taking drugs	Adjustment of medication, self-regulation
Aversion	Dislike of taking drugs, fear of side-effects, weakness, dependence	Fear of side-effects, addiction, harmful effects of drugs	Aversion to medicines, medicine as harmful	Wide range of side-effects
Alternative coping strategies	Range of alterative remedies	Traditional (herbal) remedies	Use of alternative medicine	Alternative coping strategies
Sanctions	-	Doctors warned patients and told severely of the need to take tablets regularly	-	Coercion from significant others, fear of coercion from mental health professionals
Selective disclosure	Patients did not tell doctors of altered doses	-	-	Management of information to psychiatrists
Explanation/theory (second-order interpretation)	Patients carry out a “cost-benefit” analysis of each treatment, weighing the costs/risks of each treatment against the benefits they perceive	Medicine-taking is influenced by cultural meanings and cultural resources	Patients may not articulate views that they do not perceive to be medically legitimated	The self-regulation of medication appears to have been circumscribed or inhibited by impact of the threat of social and professional sanctions

5. *Translating the studies into one another*. How do the studies relate? Reciprocal, refutational, and line of argument synthesis can be employed to during the translation.[72]

---

[71] Nicky Britten et al., “Using Meta-Ethnography to Synthesise Qualitative Research: A Worked Example,” *Journal of Health Service Reseach & Policy* 7, no. 4 (October 2002 2002): 212.

[72] Mary Dixon-Woods et al., “Synthesising Qualitative and Quantitative Evidence: A Review of Possible Methods,” *Journal of Health Services Research & Policy* 10, no. 1 (2005): 48.

6. *Synthesizing translations*. Establish the relationship between the studies selected is the most is the most complex and difficult to explain and replicate.[73] A line of argument can include a new hypothesis, a mid-range theory, an overarching model, or a form of grounded theory. A meta-ethnography may produce multiple lines of argument, which may or may not be combined into one theory.[74] See Table 5 for an example of concepts being translated across studies.

Table 5. Synthesis, Including Concepts and Second- and Third-order Interpretations[75]

Concepts	Second-order Interpretations	Third-order Interpretations
Adherence/Compliance: stable adherence; correct behavior and routine medicine-taking Self-Regulation: problematic adherence; levels of non-compliance; leaving off drugs; preference for not taking drugs; self-regulation	Patients carry out a "cost-benefit" analysis of each treatment, weighing costs/risks of each treatment against the benefits as they perceive them	Self-regulating includes the use of alternative coping strategies
Aversion: dislike of taking drugs; fear of side-effects; aversion to medicines; harmful effects of drugs Alternative Coping Strategies: range of alternative remedies; traditional remedies	Medicine-taking is influenced by cultural meanings and cultural resources	
Sanctions: patients are warned by their doctors and told severely about the need to take tablets regularly; coercion from significant others, fear of coercion from mental health professionals	Self-regulation… inhibited by… the threat of social and professional sanctions	Self-regulation flourishes if sanctions are not severe
Selective Disclosure: patients do not tell doctors of altered doses; management of information to psychiatrists	Patients may not articulate views they do not perceive to be medically legitimated	Altering coping strategies are not seen by patients as medically legitimate; Fear of sanctions and guilt produce selective disclosure

7. *Expressing the synthesis*. Presenting how the studies were selected and translated, and any new testable hypothesis

Meta-ethnography allows researchers to review the context and concepts of primary and second order interpretations in studies while facilitating the development of new theories. It is a systematic approach that preserves interpretive characteristics of primary data more than other qualitative synthesis methodologies (i.e., qualitative meta-synthesis).[76]

[73] Britten et al., "Using Meta-Ethnography to Synthesise Qualitative Research," 211; Atkins et al., "Conducting a Meta-Ethnography of Qualitative Literature: Lessons Learnt."

[74] France et al., "A Methodological Systematic Review of Meta-Ethnography Conduct to Articulate the Complex Analytical Phases."

[75] Britten et al., "Using Meta-Ethnography to Synthesise Qualitative Research," 213.

[76] María Sobrido Prietoa and José María Rumbo-Prieto, "The Systematic Review: Plurality of Approaches and Methodologies," *Enfermeria Clinica* 28 (January 2018): 391.

One potential shortcoming of this methodology is that it provides no guidance on how to sample or appraise studies, therefore they can lack transparency and replicability. It may also be difficult to synthesize descriptive studies and interpretive studies. Finally, the process is laborious, so meta-ethnography may not be suitable if a large number of studies need to be included.[77]

4. Meta-Synthesis

Meta-synthesis is a method of qualitative research synthesis. The term *meta-synthesis* is used to refer to both a specific methodology and to reference the larger family of qualitative research synthesis methods, each of which has its own epistemological assumptions.[78] This discussion will refer specifically to the methodology developed by Sandelowski and Barroso from their *Handbook for Synthesizing Qualitative Research*.[79]

Researchers in the healthcare and psychiatry fields use meta-synthesis to develop an evidence base for policy makers and practitioners because it's purpose is to reinterpret findings and broaden understanding of a particular phenomenon. The method relies on qualitative research findings with the possibility of further interpretation that then offer broader interpretations that then can build upon that research.

Meta-synthesis stands out from other qualitative research methodologies because it describes a "stimulus to thinking and creativity' rather than a 'prescriptive set of rules and procedures to be rigidly followed,'"[80] thus enabling its success in a variety of health-related topics.[81] The robust nature of the methodology may suit it to other research fields. Another strength is the distinction between *meta-summary* and *meta-synthesis* (discussed in the steps that follow). Meta-summaries can facilitate meta-synthesis, but does not necessarily need to lead to synthesis if findings are not of the proper classification, or if effect sizes are too small to suggest an explanatory theory. This distinction allows review teams to explore whether the studies they select are appropriate for synthesis, or to publish the meta-summary and meta-synthesis as two separate products.[82]

---

[77] France et al., "A Methodological Systematic Review of Meta-Ethnography Conduct to Articulate the Complex Analytical Phases."; Dixon-Woods et al., "Synthesising Qualitative and Quantitative Evidence: A Review of Possible Methods."

[78] Oliver Rudolf Herber and Julie Barroso, "Lessons Learned from Applying Sandelowski and Barroso's Approach for Synthesising Qualitative Research," *Qualitative Research* 20, no. 4 (2020): 417.

[79] Margarete Sandelowski and Julie Barroso, *Handbook for Synthesizing Qualitative Research* (New York: Springer, 2007).

[80] Sandelowski and Barroso, *Handbook for Synthesizing Qualitative Research*, xv.

[81] Herber and Barroso, "Lessons Learned from Applying Sandelowski and Barroso's Approach for Synthesising Qualitative Research," 416.

[82] Herber and Barroso, "Lessons Learned from Applying Sandelowski and Barroso's Approach for Synthesising Qualitative Research."

Sandelowski and Barroso's meta-synthesis methodology consists of the following steps[83]

1. *Conceiving the synthesis* involves several activities that facilitate the synthesis, including to "address a research purpose, consider available resources, decide on the target phenomenon of the study, and establish inclusion and exclusion criteria".[84] This includes determining the underlying assumptions and theoretical underpinnings of the proposed synthesis, which may require consulting more experienced researchers.[85] Whereupon, researchers may want to narrow the research question and limit the date range.[86]
2. *Searching and retrieving literature* involves developing inclusion and exclusion criteria, such as topical, population, temporal, and methodological parameters relevant to their research question before searching the literature. Reviewers should continually reevaluating search terms, and conducting a rigorous and comprehensive search.
3. *Appraising findings* requires reading and re-reading the studies to familiarize oneself with the content and identify findings. Appraisal tools such as the Critical Skills Appraisal Programme (CASP) and Joanna Briggs Institute's Qualitative Appraisal and Review Instrument (QARI) can help.[87] The number of questions in the appraisal tool can dramatically increase the amount of work, and depending on resources and constraints, it may be necessary to use a tool with fewer questions. No study should be excluded due to a judgement of poor quality so as to avoid introducing bias.[88]
4. *Classifying findings* is closely linked with the third step, and some parts may be performed concurrently. Reviewers should categorize findings based on whether they have transformed the information from data through some degree of interpretation. Findings can be classified into one of five categories: no finding, topical survey, thematic survey, conceptual/thematic description, and interpretive explanation (see Figure 7).

[83] Sandelowski and Barroso, *Handbook for Synthesizing Qualitative Research*.

[84] Mette S. Ludvigsen et al., "Using Sandelowski and Barroso's Meta-Synthesis Method in Advancing Qualitative Evidence," *Qualitative Health Research* 26, no. 3 (2016): 322.

[85] Ludvigsen et al., "Using Sandelowski and Barroso's Meta-Synthesis Method in Advancing Qualitative Evidence," 322; Herber and Barroso, "Lessons Learned from Applying Sandelowski and Barroso's Approach for Synthesising Qualitative Research," 417; Kenneth Finlayson and Annie Dixon, "Qualitative Meta-Synthesis: A Guide for the Novice," *Nurse Researcher* 15, no. 2 (2008): 61-63.

[86] Herber and Barroso, "Lessons Learned from Applying Sandelowski and Barroso's Approach for Synthesising Qualitative Research," 418.

[87] Ludvigsen et al., "Using Sandelowski and Barroso's Meta-Synthesis Method in Advancing Qualitative Evidence," 324.

[88] Herber and Barroso, "Lessons Learned from Applying Sandelowski and Barroso's Approach for Synthesising Qualitative Research," 420; Finlayson and Dixon, "Qualitative Meta-Synthesis: A Guide for the Novice," 64-65.

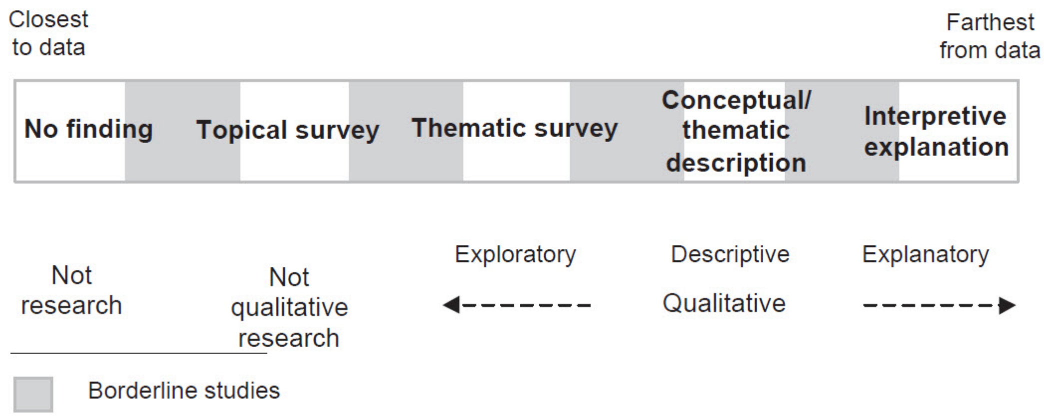

Figure 7. Typology of Findings[89]

To ensure findings are classified correctly, a team should reach consensus in the case of borderline studies.[90] It may be useful to identify one prototypical example of a study for each class of finding, so that the review team can use it for comparison.[91]

5. *Synthesizing findings into meta-summaries* involves developing a "quantitatively oriented aggregation of qualitative findings that themselves are topical or thematic summaries or surveys of data"[92] and comprises five phases: extracting findings, editing findings for readability, grouping findings, abstracting findings to facilitate further grouping, and calculating effect sizes.[93] Text findings should be extracted and numbered or grouped into statement sets according to topic/theme, which allows reviewers to quickly identify patterns and form new hypotheses. For synthesis involving a large number of studies, qualitative data analysis software can help managing the data set.[94] Sandelowski and Barroso suggest that only findings from topical or thematic surveys should be included in the meta-summary; however, some researches contend that more interpretive findings and qualitative research should be included, because meta-summary can be a first step toward completing the meta-synthesis.[95]

---

[89] Sandelowski and Barroso, "Classifying the Findings in Qualitative Studies," 908.

[90] Sandelowski and Barroso, "Classifying the Findings in Qualitative Studies," 914-916.

[91] Herber and Barroso, "Lessons Learned from Applying Sandelowski and Barroso's Approach for Synthesising Qualitative Research," 421.

[92] Sandelowski and Barroso, *Handbook for Synthesizing Qualitative Research*, 151.

[93] Herber and Barroso, "Lessons Learned from Applying Sandelowski and Barroso's Approach for Synthesising Qualitative Research," 422.

[94] Ludvigsen et al., "Using Sandelowski and Barroso's Meta-Synthesis Method in Advancing Qualitative Evidence," 325; Herber and Barroso, "Lessons Learned from Applying Sandelowski and Barroso's Approach for Synthesising Qualitative Research," 424.

[95] Herber and Barroso, "Lessons Learned from Applying Sandelowski and Barroso's Approach for Synthesising Qualitative Research," 423.

6. *Synthesizing findings into meta-synthesis* is the final step and consists of identifying themes that emerge across studies and integrating them to prompt further insight beyond that of the summary. There are several ways to carry out this step. A taxonomy may be used to classify related themes that best describe particular phenomenon. Events of interest could be placed on an event timeline to determine how they are related. Reciprocal translation could be used to translate common themes between studies. Other analytics devices include line of argument synthesis, in which maps and diagrams can be used to draw connections between concepts and themes contained in findings.[96] The range of options offers researchers the ability to select an approach they are most comfortable or familiar with. Additionally, the approach selected could be tailored based on identified findings and the results of the meta-summary.

One shortcoming of this method is the requirement for prerequisite knowledge: meta-summary and meta-synthesis would be difficult for inexperienced researchers to conduct. Prospective research teams should be well-versed in qualitative research methods and have experience conducting qualitative research.[97] Additionally, there are philosophical issues with the appropriateness of qualitative synthesis methods.[98] For example, some authors contend that the suitability of taking abstract concepts from studies and using them as the evidence of a new conclusion is questionable, especially in the field of clinical practice.[99]

5. Validity

Validity is a foundational framework that can be used to build upon by other frameworks; it is necessary to assess the effectiveness of any type of data or relationships between variables. The primary purpose of validity is to ensure that variables are measured and represented based on their intended design. Researchers can use a variety of tested procedures to evaluate and mitigate any of the aforementioned threats to the four major types of validity. Once a study or instrument is deemed valid, only then should it be used for predictive power.

Validity touches all facets of research analysis: qualitative methods, quantitative methods, and mixed methods. Requirements to ensure validity include statistical power, unviolated assumptions, reliability of measures, non-random relevancies, and non-random heterogeneity. Mitigation of bias is also significant in construction of instruments, data collection, and analysis.

---

96 Herber and Barroso, "Lessons Learned from Applying Sandelowski and Barroso's Approach for Synthesising Qualitative Research," 425.

97 Herber and Barroso, "Lessons Learned from Applying Sandelowski and Barroso's Approach for Synthesising Qualitative Research," 427.

98 Finlayson and Dixon, "Qualitative Meta-Synthesis: A Guide for the Novice," 65.

99 Ludvigsen et al., "Using Sandelowski and Barroso's Meta-Synthesis Method in Advancing Qualitative Evidence."; Herber and Barroso, "Lessons Learned from Applying Sandelowski and Barroso's Approach for Synthesising Qualitative Research," 426.

Validity pertains to the meaningfulness of research components and whether a measurement is measuring what it was designed and intended for. Validity as a framework is based on determining whether findings are accurate from the standpoint of researchers, participants, and readers. It provides procedures to ensure that instruments of measurement have been tested for rigor and have the ability to be replicated in future studies and analyses. Confirming validity is even more paramount in a mixed-methods meta-analysis, given the aggregation of several instruments of data collection in one large study.

In the study of social sciences, there are four types of validity: statistical conclusion validity, internal validity, construct validity, and external validity. In statistical conclusion validity, we question if a relationship actually exists between variables. There are some major threats to statistical conclusion validity: low statistical power, violation of assumptions, reliability of measures, reliability of treatment, random irrelevancies in the experimental setting, and random heterogeneity of respondents. Internal validity questions if there is a relationship, is it a causal relationship? It also seeks to answer if there are any confounding variables in the study. The threats to internal validity include: history, maturation, testing, instrumentation, selection, and mortality.

Construct validity follows up on the initial question of internal validity. If a relationship is causal, what are the particular cause and effect behaviors in the relationship? It seeks to answer the larger question on how well one has translated a concept and has operationalized it. There are in turn several aspects to construct validity:

- *Face validity* is the informal evaluation by experts in the field whether a measuring technique measures what it aims to measure. Face validity is subjective and does not constitute enough evidence for construct validity. However, most valid constructs should face this informal test.
- *Content validity* is the extent to which the measurement scale covers the full range of the construct the research is trying to assess.
- *Criterion validity* is the extent to which a measure accurately captures the relationship to other variables or outcomes. Criterion validity may be further broken down to *concurrent validity* (the extent to which a measure predicts criteria examined at the same time), and *predictive validity* (the extent to which measures obtained at a certain point in time predict a future criterion).
- *Convergent validity* is the extent to which two measures that are expected to be related are empirically shown to be related.
- *Discriminant validity* is the extent to which a measurement distinguishes between the construct of interest and other constructs.
- Lastly, *external validity* asks how generalizable a given causal relationship established in the research is to other contexts, such as different populations, settings, and timeframes.

G. Organizational Learning and Sensemaking

1. Organizational Learning

Organizational change and learning scholars have used meta-analysis techniques to assess the empirical literature on the receptivity for organizational change and learning in an effort to produce a representative and generalizable guide to organizational change and learning.[100] In particular, organizational learning—the process of acquiring, distributing, integrating, and creating information and knowledge among organizational members—has come to be seen as a key process that contributes to successful innovation. However, despite exponential growth in studies of organizational learning during the past 30 years, organizational learning remains an elusive concept for researchers and practitioners, characterized by a lack of evidence-based guidelines for those wishing to implement it.[101] However, as Deric documents, the use of meta-analysis to rigorously review and synthesize existing empirical studies across the field of organizational learning "appears to be changing this state of affairs."[102]

2. Organizational Sensemaking

Organizational sensemaking comprises a fluid set of concepts concerning how to approach organizational studies. The concepts describe the process by which people notice and interpret events and coordinate a response to clarify the meaning of those events. It is best adapted to events that breakdown meaning, such as crisis situations, periods of strategic change, the formation and dissolution of organizations, and the socialization of new members to a group. As such, organizational sensemaking is best suited for qualitative rather than quantitative inputs. Generally, any discussion of organizational sensemaking will include the following features as part of the broader concept.[103]

- *Retrospection*. Sensemaking is retrospective in nature. As such, meaning is constructed after reflection on an experience, which provides rationality and clarity to an outcome.
- *Plausibility*. Meaning is constructed on the basis of reasonable explanations rather than scientific discovery.
- *Cue-based*. Meaning can only be constructed on the basis of a sub-set of data because individuals and organizations become overwhelmed when confronted with expansive inputs.

---

[100] Andrew B. Burris, "A Qualitative and Quantitative Assessment of Readiness for Organizational Change Literature" (MS Engineering Management Air University, Air Force Institute of Technology, 2008) (AFIT/GEM/ENV/08-J01).

[101] Aleksandra Deric, "Organisational Learning and Learning Capability: What They Are, Why They Matter, and How to Foster Them," CQ Net, updated 10 November 2019.

[102] Deric, "Organisational Learning and Learning Capability…".

[103] Ravi S. Kudesia, "Organizational Sensemaking," in *Oxford University Research Encyclopedia* (Oxford, UK: Oxford University Press, 26 April 2017).

- *Enacted constraints*. Limits on future decisions and outcomes are a result of the outcomes of previous decisions.
- *Social*. The process of attributing meaning to an outcome is inherently social by nature. Meaning is constructed through conversations, communications, and the exchange of ideas.
- *Ongoing*. The process of determining meaning is constantly negotiated.
- *Grounded in Identity*. The process of determining meaning is a product of and a process based on who the sensemaker is and is becoming.[104]

H. Structured Analytic Techniques

Structured analytic techniques are formal exercises intended to help intelligence analysts reduce the effects of cognitive biases and improve collaboration with other analysts. While these methods pull from a variety of sources and do not necessarily originate from the intelligence community or are limited to only intelligence analysts, the U.S. intelligence community has recently been a significant player in collecting, publishing, and encourage the use of these approaches.[105]

The techniques allow analysts to use slow, reasoned analytic thinking methods to reduce the subconscious biases that plague intuitive thinking (a fast, unconscious method of thinking).[106] The methods reveal internal thought processes transparently and systematically, exposing faulty assumptions and allowing other analysts to more easily critique or build on this analysis.[107] The definitive handbook on the subject is Richards Heuer and Randolph Pherson's *Structured Analytic Techniques for Intelligence Analysis*, which details dozens of these techniques, their purposes, and when and how to conduct them.

In the wake of the September 11 attacks, these techniques have become increasingly popular methods to create alternative analysis, challenge assumptions, reduce intuitive errors, and increase analyst collaboration. U.S. Intelligence Community agencies have integrated these techniques into their tradecraft and analyst training, and expect analysts to incorporate them into their workflow. Some of the techniques benefit from using trained, experienced facilitators, but many can be performed by any analyst or team of analysts.

---

[104] "Industrial-Organizational Psychology: Organizational Sensemaking," iResearchNet.com, 2022, accessed 25 January, 2022.

[105] Richards J. Heuer Jr. and Randolph H. Pherson, *Structured Analytic Techniques for Intelligence Analysis* (Washington, DC: CQ Press, 2015).

[106] Heuer Jr. and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 5-6.

[107] Heuer Jr. and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 4.

The techniques offer a suite of tools that vary in purpose, method, and complexity. Heuer and Pherson classify the structured analytic techniques into eight families based on their objectives.

Table 6: Eight Types of Structured Analytic Techniques[108]

Technique Family	Objective	Examples
Decomposition and Visualization	Allow analysts to break down large problems into smaller components, and to map interconnections and structure visually.	checklists, timelines, matrices, network analysis, mind maps, Gantt charts
Idea Generation	Enable eliciting a lot of ideas at the beginning of a project, to explore the gamut of possibilities rather than prematurely rejecting ideas.	structured brainstorming (and several variants), morphological analysis, quadrant crunching
Scenarios and Indicators	Allow generating a series of plausible futures, determining likelihood and potential impact, and describing signposts that might indicate which future might occur.	alternative futures analysis, the cone of plausibility, indicator generation
Hypothesis Generation and Testing	Employ aspects of the scientific method and abductive reasoning to develop and test hypotheses.	several methods of hypothesis generation, diagnostic reasoning, analysis of competing hypotheses, argument mapping
Assessment of Cause and Effect	Help differentiate between causation and correlation.[109] They can also help analysts understand the reasons behind human behavior without "mirror imaging" (assuming that a person would act in the same way as the analyst in a given situation) or risking fundamental attribution error (assuming that a person's actions spring from a personality-driven decision, rather than external context).	key assumptions checks, structured analogies, role-playing, red hat analysis
Challenge Analysis	Help analysts question the analytic consensus, reframe the question, and consider a wider variety of possibilities.	premortem analysis, structured self-critique, high-impact, low-probability analysis, devil's advocacy, red team analysis
Conflict Management	Encourage analytic disagreements to become constructive and improve analysis, rather than by forcing a consensus, watering down the analysis, or noting the disagreement in a footnote.	adversarial collaboration, structured debate
Decision Support	Allow analysts to provide decision-makers with the appropriate information structured in a format that enables important choices while minimizing irrational decisions.	decision trees, force field analysis, Strengths, Weaknesses, Opportunities, Threats (SWOT) analysis

108 Table information drawn from Heuer, Jr. and Pherson, *Structured Analytic Techniques for Intelligence Analysis*.

109 Heuer Jr. and Pherson, *Structured Analytic Techniques for Intelligence Analysis*, 205-206.

The strengths of structured analytic techniques include—

- *Simplicity.* Most require little training, resources, or time; most analysts can be facilitate.
- *Adds structure to ambiguity.* Formality and rigor enhance approaching unstructured analytic problems, beyond intuition alone.
- *Collaboration.* Structured analytic techniques enable teams of analysts to work together instead of separately, gaining from a diverse set of insights instead of just one person's intuition.
- *Engagement*. Structured analytic techniques can be engaging and fun if done correctly, allowing analysts to take breaks from their normal work routine and engage creatively with the analysis.

The weaknesses of structured analytic techniques include—

- *More art than science.* Most of these methods have not been rigorously, scientifically evaluated, and would be difficult to test quantitatively because the methods are so qualitative. [110] The little testing that has occurred has had mixed reviews, with some techniques showing success and others not.[111]
- *Team requirements.* Most require a team of analysts. Moreover, they require participants to have a certain level of expertise, creativity, and open-mindedness. They also require leadership to buy into the concept of structured analytic techniques, or else they will not allow the time and training required to put on these exercises.
- *Potential overconfidence.* If conducted poorly or in bad faith, these techniques could potentially "launder" bias, using the techniques as evidence that the findings are unbiased, even if they are not. They could also lead to bias overcorrection or introducing new "noise" to analytic judgments.[112]

Given how varied these techniques are, the type of data required as input also varies. Typically, structured analytic techniques rely solely on qualitative data. Because they are meant to approach ambiguous, unstructured, limited intelligence information, they do not require large quantities of high-quality data. That said, more, better quality data will certainly lead to better analysis, but the analysts' expertise, creativity, and open-mindedness is more important than data quality.

---

[110] Welton Chang et al., "Restructuring Structured Analytic Techniques in Intelligence," *Intelligence and National Security* 33, no. 3 (10 November 2017).

[111] Stephen J. Coulthart, "An Evidence-Based Evaluation of 12 Core Structured Analytic Techniques," *International Journal of Intelligence and Counterintelligence* 30, no. 2 (8 February 2017).

[112] Chang et al., "Restructuring Structured Analytic Techniques in Intelligence."

I. Defense Research

1. Cycle of Research

In this context, the cycle of research refers to the integration of the outcomes of analyses, assessments, exercises, and wargames to support defense decision-making.

Figure 8 shows the relationship between the different event types and the quantitative (data from analyses; actions from exercises) qualitative (decisions from wargames) inputs that can be synthesized. The framework explains the differences between the various event types but offers only limited guidance on how to synthesize results across them.[113] The cycle of research is iterative where outcomes from one type of event informs inputs to subsequent events of other types, in order to validate those outcomes and discover issues or unaddressed assumptions.[114] For example, a new approach discovered during a wargame might be submitted for modeling and simulation before being attempted in a field training exercise.

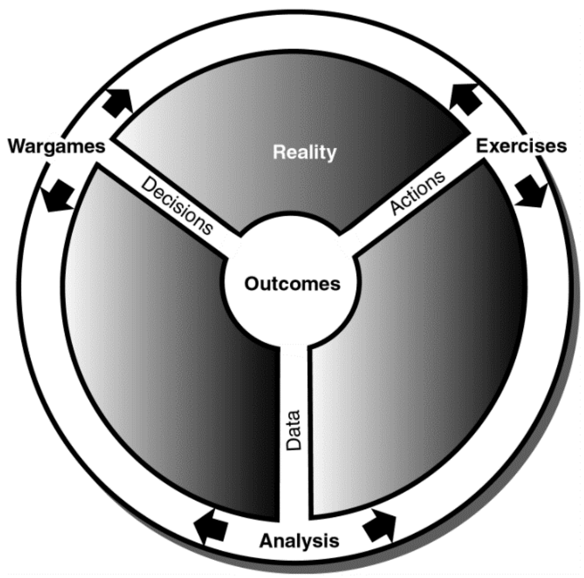

Figure 8. Cycle of Research[115]

According to Perla, wargames, analyses, and exercises all look at the same problem, but from a different perspective. Wargames focus on decisions and people, exercises focus on actions and reality of execution, and analysis focuses on data and physics. All three need to be

---

113 Peter P. Perla, *The Art of Wargaming: A Guide for Professionals and Hobbyists* (Annapolis, MD: Naval Institute Press, 1990), 255; Peter P. Perla et al., "Rolling the Iron Dice: From Analytic Wargaming to the Cycle of Research," *War on the Rocks* (21 October 2019).

114 Wargames do not validate, but provide insights that allow researchers to consider complexities related to the wargame.

115 Perla et al., "Rolling the Iron Dice."

integrated with real-life data from experiments (modeling) if the problem is to be understood.[116] Wargames provide insight into the options available to actors in a given scenario, but may lack precision and rigor. Campaign models using quantitative analysis can supplement those options with data to assess an actor's expected performance in that scenario, given the means they have available. Exercises and weapons tests further supplement those activities by providing data on how an actor's actions or systems would perform in the real world.[117]

2. Verification, Validation, and Accreditation

Verification, validation, and accreditation (VV&A) is a process used to ensure the application of modeling and simulation (M&S) results are appropriate for the intended purpose.[118] By their nature, military M&S vary significantly in character and foundation. There can be vast differences in information and data quality from one process to another. Some concepts and phenomena that serve as the foundation for M&S and analyses can be extremely reliable or poorly understood.[119] Due do the strong possibility of variance in data quality in the VV&A process, validity can at most be considered conditional.

Verification is the process of determining that a model implementation (i.e., a program) accurately represents the developer's conception and specifications.[120] Verification can be evaluated within the context of degrees of accuracy—it is not often feasible to do a line-by-line check of a model.

Validation is the process of determining (a) the manner in which and degree to which a model (and its data) is an accurate representation of the real world from the perspective of the intended uses of the model, and (b) the subjective confidence that should be placed on this assessment.[121] Descriptive validity suggests that a model has the ability to accurately explain phenomena meaningfully. Structural validity means that a model has the right objects and variables to correspond to the real world. Predictive validity means that a model has the ability to predict system behavior within some level of accuracy.

---

[116] Peter P. Perla, "Why Wargaming" (MORS Certificate in Wargaming Course, Virtual, Military Operations Research Society, 2022).

[117] Phillip Pournelle, "Can the Cycle of Research Save American Military Strategy?," (October 18 2019).

[118] US Coast Guard, *Verification, Validation, and Accredidation (Vv&a) of Models and Simulations (M&S)*, US Department of Homeland Security (Washington, DC: US Department of Homeland Security, 22 December 2006).

[119] Paul K. Davis, *Generalizing Concepts and Methods of Verification, Validation, and Accredidation (Vv&a) for Military Simulations*, RAND Corporation (Santa Monica, CA: RAND Corporation, 1992).

[120] Davis, *Generalizing Concepts and Methods of Verification, Validation, and Accredidation (Vv&a) for Military Simulations*.

[121] Davis, *Generalizing Concepts and Methods of Verification, Validation, and Accredidation (Vv&a) for Military Simulations*.

Accreditation is an official determination that a model is acceptable for a specific purpose or application.[122] Accreditation is often conditional upon the analytic plan itself that includes initial logic to establish conclusions. One important reality is that models might be good for some studies and not for others, thus accrediting a model for a whole class of studies might be unwarranted.

3. Warfighting Experimentation

In the scientific world, an experiment is "an operation or procedure carried out under controlled conditions in order to discover an unknown effect or law, to test or establish a hypothesis, or to illustrate a known law."[123] (For more on scientific experimentation, see Natural Sciences in Chapter 3.) Drawing from Richard Kass, the elements of a scientific experiment should hold true in the military world as well. Warfighting experiments involve testing whether and how some military concept or new capability affects changes to military effectiveness. In order to best determine that, the experiments need to be performed in a controlled manner—typically by comparing the difference in effectiveness with the capability or concept, and without.[124]

This section discusses how an analyst can assess whether a warfighting experiment was performed such that the results can be determined with a certain level of confidence. This, in turn, translates into determining whether an experimentation report is suitable to use in a meta-analysis. If an analyst can answer the questions in Figure 9 positively and satisfactorily, then the capability was sufficiently and rigorously tested.

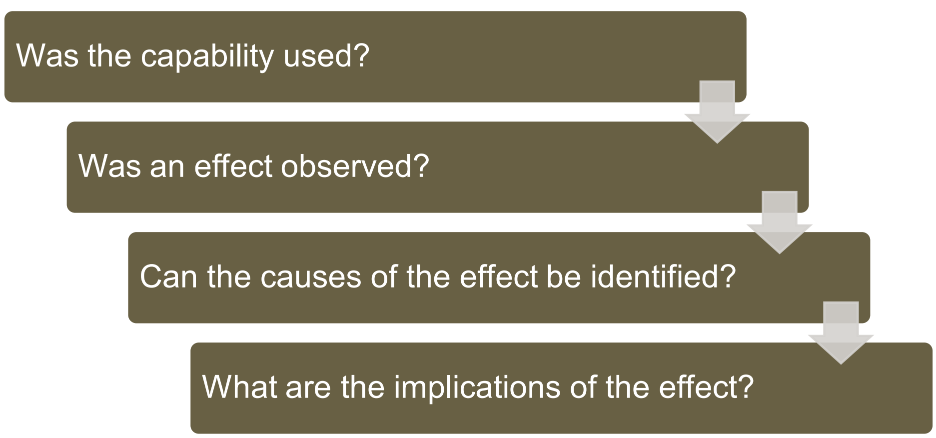

Figure 9. Requirements for a Rigorous Test

---

[122] Davis, *Generalizing Concepts and Methods of Verification, Validation, and Accredidation (Vv&a) for Military Simulations*.

[123] Merriam-Webster, "Experiment," in *Merriam-Webster* (Online, Springfield, MA: Merriam-Webster Inc., 2022).

[124] This sub-section is drawn largely from Richard A. Kass, *The Logic of Warfighting Experiments*, The Future of Command and Control, (Washington, DC: Command and Control Research Program, 2006).

a. Was the capability used?

This is a relatively obvious requirement that is frequently relevant for warfighting experiments. New capabilities are often tested as part of a different exercise or separate training, and if the capability, such as a new technology, is unavailable for some or all of the trials in the experiment, then the capability's effect on the system cannot be assessed. The assessment of the technology's preparedness for use may be an interesting subject in itself, but it is almost certainly not going to be the subject of the analysis at hand.

Other aspects of this requirement involve—

- *The nature of the experiment:* Was the experiment designed so that the system could be sensitive to the proposed capability?
- *The training for the capability:* Were the participants sufficiently and properly trained in the use of the new capability?
- *The new capability did not have the opportunity to perform: W*as the capability actually employed during the exercise or training?

b. Was an effect observed?

In order to analyze an effect of the capability, one must be able to detect a change in the system's measure of effectiveness or measure of performance during the wargame. To determine whether the capability is the agent of change, that ability must bear fruit—i.e., a change in effect needs to be observed.

There are many reasons a real effect may not be detected; for example, its effect may be "buried in the noise," which can be cause by any of a number of reasons, such as—

- *Systems may vary in performance*: Does the system, particularly a new capability, varies in performance from trial to trial?
- *Small sample size:* Are there enough trials to statistically prove a difference?
- *Operators may vary in proficiency within a trial*
- *Measurement methods may be unreliable*
- *Trial conditions vary within or between trials*

c. Can the causes of change be identified?

Just because a change in effect is observed does not mean the capability being tested is responsible for the change. Many possible sources of a change are in effect during a wargame. The analyst must take care to identify as many as possible before attempting to attribute the observed effect to the capability.

One of the main causes for erroneously attributing a change in effect to a capability being tested is related to training on the new capability. For example, if the capability is used only during the final set of trials in a wargame, it is possible that an observed effect is due only to player learning during the trials. There are ways to address this situation before the experiment begins by, for example, changing the order of trials that include the capability. Similar to the learning effect, an increase in proficiency in measurement can occur—though usually one cannot adjust for that change by changing trial order.

Another potential confounding factor is the capability's performance. New technology may perform better during trials as developers identify weaknesses or incorporate suggestions for improvement from players and operators. The analyst must account for these improvements when attempting to attribute causes of change.

d. What are the implications of the observed effect?

This requirement represents the *So what?* of the identification of an effect likely caused by the introduction of the capability. Several questions need to be addressed:

- Was the capability, operators, and system representative of a future implementation?
- Was the scenario realistic?
- Was the observed effect an "important" one?

The answers to these questions may be a matter of opinion of subject matter experts, warfighters, and other experts, and thus difficult for the wargame analyst to address. It is no less difficult for one attempting a meta-analysis, as those opinions may not be consistent from wargame to wargame.

Potential weaknesses in Kass's framing include conflating experiments with demonstrations, a much more narrowly scoped form of activity that experimentation. Kass's framing is also less applicable to broader experimental campaigns or discovery experimentation, being more focused on the cause-and-effect relationship within single experiments.

4. Existing Approaches to Measuring Confidence

In addition to frameworks, the sponsor asked IDA to review approaches to measuring levels of confidence. This is a germane question when the sponsor encounters a situation where analysts are required to judge whether there is enough evidence for a conclusion, or how much confidence to put into a conclusion given the information available.

The two areas the IDA team reviewed for measuring confidence was the literature around statistical confidence, and the levels of confidence developed by the intelligence community. Most areas of applied research employ concepts about confidence and uncertainty from statistics. The levels of confidence from the intelligence community were designed to communicate with senior decision makers. Given the sponsor's preference not to focus on confidence in systems-level testing, the IDA team did not examine confidence levels from testing and evaluation.

There are limitations of both approaches for the defense experimentation community: not all output produced by the services, DoD agencies, and other entities allow translation into statistical confidence. For its part, the levels for confidence from the intelligence agencies are around topics where the issue is access to information in the face of potential deception, and sometimes around predicting single-point events in time. Both circumstances are narrower than the requirements of the joint experimentation and analysis community.

A. Statistical Significance

Statistical analysis is a means of investigating trends, patterns, and relationships using quantitative data. Scientists, governments, businesses, and others use statistical analysis to understand a population through sampling.[125] The study of statistics provides insight into a population by studying only a portion of that population; unlike parameters, which represent an entire population, statistics are data points that represent only a portion of the population. Statistics can be used to describe a sample, estimate a population, or to prove/disprove a null hypothesis (e.g., t score, chi-square).[126]

Accurate statistics rely on representative and high quality data (low measurement error). Often, large sample sizes are required to ensure data are representative or to overcome measurement error. High levels of measurement error or unrepresentative samples may skew statistics.

---

125 "The Beginner's Guide to Statistical Analysis: Five Steps and Examples," Scribbr, 2022.

126 College of Physics and Astronomy, "Measurements and Error Analysis: The Uncertainty Measurements," University of North Carolina at Chapel Hill, 2011.

All measurements carry some degree of uncertainty from a variety of sources. Uncertainty analysis or error analysis is the process of evaluating the uncertainty associated with a measurement result. Associated with this analysis are the concepts of precision (a measure of the degree of consistency among independent measurements of the same quantity), accuracy (the closeness of agreement between a measured value and a true or accepted value), and confidence, which accounts for both precision and accuracy.[127] Measures of uncertainty include:

- *Standard error* is the simplest measure of uncertainty, and is displayed as a range around a given statistic (e.g., for a statistic 5.00 +/- 0.05 units, 0.05 is the standard error).
- *Coefficients of variation* place standard error in relation to the underlying statistic by calculating it as a percentage of the estimate.
- *Confidence intervals* use standard error to derive a range in which the value of a parameter is likely to lie.
- *Statistical significance* demonstrates the strength of observed changes or relationships to variables tested.[128]

The strength of using statistics is that they allow researchers to draw insights about large populations from a smaller sample representation of that population. Statistics allow analysts to reach quantitative, concise findings from a large pool of data.[129] As a result, statistics enable researchers to draw generalizable conclusions from the sample. Statistics also facilitate discovering patterns and correlations that may not have been apparent otherwise. Additionally, other research teams can replicate statistical methods to verify results, or use them to re-test hypothesis and discover changes over time.

Weaknesses of using statistics typically relate to the source data and to interpreting the relationships that underpin those data. For instance, statistics can be biased due to small or non-representative sample sizes: analysts cannot accurately generalize from the selected population.[130] Audiences may find it difficult to assess the quality of the statistics because they do not know the provenance of the underlying data. Additionally, while statistics are useful at revealing correlations among variables, some may infer causal relationships if one variable precedes another. Confounding variables may also complicate relationships between variables, something else that might not be immediately apparent to an audience. Analysts with agendas can manip-

---

[127] College of Physics and Astronomy, "Measurements and Error Analysis."

[128] Office of National Statistics, "Uncertainty and How We Measure It for Our Surveys," UK Office of National Statistics, 2022; Dr. Saul McLeod, "What Are Confidence Intervals in Statistics," Simply Psychology, updated 2021.

[129] Joel Best, *Damned Lies and Statistics: Untangling Numbers from the Media, Politicians, and Activists* (Berkeley, CA: University of California Press, 8 May, 2001), 16.

[130] Best, *Damned Lies and Statistics*, 35.

ulate or misrepresent statistics to highlight or undersell a particular aspect that supports a particular perspective.[131] For example, in a population of 100 people, if the prevalence of a specific factor increased from one person to two people, statistics could present that as either a one percent increase or a one-hundred percent increase, depending on phrasing. Human beings are inclined to draw connections between disparate concepts and are bad at understanding abstract numbers, so are prone to accepting this sort of statistical manipulation.[132]

B. Intelligence Analysis

U.S. intelligence community analysts use confidence levels to express uncertainty with regard to their assessments and judgments. Analysts judge their confidence "based on the logic and evidentiary base that underpin it, including the quantity and quality of source material, and their understanding of the topic."[133] Confidence levels allow analytic products to appropriately couch the likely accuracy of the assessments, the level of assumptions that underpin it, and the risks of the assessment being incorrect.[134] Words used to define confidence include:

- Confidence measurements, such as *high confidence* or *low confidence*
- Verbs of uncertainty, such as *judge*, *assess*, *believe*, or *estimate*
- Modal auxiliary verbs, such as *will*, *would*, *could*, *may*, or *might*
- Indirect judgments, such as *indicates*, *reflects*, *reveals*, or *suggests*[135]

Agencies that use confidence levels define them primarily based on the reporting quality. For example, the National Air and Space Intelligence Center (NASIC) defines its confidence levels as follows:

- *High Confidence:* "NASIC has high confidence in this assessment because it is based on direct or high-quality information from multiple sources or from a single highly reliable source, and/or that the nature of the issue makes it possible to render a solid judgment."
- *Medium Confidence:* "NASIC has medium confidence in this assessment because it is information that is indirect, or derived from multiple sources, or from a single reliable source. Information is deemed credible and plausible but not corroborated sufficiently to warrant a higher level of confidence."

---

131 Best, *Damned Lies and Statistics*, 19.

132 Best, *Damned Lies and Statistics*, 20.

133 James A. Clapper, *Sourcing Requirements for Disseminated Analytic Products*, Office of the Director of National Intelligence (Washington, DC: Office of the Director of National Intelligence, 22 January 2015), 3.

134 Clapper, *Sourcing Requirements for Disseminated Analytic Products*, 3.

135 James A. Clapper, *Applying the Intelligence Community's Analytic Tradecraft Standards: A Guide to Best Practices*, 2nd ed. (Washington, DC: Office of the Director of National Intelligence, September, 2015), 19-20.

- *Low Confidence:* "NASIC has low confidence in this assessment because it is based on information that is scant, questionable, or fragmented, and it is difficult to make solid analytic inferences."[136]

The primary strength of confidence levels is that they allow customers to understand intelligence gaps and plan for and mitigate risks. Confidence levels address the five elements of analytic uncertainty: information, intelligence/information gaps, time horizon, historical context, and methodology.[137] When confidence levels follow Intelligence Community guidance and include implications for faulty assumptions and indicators that would alter judgments, the analysis gains rigor and potential future paths for research.[138]

The main weakness of confidence levels is that they can be ambiguous. Some of these terms are poorly and inconsistently defined, with meanings that change based on analyst personality.[139] Moreover, they are susceptible to linguistic ambiguity; mixing and matching different terminology can confuse users and muddle the expression of confidence. Customers may mistake an expression of confidence for an expression of likelihood.[140]

How an analyst derives confidence levels depends on how the analyst derives the assessment itself. Quantitatively-driven assessments may have a mathematical process for defining each confidence level; conversely, qualitative assessments may be more subjective. Given that confidence levels are meant to measure uncertainty, the amount or quality of data that underpins the assessment affects the confidence level.[141] Analysts may use structured analytic techniques and statistical methods to define qualitative and quantitative confidence levels, respectively.

C. Implications of Measuring Confidence

The approaches to measuring confidence reviewed in this chapter, while necessary and useful in their domains, have limited utility to J7's desired context. Because the content that the IDA team expects J7 to synthesize will only partially involve quantitative results, measures of statistical significance will often not apply to the questions being asked.

To the extent that individual studies and analyses producing quantitative results will incorporate statistical significance in their findings, the team finds this approach to understanding confidence will be incorporated in material that J7 may be called up to review. In the case of confidence levels from intelligence analysis, the team found that distinguishing levels of confidence, plus a heuristic to add to the confidence level, was useful. As seen in the synthesis step of the

---

136 National Air and Space Intelligence Center, "Confidence Definitions," National Air and Space Intelligence Center, 2022.

137 Clapper, *Applying the Intelligence Community's Analytic Tradecraft Standards*, 17.

138 Clapper, *Sourcing Requirements for Disseminated Analytic Products*, 3.

139 Clapper, *Applying the Intelligence Community's Analytic Tradecraft Standards*, 20.

140 Clapper, *Sourcing Requirements for Disseminated Analytic Products*, 3.

141 Clapper, *Applying the Intelligence Community's Analytic Tradecraft Standards*, 16.

framework in Chapter 2, the team adapted this approach to confidence levels to how the defense community might understand confidence when making a judgment across multiple sources of information.

5. Knowledge Management Observations & Implications

This chapter discusses some of the knowledge management issues that came up during developing the meta-synthesis framework. For one, the sponsor asked IDA to review JExNET, one of the key databases that the meta-synthesis framework will ultimately rely upon. The Joint community is developing JExNET, which presented IDA with an opportunity to offer feedback on its stage of development. The IDA team also briefly reviewed a knowledge management practice from medicine that is potentially relevant to the implementation of databases designed to feed the style of meta-synthesis framework presented in Chapter 2.

A. JExNET

Given the lack of complete and accurate reporting in data fields, it is difficult to recommend using JExNET as a foundational data source with the analytic framework. However, assuming that scraped data entries are a small fraction of data within JExNET, addressing its limitations as a foundational data source is practical. IDA recommends improving data reporting and report submission. It may also be valuable to knowledge management from adapting certain knowledge management practices from medicine (discussed below in detail).

Because an analytic framework is limited by the availability of quality data, providing analysts with a ready source of quality data is imperative to the usefulness of any framework. J7 tasked IDA with ideas and recommendations for improving JExNET as a way of improving the analytic outputs of the J7. To this end, IDA assessed the entries in JExNET to determine existing data availability that could be used with a meta-synthesis framework. The information about JExNET is a snapshot of the database downloaded on 8 October 2021. JExNET likely has more content relative to when IDA reviewed it; however, this analysis is still informative. The intention was to tie some steps of the meta-synthesis framework to a central repository for all DoD learning events in order to help reduce the burden on the analyst to seek out information sources.

JExNET allows the user to search its records and to filter results through several of its data categories, which is useful for locating specific reports or events related to the same theme or topic. JExNET contains event data from two types of sources: direct inputted data and data it scrapes from other defense databases.

After exporting JExNET data on 8 October 2021, the research team applied a series of sorts and filters to the data before counting events in sub-categories. The data were first filtered by event type to generate a total count of events by event type. Within each event type, the following 25 fields were reviewed to determine whether the field was complete and deemed useful:[142]

- Event Title
- Event Series
- Title Classification
- Event Start Date
- Event End Date
- Event Type
- Event Classification
- Event Sponsor
- Event Objective
- Event Threat
- Event Joint Concept
- Event Office of Primary Responsibility (OPR)
- Additional Event Participants
- Catalyst
- National Defense Strategy (NDS)
- Full [Joint] Function
- [Time] Horizon
- Challenge
- Point of Contact (POC) Name
- POC Organization
- POC Phone Number
- POC NIPR
- POC SIPR
- Contact POC for Report
- Report Uploaded

Using the number of incomplete entries and the total number of entries, IDA calculated the percent of each event type whose data was categorized as "incomplete" or "not useful" per field. The number of "incomplete" or "not useful" data fields were then totaled across all event types in order to identify fields that were most often incomplete and where incomplete data may undermine JExNET's utility to an analyst.

For this assessment, the feature to download all data as an Excel spreadsheet was most useful because it offered a way to check the completeness of the event entries. Unfortunately, some of the key data fields were often empty, reducing the utility of the knowledge platform.

---

[142] Completeness and utility were judgement calls made by the reviewer based on their assessment. If the OPR listed a service, but did not provide an office, the entry was considered incomplete.

Examples of key data fields that were missing information during this particular examination included:

- Reports—Less than 12% of all events had a report uploaded to the platform; the analyst would need to contact the POC for the event or track down the document through other channels. Without the event reports, it is impossible for the analyst to identify findings necessary for synthesis.
- POCs—Less than 75% of the events had a POC listed. Of those, less than 75% had a phone number or NIPR email address listed.
- Joint concepts—Roughly 55% of events were linked to a joint concept.
- National Defense Strategy (NDS) operating problems—Roughly 55% of events were linked to an NDS key operating problem.
- Meta-data that could be used to scope searches—Some fields had so few valid entries for the events that they were essentially useless for searching/sorting purposes. For example, the fields *time horizon* and *challenge* had data for less than 5% of events.

The level of completeness for JExNET fields, such as the *event threat*, *event objective*, *event joint concept* and *contact information*, is left to the discretion of the POC entering information for the event. J7 analysts can complete J7 initiative fields—*event catalyst*, *key operational problem*, *joint function*, *DoD focus*, *service interaction*, and *event challenge*—by working with the event POC.[143] However, in practice, these fields do not always appear complete either.

Several explanations might exist for why the event entries are incomplete in many fields. The majority of data that populate JExNET is scraped from other repositories.[144] However, due to incongruity in the format of reports, the categories for sorting on each individual repository are not necessarily not the same, leading to some fields being incomplete once the data are pulled to JExNET. The incomplete entries may be due to a lack of clear guidance on JExNET about what those fields mean and what a valid entry would be. Other omissions could be due to problems pulling the information in those fields from the various electronic sources, or to incomplete data.

Due to time constraints, IDA did not check the veracity of the data in the data fields. Errors in the information that was in the database could reduce the usefulness of event data. Infor-

---

143 According to "JExNET Reporting Requirements", an unclassified pdf that was acquired by the IDA team during our research.

144 Mr. Charles W. Robinson: *Scraped Data in Jexnet*, 2 February 2022. It is worth noting that Joint Staff J7 acknowledge they are already transitioning away from JExNET to Advana, "a modern data platform that pulls data from hundreds of business systems to make data discoverable, understandable, and usable for [Defense] analytics…" Lissbeth McCrodden, "'Advana' Defense Analytics Platform—Department of Defense," Defense Acquisition University, updated 16 July 2021.

mation like event description (captured in the *event objective* field) could be sparse or misleading, which might lead the analyst to dismiss events germane to the question of interest. Information such as the report link and POC details could be incorrect or outdated, which would require the analyst to track down that information.

B. Medicine

While reviewing frameworks for the work described in Chapter 3, the IDA study team reviewed knowledge management practices in medicine that may be useful to DoD.

One set of practices that may be of interest comes from the Cochrane Collaboration. The Cochrane Collaboration is a consortium of researchers, based in the United Kingdom but with researchers and supporters around the world, that help develop and execute a research approach.[145] One notable aspect of the Cochrane approach is that after a researcher reviews on a particular question, he or she uploads that review itself into the system so that future researchers may see it.

This could be a useful practice for J7: once J7 analysts respond to RFIs, they could upload the results of those RFIs into JExNET so that future analysts can see what was done on a particular topic or question and which data sources fed the conclusions. This could be doubly useful in an environment with high turnover like military staffs.

Another set of possibly useful practices is drawn from the medical research community. The first is the PICO model for developing a good clinical research question prior to starting one's research. The model comprises the elements of Patient/Problem, Intervention, Comparison, Outcome, and a seven-step approach to the "perfect search." For more on the PICO elements, see Chapter 3. DoD could apply a similar model as a first step to developing/planning an event. The PICO model seven-step approach informed the analytic approach that IDA used to develop the Analyst Notebook for this task.

Finally, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), which medical researchers use to ensure complete and transparent reporting of research data and analyses, might serve the J7. PRISMA provides authors with both reporting requirements and a standardized vehicle for reporting. While not a quality assessment instrument, PRISMA gives journal peer reviewers and editors the information in a standard format to help them conduct critical appraisals of published data and analyses. For more on PRISMA, see Appendix E. DoD could refine its JExNET reporting requirements informed by PRISMA to improve the usefulness of event metadata and reports.

[145] "Cochrane," The Cochrane Collaboration, 2022.

6. Exercising JExNET

In addition to the brief audit of JExNET fields documented in Chapter 5, IDA also delved more deeply into the database's content through an effort that directly supported a J7 review of whether concept required capabilities (CRCs) were addressed in JExNET entries. Although the CRC descriptions and the content of the JExNET entries are classified, below is an unclassified overview of the work and its major implications.

As it stands, JExNET contains insufficient content to help the joint community even on the relatively "easy" question of whether database entries may apply to individual CRCs. IDA did not apply the meta-synthesis framework presented in Chapter 2, because the framework is database-agnostic and does not go down to the level of detail for addressing search terms or search strategies for any individual database. However, to the extent that JExNET might play a future role as a major source of information for defense analysts wishing to look across events, the dearth of content in JExNET still has implications for the meta-synthesis framework that IDA developed for this project.

This chapter describes the CRC review IDA conducted for J7 using JExNET, discusses major thoughts about JExNET's utility, and offers potential implications for the meta-synthesis framework.

A. CRC Review

J7 provided IDA with 10–12 individual CRCs for each of the following functions: information advantage, command and control (C2), fires, and contested logistics. J7 also provided IDA with a spreadsheet to document the experiment or project name, whether the experiment or project aligned with each individual CRC, and a brief justification for why it aligned. After finding that targeted searches in JExNET did not return complete datasets, IDA downloaded the data in the JExNET data fields to enable easier content searches, and offered feedback through the spreadsheet and through a PowerPoint briefing. The briefing detailed additional insights on whether JExNET's entries aligned with J7's CRCs, as well as on a percentage of the total items in the database that the team judged as aligning with each individual CRC. Finally, IDA identified examples of high-profile events that were missing from JExNET, even as they were advertised in unclassified sources.

Several challenges arose to determining whether JExNET entries aligned with CRCs. One was how to define the many qualitative terms that appeared in the individual CRCs or in the functions themselves. For example, was C2 technically in everything? Second, it was difficult to determine the cut-off for when an experiment or event was aligned versus not aligned—

understanding had to fit into a binary judgment. This was especially difficult in cases where older events had to be judged against newer concepts. A third factor complicating the examination was the amount of missing information in individual JExNET entries. The level of missing or inconsistently inputted metadata and other content meant that IDA only had information for cursory judgments about whether an event may have aligned with a particular CRC. See Figure 10 for a graphical representation of the down-scoping of the universe of relevant events to events identified as relevant from a search in JExNET.

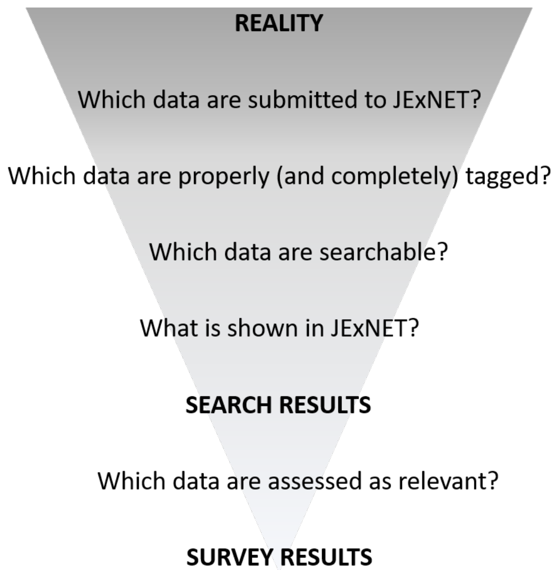

Figure 10. Inverse Relationship between the Universe of Relevant Events and Events Identified as Relevant from JExNET

A fourth challenge was the limited search capabilities within the database itself. IDA often used data about the entries downloaded into a separate spreadsheet, using Excel's capability to search through this downloaded metadata. Some members resorted to manually examining large numbers of individual entries rather than relying on search results. Lastly, a fifth challenge was entries without reports or other attachments—either because they were not included or occasionally because they were available only at a higher level of classification.

J7 also asked IDA to review whether JExNET entries aligned with mission-level concept gaps. *Gaps* could be understood as either capability gaps or gaps in DoD's attention. However, the IDA team determined that trying to use JExNET entries to identify gaps was not feasible for several reasons. Because the entries in JExNET did not appear to fully capture the work done within DoD on the mission-level concepts, no reasonable judgment could be made on whether there were either capability or DoD attention gaps for any particular mission-level concept.

Additionally, the team wrestled with defining gaps when no requirement or standard was provided. The resourcing or attention paid to any DoD concept will have gaps, because it is more or less impossible for DoD to perfectly resource or attend to the development of any single concept. If everything has gaps, it may be more useful to note which CRCs have relatively more gaps; however, any relative assessment of gaps across even part of the defense enterprise went well beyond the scope of the project and the information available within JExNET.

B. Impressions of JExNET's Utility

A significant lesson from IDA's JExNET exercise is the limited utility that this database offers for understanding the range or defense experiments and other activities occurring within DoD. The amount of missing information and the considerable effort it will likely take to populate it to a useful level is not trivial. Many DoD events simply are not in the database. Even events with entries are often sparsely populated.

Another major impression is that even substantially investing in JExNET will not be enough to further organizational learning around defense experimentation, in light of other IDA work for J7 on joint experimentation. Even were the database record and artifacts uploaded into JExNET afterwards more complete, disaggregated information about DoD events, will not equate to understanding concepts mentioned in the database. Nor will a database in and of itself create organizational learning. The entries themselves do not tell DoD whether enough has been done on a concept to support decisions, nor what should be done next to further organizational learning. In short, a knowledge management system may be a key component of a learning organization that can incorporate what has come before to adapt and innovate. But without investing in other elements—such as human capital and learning—knowledge management systems by themselves may not produce the desired knowledge or learning.

Given sufficient time, an analyst might be able to answer certain questions using JExNET. To the extent that continued efforts to populate JExNET may better reflect overall DoD experiments and events over time, this could include the relative amount of attention that might be being paid to certain topics over others. Again, this relies on sufficient progress in populating the database with a representative sample of meaningful content. Another question that could be answered using JExNET is how to find additional detail about an event. (This also depends on data completeness and may be limited by outdated POC information.)

On the other hand, JExNET will not be able to provide several types of information. This includes aggregate DoD focus, because JExNET is unlikely to ever be a complete record of DoD events. JExNET itself will also never be able to determine the quality of the events that are inputted because such a determination requires a level of judgment that no database—even one augmented with machine learning—is yet able to make. Also, JExNET cannot communicate what is meant by concept terms, only how entries are metadata tagged. Especially with the inexact and changing nature of joint terms and concepts, such understanding requires human-level interpretation and mastery of context that narrow artificial intelligence is not capable of at this time.

C. Implications for the Meta-Synthesis Framework

IDA's attempt to exercise JExNET revealed or reinforced several implications for the meta-synthesis framework. Again, the meta-synthesis framework presented in Chapter 2 was not designed around JExNET. Nor does it assume using a single database to arrive at an answer across a portfolio of events or provide database-specific search strategies. Instead, the framework prompts the user to document data sources (including individual databases) and the user's search terms. Yet this exercise revealed implications for the framework even though the IDA team did not try to apply the framework in this particular instance.

1. Organization context and access to POCs will likely improve understanding.

The IDA team queried JExNET to understand how individual events aligned with CRCs; but the effort would have benefited from two things: greater knowledge of the organizational context for those events, and access to POCs who could provide context, help interpret terminology, point to relevant work, and make greater sense of the work that was done. Had the time and opportunity been available for this access, it would have likely improved the team's understanding for the CRC task. With respect to the framework, this experience underscored the importance for better meta-synthesis of Phase I (Frame the Question) tasks—defining key terms, developing points of contact, and reviewing foundational documents.

2. Complete documentation of events is critical to systematic learning or meta-synthesis.

The scale of missing information within JExNET was an object lesson on the importance of DoD having a knowledge management solution that can comprehensively capture the information generated across experiments, studies, analyses, wargames, exercises, demonstrations, and other events. Given JExNET's magnitude of incomplete data, the implication for the framework will be that analysts will need to search multiple other data sources for a substantive meta-synthesis on any question that attempts to gather information across DoD. If databases remain less-than-complete repositories of what the DoD has produced, then human POCs will remain even more important sources of institutional knowledge and brokers who understand where to find additional information on completed events.

3. Understanding meaning and nuance are still human tasks.

The JExNET exercises also underscored the extent to which making meaning out of terms and phrases still depends on human understanding and expertise in DoD operations, experiments, studies, wargames, and other endeavors. Although future improvements to JExNET's search capabilities and the completeness of the information scraped in or manually input are important, many judgments about its content cannot yet be automated. Some judgments cannot be made even by most humans, or even by most humans within DoD, but require a subset of highly specialized individuals with domain understanding. Referring back to the question of whether every event DoD conducts involves C2 illustrates the level of judgement required to interpret terminology. The tendency for DoD terminology to change—and for the thinking around concepts to evolve and mean different things to different stakeholders—further complicates an already difficult task. For the meta-synthesis framework, this points to the need to define key terms, document their own expertise, and evaluate how well the constructs of interest were effectively examined in the actual event.

7. Conclusions and Recommendations

In addition to the meta-synthesis framework that IDA developed for this task are several major conclusions to draw from this project. Creating a usable meta-synthesis framework for J7 highlighted many challenges that DoD faces when attempting to learn across events, and led the research team to a few recommendations for J7 to consider.

A. Major Conclusions

1. Significant challenges exist in synthesizing learning across DoD.

This project clearly demonstrated the magnitude of the challenge that J7 and the broader DoD have in trying to learn across a large number of disparate events. These events were conducted for different purposes, by different organizations, often led by a Service instead of the Joint Staff, with many unstated assumptions, across time, and over an evolving context. The scope, scale, and sheer volume of information generated for and by DoD about individual experiments, studies, exercises, and wargames is unusual for any military in human history.

A meta-synthesis framework or other similar tool is essential to support learning across events and contexts. Regardless of whether the particular framework presented in this report is the right one, J7 needs to adopt some common and repeatable approach to avoid conflating *activity* with *learning*.

2. Unavoidable trade-offs between rigor and speed.

While in a perfect world both rigor and speed would be possible, there are unavoidable trade-offs between the two. The framework and accompanying Analyst Notebook illustrate the effort it would take an analyst or team of analysts to conduct any systematic and rigorous meta-synthesis. The project's experience with evaluating the relevance of existing reports to a range of CRCs clearly illustrate the enormous time pressures that J7 is under to answer even perfunctory questions. There will be no getting around giving up either a substantial amount of rigor, or a substantial amount of speed during any particular look.

A plausible approach is to employ an initial meta-synthesis effort that focuses on speed to get a quick look at the landscape, followed by a more targeted meta-synthesis to look deeply and rigorously at a particular question.

3. Fully document results and make the information available.

The attempt to exercise JExNET against the CRCs demonstrated in painful ways that any database, archive, or knowledge management system will only be as useful as the completeness of its information capture. If JExNET will be a major source of information for J7 in future queries or meta-syntheses, the amount of missing information significantly limits its ability to contribute to understanding right now. Such gaps, if unrecognized, can introduce false and misleading impressions about "what is known." Again, the meta-synthesis framework is not built around JExNET or any other individual data source, but its application needs well-populated and complete archival records of the results of experiments, studies, and wargames.

4. Automation will not solve problems that require judgment and interpretation.

The United States is experiencing significant advances in machine learning and other forms of automation, but the complexity of military judgment required to synthesize disparate information to support joint understanding is highly unlikely to be a fit for machine learning. Machine learning's dependence on clean data at scale, the requirement to have inputs quantifiable to multiply against numeric weights, and brittleness in the face of novelty do not make it a good match for many anticipated types of meta-synthesis for DoD.

5. Organizational learning is the key problem rather than knowledge management or evidence assessment.

A final, key lesson from this effort is that the core issue and challenges that J7 faces are not the conceptually straightforward ones of knowledge management or assessing evidence to support a position. Instead, the challenges are the substantially more difficult ones of supporting institutional memory (what has happened), organizational learning (what do we now know), and organizational sensemaking (what new patterns of behavior the organization should adopt in the face of new contexts). A meta-synthesis framework, while an essential step, would only address the tip of such an iceberg.

B. Recommendations

These are IDA's recommendations for the sponsor's consideration.

1. Continue efforts to populate expected archives of information.

Because any meta-synthesis framework will only be as useful as the availability of information on past events, continuing to populate databases and archives will be foundational to any future meta-synthesis or organizational learning. This may require significant resources and cannot be assumed.

2. Continue to adapt the meta-synthesis framework presented in this report.

IDA staff outside the project attempted to apply the meta-synthesis framework in order to provide feedback on improvements. IDA recommends that the sponsor also use and adapt the framework or its elements to evolve it into a form most useful to J7.

Time constraints limited the sponsor's ability to exercise the framework presented in this report. The best version of the framework will be the one most adapted to J7's staff, organizational processes, and other specific attributes. For best results, emphasize repeated tests and adaptations around a user, changing design to incorporate more and more of a user's ongoing feedback.

3. Consider how to trade-off rigor versus speed.

The framework in this report was built around allowing for a rigorous meta-synthesis. It has some features built in that allow users to scope their efforts, either by skipping some aspects of the framework or by scoping the number and types of events they expect to examine. However, J7 staff may find it too time- and labor-intensive to be practical for J7's timelines, and so may want to consider further tradeoffs in rigor for speed.

As discussed above, one option is to have a faster process that narrows a space for a more rigorous and thorough examination. Another option is to plan a few deep and rigorous meta-syntheses during a year in order to allow the time required a comprehensive series of analytic steps. Yet another option is to apportion quick-turn questions that require speed to J7 analysts and onsite contractors, and longer-turn meta-analyses to IDA or other organizations organized around research.

4. Examine J7's existing structure and resourcing for organizational learning.

One important issue was beyond the scope of this project. IDA's attempt to create a meta-synthesis framework for J7 raised questions about the sponsor's deeper problems in organizational learning. This judgment was also informed by other work that IDA is conducting for both J7 and the Joint Staff writ large. However, it is in J7's interest to look specifically at whether J7 is organized and equipped to support institutional memory, organizational learning, and organizational sensemaking.

Appendix A: Analyst Questionnaire

IDA developed a questionnaire to help the research team understand Joint Staff J7's analytic requirements from the analyst perspective. Responses from two J7 analysts follow.

1. How often do you get requests for information, analysis, or recommendations on specific future capabilities?
2. When you are asked to provide analysis on a specific subject, how far out is the deadline? what percentage of that time are you able to devote to answering the request?
3. Do you have a specific process you use to provide an answer to requests for analysis? What is your process?
4. When looking for specific information about a future technology or capability, where do you look for relevant data? Please list all sources you have used more than once in the past.
5. What are the biggest challenges you encounter in providing accurate answers back to leadership?
6. How well does JExNET support your needs as an analyst? What would you change about JExNET?
7. Does your organization have standard operating procedures (SOPs) in place for analyzing developing capabilities and their likelihood of future success?

Respondent A, JS J7

1. **How often do you get requests for information, analysis, or recommendations on specific future capabilities?** *These types of RFIs were mostly suspended while the branch stood-down. On past exposure I believe the branch was handling about 6-8 high level RFIs for senior leaders a year. Various smaller tasks (maybe 10) were also accomplished. This is anecdotal as I was not working in the branch during the time.*
2. **When you are asked to provide analysis on a specific subject, how far out is the deadline? what percentage of that time are you able to devote to answering the request?** *Unknown. From what I've heard suspense timelines are generally about a week even on big projects. Again anecdotal.*
3. **Do you have a specific process you use to provide an answer to requests for analysis? What is your process?** *No, no existing process.*
4. **When looking for specific information about a future technology or capability, where do you look for relevant data? Please list all sources you have used more than once in the past.** *JExNET. POCs listed as sources in JExNET. Interviews with SMEs.*
5. **What are the biggest challenges you encounter in providing accurate answers back to leadership?** *Unknown (not done this)*
6. **How well does JExNET support your needs as an analyst? What would you change about JExNET?** *Meta-data tagging for improved search functionality. Starting conditions and assumptions clearly listed. Epoch easily accessible.*
7. **Does your organization have SOPs in place for analyzing developing capabilities and their likelihood of future success?** *Not at this time.*

Respondent B, JS J7

1. **How often do you get requests for information, analysis, or recommendations on specific future capabilities?** *None. With that being said, as the Joint Wargaming and Experimentation Division (JWED) stands up the experimentation branch and we try to align Service Title 10 activities to align with Joint Warfighting Concept gaps linked to Concept Required Capabilities, I suspect there will be more incidents for RFI.*
2. **When you are asked to provide analysis on a specific subject, how far out is the deadline? what percentage of that time are you able to devote to answering the request?** *Experimentation Branch is just standing up. The Force Integration Functional Capability Board (FCB) did ask about who within Joint Staff J7 is working Non-Kinetic M&S and analysis tool efforts. I did answer that RFI and other POC's I work with. With that being said, as JWED stands up the experimentation branch and we try to align Service Title 10 activities to align with Joint Warfighting Concept gaps linked to Concept Required Capabilities, I suspect there will be more incidents for RFI.*
3. **Do you have a specific process you use to provide an answer to requests for analysis? What is your process?** *Low fidelity processes have been established for internal communications. We report to our leadership through a Weekly Activity Report and Quads for their information. Outreach to external organizations are either through Joint Staff Task Management Tool (TMT) taskers; tasks via chain of command or sporadic emails. With that being said, as JWED stands up the experimentation branch and we try to align Service Title 10 activities to align with Joint Warfighting Concept gaps linked to Concept Required Capabilities, I suspect there will be more incidents for RFI.*
4. **When looking for specific information about a future technology or capability, where do you look for relevant data? Please list all sources you have used more than once in the past.** *Defense Technology Information Centers (DTIC), Office of the Under Secretary of Defense (OUSD) for Research and Experimentation (R&E) data call, Joint Staff TMT.*
5. **What are the biggest challenges you encounter in providing accurate answers back to leadership?** *For now – we have a very high pers tempo on standing up the Experimentation Branch and that greatly limits time availability to spend much time on answers back to leadership.*
6. **How well does JExNET support your needs as an analyst? What would you change about JExNET?** *I have limited use of JExNET.*
7. **Does your organization have SOPs in place for analyzing developing capabilities and their likelihood of future success?** *They are being established.*

Appendix B: Existing Analytic Frameworks

These slides summarize IDA's initial review of existing analytic frameworks and were presented to the Joint Staff J7 on 28 July 2021.

Appendix C: Analyst Notebook

The Analyst Notebook is a stand-alone document that includes empty, phased worksheets, and introductory front matter for each worksheet.

Appendix D: Senior Leader Guide

A. Introduction

IDA's meta-synthesis framework provides Joint Staff analysts asked to respond to requests for information (RFIs) related to status updates of future capabilities and technologies with the tools necessary to methodically, transparently, and consistently address them. The framework is effective regardless where the analyst sits, how much experience the analyst has, how much time the analyst is given to respond, or what classification level they conduct their research. The resulting worksheets provide senior leaders with the documentation necessary to understand the analyst's confidence in their answer/recommendation and, if necessary, assess the validity of the answer/recommendation themselves.

The framework is made up of five phases. While the framework itself is simple, the requirements in each phase are complex. The five phases are—

1. Frame the question
2. Select the data
3. Assess the data
4. Synthesize the data
5. Communicate

Each phase generates a worksheet, which, if properly used, documents an analyst's research process and aids in synthesis of data across events and measurement of confidence in findings/recommendations. Figure D-1Figure depicts the overall break-down of each phase.

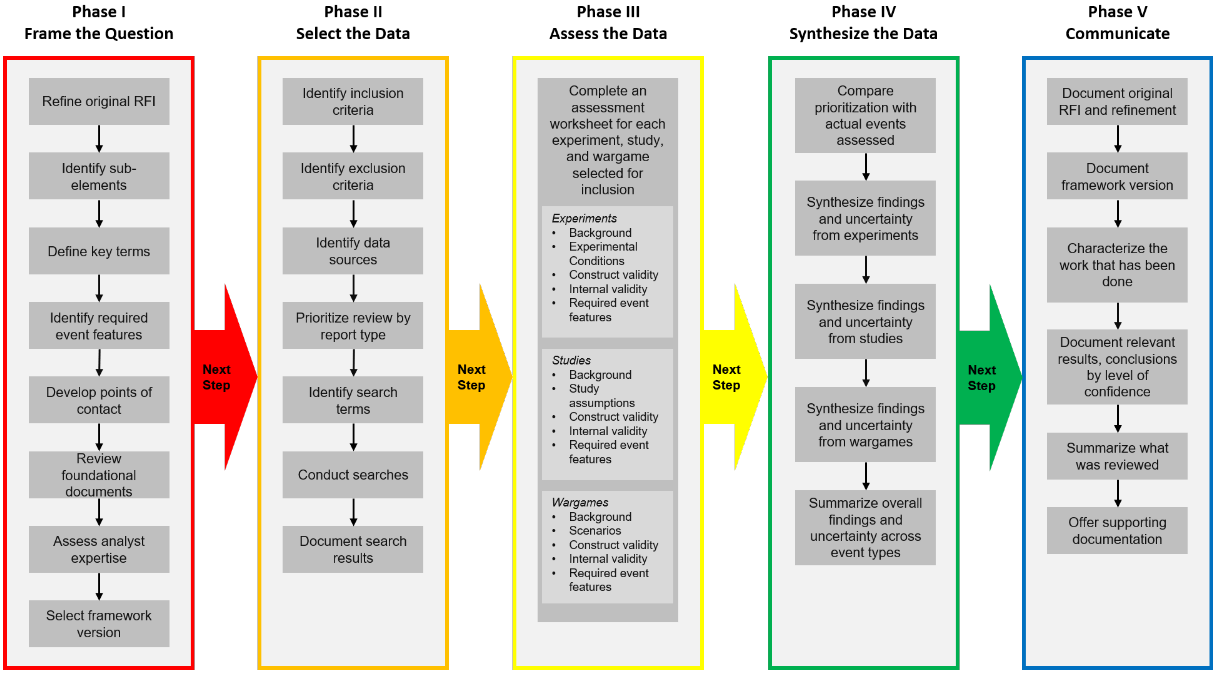

Figure D-1. Meta-Synthesis Framework Flow Chart

In response to every RFI, the analyst will provide senior leaders with the Phase V (Communicate) worksheet. The senior leader can ask for the supporting worksheets (Phases I–IV) and a Status report.

B. How to Use the Framework Worksheets

The Phase V (Communicate) worksheet provides the senior leader not only with a summary of findings and recommendations, but also sheds light on the research conducted to develop the findings and recommendations. This measure of transparency is often enough for a senior leader to know whether or not to trust the findings and recommendations. If the senior leader still requires more evidence to trust the findings and recommendations, they should review the Phase I through Phase IV worksheets as well. Regardless of the depth of review, the senior leader should consider the following items.

Phase I

- Does the scope of the question researched by the analyst align with the RFI posed by the senior leader? If not, can any of the findings/recommendations be applied to RFI originally posed?
- Were there any event features that were not marked as required but should have been?
- Where there any key offices or foundational documents not consulted in the review that would limit the relevance of the findings/recommendations?

- Did reviewers lack necessary qualifications to conduct the review properly?

Phase II

- Do the criteria make sense? Are there missing or unnecessary inclusion or exclusion criteria that limit the relevance of the findings/recommendations?
- Were there any data sources not included that would limit the relevance of the findings/recommendations?
- Were there any key search terms missing that could limit the relevance of the findings/recommendations?
- Was the number and distribution of events reviewed sufficiently inclusive?

Phase III (Experiments)

- Should any of the included experiments been excluded because of the intent, constructs used, inputs, or experimental conditions?
- If a shortened assessment was conducted, do you trust the construct and internal validity for every experiment included in the assessment?
- Do you agree with the analyst's assessment of applicability for each experiment included?

Phase III (Studies)

- Should any of the included experiments been excluded because of the intent, constructs used, inputs, or assumptions?
- If a shortened assessment was conducted, do you trust the construct and internal validity for every study included in the assessment?
- Do you agree with the analyst's assessment of applicability for each study included?

Phase III (Wargames)

- Should any of the included wargames been excluded because of the intent, constructs used, inputs, or scenarios used?
- If a shortened assessment was conducted, do you trust the construct and internal validity for every wargame included in the assessment?
- Do you agree with the analyst's assessment of applicability for each wargame included?

Phase IV

- Do you agree with the analyst's characterization of experiments, studies, and wargames?

- Do you agree with the analyst's summary of the most relevant/credible results/conclusions?
- Were there any sources of uncertainty not considered by the analyst?
- Do you agree with the analyst's assessment of the representation of required event features?
- Do you agree with the analyst's assessment of confidence in the results/conclusions?

C. How to Use the Status Report

The status report provides senior leaders with a snapshot of research activity and a status update for all RFIs in a given period of time. Analysts provide a tally of RFIs pending per phase (see Figure D-2).

Figure D-2. Status Report Summary of RFIs

D. Limitations

The amount of time will vary, but a typical systematic review takes approximately six to eighteen months.[146] The phased worksheets are not a substitute for a comprehensive review of the literature. Rather they offer the defense analyst a way to quickly, systematically, and transparently identify and review key events in order to provide senior leaders conclusions and recommendations they can trust.

Locating and retrieving relevant event reports for such a review is challenging, but is critical to the success of defense analyst responding to a senior leader request. No framework can make up for a lack of, lack of access to, or an inability to identify relevant data. This can only be remedied by providing defense analysts with access to repositories that represent the full breadth of relevant events. Additionally, those repositories much provide complete and transparent access to event data through. One way to ensure that is through robust reporting requirements.

[146] Valerie Smith et al., "Methodology in Conducting a Systematic Review of Systematic Reviews of Healthcare Interventions," *BMC Medical Research Methodology* 11, no. 15 (3 February 2011),

Appendix E: QUOROM & PRISMA Best Practices

The *QUality Of Reporting Of Meta-analyses* (QUOROM) and *Preferred Reporting Items for Systematic reviews and Meta-Analyses* (PRISMA) statements are related methodologies designed to improve the quality of meta-analysis and, in the case of PRISMA, systematic reviews. The former, developed in 1996 and published in 1999, focuses solely on the meta-analysis of randomized, controlled studies. The QUOROM statement comprised 21 reporting requirements related to searches, selection, validity assessment, data abstraction, study characteristics, and quantitative data synthesis. The PRISMA statement, a refinement of the QUOROM statement and published in 2009, is relevant to randomized and non-randomized controlled studies. Focused on ensuring transparent and complete reporting, the checklist was expanded to 27 items. The authors of the PRISMA statement recognize that it may require modification or the incorporation of additional items to be even more broadly applicable.[147] This section will focus on the PRISMA statement as the current state of practice in the medical research community.

PRISMA comprises four phases—identification, screening, eligibility, and inclusion—and a 27-point reporting checklist. During the *identification phase*, reports are identified through database searches and other means. The number of results returned for each search should be documented and reported. During the *screening phase*, search results are deconflicted and the resulting set are screened against exclusion criteria. The number of reports after deduplication as well as the number of reports screened and excluded should be documented and reported. During the *eligibility phase*, reports are assessed for eligibility based on inclusion criteria. The number of reports assessed for eligibility and excluded (along with the reason for exclusion) should be documented and reported. During the final phase—*inclusion*—the number of studies included in qualitative synthesis (i.e., systematic review) and quantitative synthesis (i.e., meta-analysis) should be documented and reported.[148] See Figure G-1 (adopted from Moher et al.) for a schematic of the types of data documented in each phase.

The systematic review with or without a meta-analysis is then reported out per a 27-point checklist (see Table G-1, adopted from Moher et al.). The checklist ensures transparency and completeness in reporting; thus, improving the utility of the review or analysis.[149]

---

[147] Moher et al., "The QUOROM Statement."; Liberati et al., "The PRISMA Statement: Explanation and Elaboration." For more info, see PRISMA, "PRISMA."

[148] David Moher et al., "Preferred Reporting Items for Systematic Reivews and Meta Analyses: The PRISMA Statement," *BMJ* 339 (21 July 2009): 8, https://doi.org/10.1136/bmj.b2535,

[149] Liberati et al., "The PRISMA Statement: Explanation and Elaboration," 3-22.

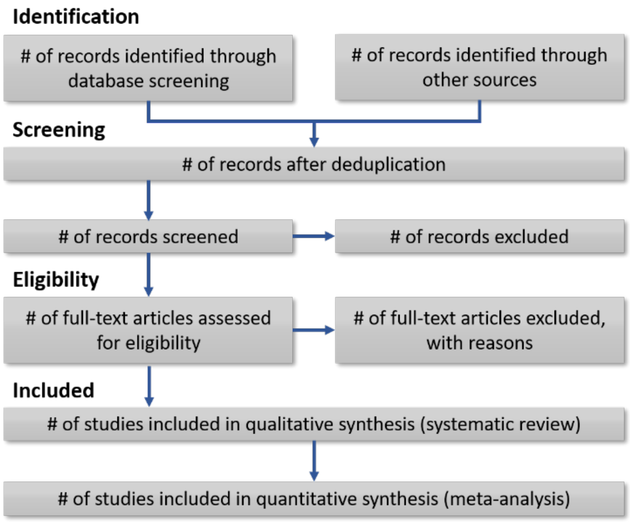

Figure G-1. PRISMA Flow Diagram

Table G-1. Checklist of items to include when reporting a systematic review (+/- meta-analysis)

Section	Topic	Item #
Title	Title	1
Abstract	Structured summary	2
Introduction	Rationale	3
	Objectives	4
Methods	Protocol and registration	5
	Eligibility criteria	6
	Information sources	7
	Search	8
	Study selection	9
	Data collection process	10
	Data items	11
	Risk of bias in individual studies	12
	Summary measures	13
	Synthesis of results	14
	Risk of bias across studies	15
	Additional analyses	16
Results	Study selection	17
	Study characteristics	18
	Risk of bias within studies	19
	Results of individual studies	20
	Synthesis of results	21
	Risk of bias across studies	22
	Additional analyses	23
Discussion	Summary of evidence	24
	Limitations	25
	Conclusions	26
Funding	Funding	27

Appendix F: References

Adcock, Robert, and David Collier. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95, no. 3 (17 January 2002). https://doi.org/10.1017/S0003055401003100. http://www.jstor.org/stable/3118231.

Ahn, EunJin, and Hyun Kang. "Introduction to Systematic Review and Meta-Analysis." *Korean Journal of Anesthesiology* 71, no. 2 (April 2018). https://doi.org/10.4097/kjae.2018.71.2.103.

APA Dictionary of Psychology. Washington, DC: American Psychological Association, 2022.

Amison, Major General Darrell. *Defence Experimentation for Force Development Handbook, Version 2.* UK Ministry of Defence (Bristol, UK: UK Ministry of Defence, January 2021).

"Modeling and Simulation Glossary: Data." US Army, 2022, accessed 10 February, 2022, https://www.ms.army.mil/contact_us.html.

Atkins, Salla, Simon Lewin, Helen Smith, Mark Engel, Atle Fretheim, and Jimmy Volmink. "Conducting a Meta-Ethnography of Qualitative Literature: Lessons Learnt." *BMC Medical Research Methodology* 8, 21 (April 2008). https://doi.org/https://doi.org/10.1186/1471-2288-8-21. https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-8-21.

"The Beginner's Guide to Statistical Analysis: Five Steps and Examples." Scribbr, 2022, accessed 27 January, 2022, https://www.scribbr.com/category/statistics/.

Best, Joel. *Damned Lies and Statistics: Untangling Numbers from the Media, Politicians, and Activists.* Berkeley, CA: University of California Press, 8 May, 2001.

Bhattacherjee, Anol. *Social Science Research: Principles, Methods, and Practices.* Textbooks Collection. 2nd ed. Hillsborough County, FL: University of South Florida, 2012. https://digitalcommons.usf.edu/cgi/viewcontent.cgi?article=1002&context=oa_textbooks.

Black, Max. *Critical Thinking: An Introduction to Logic and Scientific Method.* Aukland, NZ: Muriwai Books, 2018. https://ebookcentral.proquest.com/lib/tiso-ebooks/detail.action?docID=5835506.

"Quality Resource, Glossary: Design of Experiments (Doe)." American Society for Quality, 2022, accessed 31 January, 2022, https://asq.org/quality-resources/design-of-experiments.

"Live Science / What Is Science." Future US, Inc., 2017, accessed 23 June, 2021, https://www.livescience.com/20896-science-scientific-method.html.

Britten, Nicky, Rona Campbell, Catherine Pope, Jenny Donovan, Myfawny Morgan, and Roisin Pill. "Using Meta-Ethnography to Synthesise Qualitative Research: A Worked Example." *Journal of Health Service Reseach & Policy* 7, no. 4 (October 2002 2002): 209-215. https://journals.sagepub.com/doi/10.1258/135581902320432732.

Brown, Sharon A., Sandra L. Upchurch, and Gayle J. Acton. "A Framework for Developing a Coding Schem for Meta-Analysis." *Western Journal of Nursing Research* 25, no. 2 (1 March 2003). https://doi.org/10.1177/0193945902250038. https://doi.org/10.1177/0193945902250038.

Burris, Andrew B. "A Qualitative and Quantitative Assessment of Readiness for Organizational Change Literature." MS Engineering Management, Air University, Air Force Institute of Technology, 2008 (AFIT/GEM/ENV/08-J01).

Caldwell, John, Katherine Fisher, Thomas Greenwood, Robert Richbourg, Robert Stimets, and Cameron Swann. *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics.* Institute for Defense Analyses (Alexandria, VA: Institute for Defense Analyses, 1 June 2021).

Card, Noel A. *Applied Meta-Analysis for Social Science Research.* Methodology in the Social Sciences. Edited by Todd D. Little. New York, NY: Guilford Press, 18 July, 2011.

Carter, Nancy, Denise Bryant-Lukosius, Alba DiCenso, Jennifer Blythe, and Alan J. Neville. "The Use of Triangulation in Qualitative Research." *Oncology Nursing Forum* 41, no. 5 (September 2014). https://doi.org/10.1188/14.ONF.545-547. https://pubmed.ncbi.nlm.nih.gov/25158659/.

"Analytical Study Design in Medical Research." Peseapro Scientific Services Ltd., Updated 3 February 2014, 2017, accessed 23 June, 2021, https://www.manuscriptedit.com/scholar-hangout/overview-different-analytical-study-designs-medical-research/.

"Analytical Study Design in Medical Research: Cohort Study." Peseapro Scientific Services Ltd., Updated 3 March 2014, 2017, accessed 23 June, 2021, https://www.manuscriptedit.com/scholar-hangout/tag/retrospective-cohort-study/.

Chang, Welton, Elissabeth Berdini, David R. Mandel, and Philip E. Tetlock. "Restructuring Structured Analytic Techniques in Intelligence." *Intelligence and National Security* 33, no. 3 (10 November 2017). https://doi.org/10.1080/02684527.2017.1400230. https://www.tandfonline.com/doi/full/10.1080/02684527.2017.1400230.

"How Does the Cross-Sectional Research Method Work?" About, Inc., Updated 10 October 2019, 2019, accessed 23 June, 2021, https://www.verywellmind.com/what-is-a-cross-sectional-study-2794978.

"The Pros and Cons of Longitudinal Research." About, Inc., Updated 2 May 2020, 2020, accessed 23 June, 2021, https://www.verywellmind.com/what-is-longitudinal-research-2795335.

Clapper, James A. *Applying the Intelligence Community's Analytic Tradecraft Standards: A Guide to Best Practices.* 2nd ed. Washington, DC: Office of the Director of National Intelligence, September, 2015.

Clapper, James A. *Sourcing Requirements for Disseminated Analytic Products.* Office of the Director of National Intelligence (Washington, DC: Office of the Director of National Intelligence, 22 January 2015). https://www.dni.gov/files/documents/ICD/ICD%20206.pdf.

"Cochrane." The Cochrane Collaboration, 2022, accessed 20 January, 2022, https://www.cochrane.org/.

"Measurements and Error Analysis: The Uncertainty Measurements." University of North Carolina at Chapel Hill, 2011, accessed 27 January, 2022, https://www.webassign.net/question_assets/unccolphysmechl1/measurements/manual.html.

"Construct Validity." In *International Encyclopedia of Public Health*, edited by Harald Kristian Heggenhougen. Cambridge, MA: Elsevier / Academic Press, 2008. https://www.sciencedirect.com/topics/social-sciences/construct-validity.

Coulthart, Stephen J. "An Evidence-Based Evaluation of 12 Core Structured Analytic Techniques." *International Journal of Intelligence and Counterintelligence* 30, no. 2 (8 February 2017). https://doi.org/10.1080/08850607.2016.1230706. https://www.tandfonline.com/doi/full/10.1080/08850607.2016.1230706.

Cowles, Henry M. *The Scientific Method: An Evolution of Thinking from Darwin to Dewey.* Cambridge, MA: Harvard University Press, 2020. https://ebookcentral.proquest.com/lib/tiso-ebooks/detail.action?docID=6126334.

Cox, D.R., and Nancy Reid. *The Theory of the Design of Expeirments.* Crc Monographs on Statistics and Applied Probability. London, UK: Chapman & Hall, 2000.

Creswell, John W., and Dana L. Miller. "Determining Validity in Qualitative Inquiry." *Theory into Practice* 39, no. 3 (24 June 2010). https://doi.org/10.1207/s15430421tip3903_2. https://www.tandfonline.com/doi/pdf/10.1207/s15430421tip3903_2.

Davis, Paul K. *Generalizing Concepts and Methods of Verification, Validation, and Accredidation (Vv&a) for Military Simulations.* RAND Corporation (Santa Monica, CA: RAND Corporation, 1992). https://www.rand.org/pubs/reports/R4249.html.

Department of Defense. *Test & Evaluation Management Guide.* 6th ed. Washington, DC: Department of Defense, August, 2016. https://www.dau.edu/guidebooks/Shared%20Documents/Test_and_Evaluation_Mgmt_Guidebook.pdf.

"Organisational Learning and Learning Capability: What They Are, Why They Matter, and How to Foster Them." CQ Net, Updated 10 November 2019, 2019, accessed 25 January, 2022, https://www.ckju.net/en/dossier/organisational-learning-and-learning-capability-what-they-are-why-they-matter-and-how-foster-them.

Dixon-Woods, Mary, Shona Agarwal, David Jones, Bridget Young, and Alex Sutton. "Synthesising Qualitative and Quantitative Evidence: A Review of Possible Methods." *Journal of Health Services Research & Policy* 10, no. 1 (2005): 45-53.

Drost, Ellen A. "Validity and Reliability in Social Science Research." *Education Research and Perspectives* 38, no. 1 (June 2011). https://www3.nd.edu/~ggoertz/sgameth/Drost2011.pdf.

"Duke University Medical Center Library & Archives: Systematic Reviews." Duke University, Updated 18 January 2022, 2022, accessed 7 May, 2021, https://guides.mclibrary.duke.edu/sysreview.

Encyclopedia Britannica. "Scientific Method." In *Encyclopedia Britannica*. Chicago, IL: Encyclopedia Britannica, Inc., 2021.

Farrugia, Patricia, Bradley A. Petrisor, Forough Farrokhyar, and Mohit Bhandari. "Practical Tips for Surgical Research: Research Questions, Hypotheses, and Objectives." *Canadian Journal of Surgery* 53, no. 4 (August 2010). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2912019/.

Finckh, Axel, and Martin R. Tramer. "Primer: Strengths and Weaknesses of Meta-Analysis." *Nature Clinical Practice Rheumatology* 4, no. 3 (March 2008). https://doi.org/10.1038/ncprheum0732. https://www.nature.com/articles/ncprheum0732.pdf.

Finlayson, Kenneth, and Annie Dixon. "Qualitative Meta-Synthesis: A Guide for the Novice." *Nurse Researcher* 15, no. 2 (2008).

France, Emma F., Isabelle Uny, Nicola Ring, Ruth L. Turley, Margaret Maxwell, Edward A.S. Duncan, Ruth G. Jepson, Rachel J. Roberts, and Jane Noyes. "A Methodological Systematic Review of Meta-Ethnography Conduct to Articulate the Complex Analytical Phases." *BMC Research Methodology* 19, no. 1 (February 2019 2019). https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-019-0670-7.

Glass, Gene V. "Primary, Secondary, and Meta-Analysis of Research." *Educational Researchers* 5, no. 10 (November 1976). https://doi.org/10.2307/1174772. https://www.jstor.org/stable/1174772.

"Construct Validity—Simple Definition, Statistics Used." Statistics How To, Updated 30 December 2014, 2022, accessed 10 February, 2022, https://www.statisticshowto.com/construct-validity/.

"Restrospective Study: Case-Control and Case-Series." Statistics How To, Updated 22 August 2016, 2021, accessed 23 June, 2021, https://www.statisticshowto.com/retrospective-study/#retrospective.

Gopalakrishna, S., and P. Ganeshkumar. "Systematic Reviews and Meta-Analysis: Understanding the Best Evidence in Primary Healthcare." *Journal of Family Medicine and Primary Care* 2, no. 1 (January 2013). https://doi.org/10.4103/2249-4863.109934. https://pubmed.ncbi.nlm.nih.gov/24479036/.

Gordon, Howard R.D. *Meta-Analysis Research: A Potential Choice for Cte Researchers and Consumers.* Marshall University, Department of Adult & Technical Education (Huntington, WV: Marshall University, 26 September 2007). https://files.eric.ed.gov/fulltext/ED498416.pdf.

Gottdenker, Paula. "Francesco Redi and the Fly Experiments." *Bulletin of the History of Medicine* 53, no. 4 (Winter 1979). https://www.jstor.org/stable/pdf/44450950.

Guyatt, Gordon, John Cairns, and David Churchill. "Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine." *Journal of the American Medical Association* 268, no. 17 (4 November 1992). https://doi.org/10.1001/jama.1992.03490170092032. https://jamanetwork.com/journals/jama/article-abstract/400956.

Haque, Mohammad Nazmul, and Pablo Moscato. "From Ensemble Learning to Meta-Analytics: A Review on Trends in Business Applications." Chap. 1 In *Business and Consumer Analytics: New Ideas*, edited by Pablo Mostcato and Natalie de Vries. New York, NY: Springer Publishing, 2019.

Harden, Angela. "Mixed-Mehtods Systematic Reviews: Integrating Quantitative and Qualtiative Findings." *FOCUS*, no. 25 (2010). https://ktdrr.org/ktlibrary/articles_pubs/ncddrwork/focus/focus25/Focus25.pdf.

Heale, Roberta, and Dorothy Forbes. "Understanding Triangulation in Research." *Evidence-Based Nursing* 16, no. 98 (August 2013). https://ebn.bmj.com/content/16/4/98.

"Introduction to Study Design: Cross-Sectional Studies." Public Health Action Support Team CIC, 2020, accessed 23 June, 2021, https://www.healthknowledge.org.uk/e-learning/epidemiology/practitioners/introduction-study-design-css.

Herber, Oliver Rudolf, and Julie Barroso. "Lessons Learned from Applying Sandelowski and Barroso's Approach for Synthesising Qualitative Research." *Qualitative Research* 20, no. 4 (2020): 414-431. https://doi.org/10.1177/1468794119862440.

Heuer Jr., Richards J., and Randolph H. Pherson. *Structured Analytic Techniques for Intelligence Analysis.* Washington, DC: CQ Press, 2015.

Hunt, Morton. *How Science Takes Stock: The Story of Meta-Analysis.* New York, NY: Russell Sage Foundation, 1997.

Hunter, John L., and Frank L. Schmidt. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings.* 2nd ed. Thousand Oaks, CA: Sage Publishing, 2004.

"Industrial-Organizational Psychology: Organizational Sensemaking." iResearchNet.com, 2022, accessed 25 January, 2022, https://psychology.iresearchnet.com/industrial-organizational-psychology/organizational-development/organizational-sensemaking/.

Ipsos. "Big Data." In *Ipsos Encyclopedia*. Paris, France: Ipsos Group, 2022. https://www.ipsos.com/en/ipsos-encyclopedia-big-data.

Irwin, MG Lew. *Joint War Gaming & Experimentation*. Washington, DC: Department of Defense, 2019.

Jensen, Kathy A. *Seven Steps to the Perfect Pico Search: Evidence-Based Nursing Practice.* EBSCO Health (Ipswich, MA: EBSCO Health, 2018). https://www.ebsco.com/sites/g/files/nabnos191/files/acquiadam-assets/7-Steps-to-the-Perfect-PICO-Search-White-Paper_0.pdf.

Joint Staff Joint Force Development. *Joint Operations*. Washington, DC: Department of Defense, 2018.

Joint Staff Joint Force Development. *Joint Planning*. Washington, DC: Department of Defense, 2020.

Kass, Richard A. *The Logic of Warfighting Experiments.* The Future of Command and Control. Washington, DC: Command and Control Research Program, 2006. http://www.dodccrp.org/files/Kass_Logic.pdf.

"Controlled Experiments." Khan Academy, Inc., 2021, accessed 23 June, 2021, https://www.khanacademy.org/science/high-school-biology/hs-biology-foundations/hs-biology-and-the-scientific-method/a/experiments-and-observations.

"The Scientific Method." Khan Academy, Inc., 2021, accessed 23 June, 2021, https://www.khanacademy.org/science/high-school-biology/hs-biology-foundations/hs-biology-and-the-scientific-method/a/the-science-of-biology.

Kudesia, Ravi S. "Organizational Sensemaking." In *Oxford University Research Encyclopedia*. Oxford, UK: Oxford University Press, 26 April 2017. https://oxfordre.com/psychology/view/10.1093/acrefore/9780190236557.001.0001/acrefore-9780190236557-e-78.

Labbe, Paul, and Richard A. Nunes Vaz. "Guide for Understanding and Implementing Defense Experimentation (Guidex), Version 1.1." *ResearchGate* (February 2006). https://doi.org/10.13140/2.1.4937.6648. https://www.researchgate.net/publication/265315028.

Lau, Francis. "Methods for Correlational Studies." Chap. 12 In *Handbook of Ehealth Evaluation: An Evidence-Based Approach*, edited by Francis Lau and Craig Kuziemsky. Victoria, BC: University of Victoria, 2016.

"Oxford Bibliographies, Panel Studies." Oxford University Press, Updated 26 August 2013, 2021, accessed 23 June, 2021, https://www.oxfordbibliographies.com/view/document/obo-9780199756384/obo-9780199756384-0108.xml.

"Scientific Method." Biology Dictionary, Updated 6 November 2020, 2020, accessed 23 June, 2021, https://biologydictionary.net/scientific-method/.

Liberati, Alessandro, Douglas G. Altman, Jennifer Tetzlaff, Cynthia Mulrow, Peter C. Gotzsche, John P.A. Ioannidis, Mike Clarke*, et al.* "The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration." *PLOS Medicine* 6, no. 7 (21 July 2009). https://doi.org/https://doi.org/10.1136/bmj.b2700. https://www.sciencedirect.com/science/article/pii/S0895435609001802.

Lipps, Oliver. "Panel Surveys: Advantages and Disadvantages Compared with Repeated Cross-Sectional Surveys." *FORS Guide*, no. 14 (2021). https://doi.org/10.24449/FG-2021-00014. https://forscenter.ch/fors-guides/panel-surveys-advantages-and-disadvantages-compared-with-repeated-cross-sectional-surveys/.

Lipsey, Mark W., and David B. Wilson. *Practical Meta-Analysis.* Applied Social Research Methods. Vol. 49, Thousand Oaks, CA: Sage Publishing, 2000.

Ludvigsen, Mette S., Elisabeth O. C. Hall, Gabriele Meyer, Liv Fegran, Hanne Aagaard, and Lisbeth Uhrenfeldt. "Using Sandelowski and Barroso's Meta-Synthesis Method in Advancing Qualitative Evidence." *Qualitative Health Research* 26, no. 3 (2016): 320-329. https://doi.org/10.1177/1049732315576493.

MacGill, Markus, and Deborah Weatherspoon. "What Is a Systematic Review in Research?". *Medical News Today* (25 February 2019). https://www.medicalnewstoday.com/articles/281283.

"Redi's Experiment." Dreamstime LLC, 2021, accessed 23 June, 2021, https://www.dreamstime.com/to-test-hypothesis-francesco-redi-placed-fresh-meat-open-containers-as-expected-rotting-meat-accracted-flies-image178248228#res26615551

Maruyama, Geoffrey, and Carey S. Ryan. *Research Methods in Social Relations.* 8th ed. Hoboken, NJ: Wiley-Blackwell, 2014.

Mayville Jr., LTG William C. *Guidance for Developing and Implementing Joint Concepts.* Department of Defense (Washington, DC: Department of Defense, 17 August 2016). https://www.jcs.mil/Portals/36/Documents/Library/Instructions/CJCSI%203010.02E.pdf?ver=2018-10-26-171040-997.

"'Advana' Defense Analytics Platform—Department of Defense." Defense Acquisition University, Updated 16 July 2021, 2021, accessed 31 January, 2022, https://media.dau.edu/media/t/1_bqg1km0j.

"What Are Confidence Intervals in Statistics." Simply Psychology, Updated 2021, 2019, accessed 27 January, 2022, https://www.simplypsychology.org/confidence-interval.html

Mediratta, Anuj. *Big Data: Terms, Definitions, and Applications.* Dell Technologies (Round Rock, TX: 2015). https://education.emc.com/content/dam/dell-emc/documents/en-us/2015KS_Mediratta-Big_Data_Terms,_Definitions_and_Applications.pdf.

Meline, Timothy. "Selecting Studies for Systematic Review: Inclusion and Exclusion Criteria." *Contemporary Issues in Communication Science and Disorders* 33 (Spring 2006). https://pubs.asha.org/doi/pdf/10.1044/cicsd_33_S_21.

Merriam-Webster. Online ed. Springfield, MA: Merriam-Webster Inc., 2022.

Miller, Delbert C., and Neil J. Salkind. *Handbook for Research Design and Social Movement.* 6th ed. Thousand Oaks, CA: Sage Publications, January, 2002.

Moher, David, Deborah J. Cook, Susan Eastwood, Ingram Olkin, Drummond Rennie, and Donna F. Stroup. "Improving the Quality of Reports of Meta-Analyses of Randomized Controlled Trials: The QUOROM Statement." *Lancet* 354, no. 9193 (27 November 1999). https://doi.org/10.1016/s0140-6736(99)04149-5. https://pubmed.ncbi.nlm.nih.gov/10584742/.

Moher, David, Alessandro Liberati, Jennifer Tetzlaff, and Douglas G. Altman. "Preferred Reporting Items for Systematic Reivews and Meta Analyses: The PRISMA Statement." *BMJ* 339 (21 July 2009). https://doi.org/10.1136/bmj.b2535. https://www.bmj.com/content/bmj/339/bmj.b2535.full.pdf.

Moravscik, Andrew. *Transprency in Qualitative Research.* Thousand Oaks, CA: SAGE Publications, 2019. https://www.princeton.edu/~amoravcs/library/TransparencyinQualitativeResearch.pdf.

"Confidence Definitions." National Air and Space Intelligence Center, 2022, accessed 1 February, 2022, http://www.nasic.ic.gov/vps/presgen?corid=1718.

"NCI Library > Longitudinal Cohort Study." National Cancer Institute, 2021, accessed 23 June, 2021, https://www.cancer.gov/publications/dictionaries/cancer-terms/def/longitudinal-cohort-study.

"NCI Library > Retrospective Cohort Study." National Cancer Institute, 2021, accessed 23 June, 2021, https://www.cancer.gov/publications/dictionaries/cancer-terms/def/retrospective-cohort-study.

"Scientific Writing Portal: Creating Research Questions." Norwegian University of Life Sciences (NMBU), 2022, accessed 31 January, 2022, https://nmbu.instructure.com/courses/2280/pages/creating-research-questions.

Noblit, George W. , and R. Dwight Hare. *Meta-Ethnography: Synthesizing Qualitative Studies.* Newbury Park, California: Sage Publications, 1988.

"Oracle Cloud Infrastructure: Big Data." Oracle, 2022, accessed 10 February, 2022, https://www.oracle.com/big-data/what-is-big-data/.

"Uncertainty and How We Measure It for Our Surveys." UK Office of National Statistics, 2022, accessed 27 January, 2022, https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/uncertaintyandhowwemeasureit#what-is-uncertainty.

Office of the Secretary of Defense for Research and Engineering. *Department of Defense Experimentation Guidebook: Prototypes and Experiments.* Department of Defense (Washington, DC: Department of Defense, 18 November 2021). https://www.dau.edu/tools/Lists/DAUTools/Attachments/381/DoD%20Experimentation%20Guidebook%20v2.0%202021.pdf.

Patino, Cecilia Maria, and Juliana Carvalho Ferreira. "Internal and External Validity: Can You Apply Research Study Results to Your Patients?". *Journal of Brasilian Pneumology* 44, no. 3 (May-June 2018). https://doi.org/10.1590/S1806-37562018000000164. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6188693/.

Perla, Peter P. *The Art of Wargaming: A Guide for Professionals and Hobbyists.* Annapolis, MD: Naval Institute Press, 1990.

Perla, Peter P. "Why Wargaming." MORS Certificate in Wargaming Course, Virtual, Military Operations Research Society, 2022.

Perla, Peter P., Web Ewell, Christoher Ma, Justin Peachey, Jeremy Sepinsky, and Basil Tripsas. "Rolling the Iron Dice: From Analytic Wargaming to the Cycle of Research." *War on the Rocks* (21

October 2019). https://warontherocks.com/2019/10/rolling-the-iron-dice-from-analytical-wargaming-to-the-cycle-of-research/.

Pournelle, Phillip. "Can the Cycle of Research Save American Military Strategy?". (October 18 2019). Accessed January 2022. https://warontherocks.com/2019/10/can-the-cycle-of-research-save-american-military-strategy/.

"PRISMA: Transparent Reporting of Systematic Reviews and Meta-Anlayses." PRISMA, 2021, accessed 31 January, 2022, http://www.prisma-statement.org/.

"Evidence Based Medicine (Emb) Toolkit: Discuss Emb: What Is the Best Evidence and How to Find It." BMJ Best Practice, 2022, accessed 25 April, 2022, https://bestpractice.bmj.com/info/us/toolkit/discuss-ebm/what-is-the-best-evidence-and-how-to-find-it/.

Robinson, Mr. Charles W. *Jexnet Knowledge Platform, Version 1.5.* Joint Staff J7 (Suffolk, VA: Joint STaff J7, 9 October 2020).

Rubin, Allen, and Earl R. Babbie. *Research Methods for Social Work.* Brooks/Cole Empowerment Series. 8th ed. Boston, MA: Cengage, 2013.

Salkind, Neil J. "Exclusion Criteria." In *Encyclopedia of Research Design*. Thousand Oaks, CA: SAGE Publishing, 2010. https://methods.sagepub.com/reference/encyc-of-research-design/n137.xml.

Salkind, Neil J. "Inclusion Criteria." In *Encyclopedia of Research Design*. Thousand Oaks, CA: SAGE Publishing, 2010. https://methods.sagepub.com/reference/encyc-of-research-design/n183.xml.

Sandelowski, Margarete, and Julie Barroso. "Classifying the Findings in Qualitative Studies." *Qualitative Health Research* 13, no. 7 (September 2003): 905-923. https://doi.org/10.1177/1049732303253488.

Sandelowski, Margarete, and Julie Barroso. *Handbook for Synthesizing Qualitative Research.* New York: Springer, 2007.

"Steps of the Scientific Method." Science Buddies, 2021, accessed 23 June, 2021, https://www.sciencebuddies.org/science-fair-projects/science-fair/steps-of-the-scientific-method.

Setia, Maninder Singh. "Methodology Series, Module 1: Cohort Studies." *Indian Journal of Dermatology* 61, no. 1 (January-February 2016). https://doi.org/10.4103/0019-5154.174011. https://pubmed.ncbi.nlm.nih.gov/26955090/.

Setia, Maninder Singh. "Methodology Series, Module 2: Case-Control Studies." *Indian Journal of Dermatology* 61, no. 2 (March-April 2016). https://doi.org/10.4103/0019-5154.177773. https://pubmed.ncbi.nlm.nih.gov/27057012/.

Setia, Maninder Singh. "Methodology Series, Module 3: Cross-Sectional Studies." *Indian Journal of Dermatology* 61, no. 3 (May-June 2016). https://doi.org/10.4103/0019-5154.182410. https://pubmed.ncbi.nlm.nih.gov/27293245/.

Smith, Valerie, Declan Devane, Cecily M. Begley, and Mike Clark. "Methodology in Conducting a Systematic Review of Systematic Reviews of Healthcare Interventions." *BMC Medical Research Methodology* 11, no. 15 (3 February 2011). https://doi.org/10.1186/1471-2288-11-15. https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-11-15.

Sobrido Prietoa, María , and José María Rumbo-Prieto. "The Systematic Review: Plurality of Approaches and Methodologies." *Enfermeria Clinica* 28 (January 2018): 387-393. https://doi.org/https://doi.org/10.1016/j.enfcli.2018.08.008.

"Social Psychology Research Methods: Meta-Analysis." iResearchNet.com, 2022, accessed 26 May, 2021, https://psychology.iresearchnet.com/social-psychology/social-psychology-research-methods/meta-analysis/.

Suchmacher, Mendel, and Mauro Geller. "Study Type Determination." Chap. 1 In *Practical Biostatistics*, edited by Mendel Suchmacher and Mauro Geller. Waltham, MA: Elsevier Academic Press, 2012.

"Systems Engineering Guide: Verification and Validation of Simulation Models." MITRE, 2021, accessed 31 January, 2022, https://www.mitre.org/publications/systems-engineering-guide/se-lifecycle-building-blocks/other-se-lifecycle-building-blocks-articles/verification-and-validation-of-simulation-models.

"The Scientific Method (Archived)." The Darwin Trust of Edinburgh, 2019, accessed 23 June, 2021, http://archive.bio.ed.ac.uk/jdeacon/statistics/tress2.html.

"Library Guides: Systematic Reviews—Inclusion and Exclusion Criteria." University of Melbourne, Updated 18 October 2021, 2021, accessed 31 January, 2022, https://unimelb.libguides.com/c.php?g=492361&p=3368110.

"Library Guides: Conducting a Literature Review—Benefits of Conducting a Literature Review." University of Northern Florida, Updated 13 July 2021, 2022, accessed 31 January, 2022, https://libguides.unf.edu/litreview/benefits.

US Coast Guard. *Verification, Validation, and Accredidation (Vv&a) of Models and Simulations (M&S).* US Department of Homeland Security (Washington, DC: US Department of Homeland Security, 22 December 2006). https://media.defense.gov/2017/Mar/13/2001710628/-1/-1/0/CI_5200_40.PDF.

Whittemore, Robin, and Kathleen Knafl. "The Integrative Review: Updated Methodology." *Journal of Advanced Nursing* 52, no. 5 (December 2005). https://doi.org/10.1111/j.1365-2648.2005.03621.x. https://pubmed.ncbi.nlm.nih.gov/16268861/.

Wojton, Heather, Kelly M. Avery, Laura J. Freeman, Samuel H. Parry, Gregory S. Whittier, Thomas H. Johnson, and Andrew C Flack. *Handbook on Statistical Design & Analysis Techniques for Modeling & Simulation Validation.* Institute for Defense Analyses (Alexandria, VA: Institute for Defense Analyses, February 2019).

Zhang, Cha, and Yunqian Ma. *Ensemble Machine Learning : Methods and Applications.* New York, NY: Springer Publishing, 2012. http://ebookcentral.proquest.com/lib/tiso-ebooks/detail.action?docID=883900.

Appendix G: Glossary

IDA developed a glossary of key terms relevant to a discussion of analytic frameworks and military campaigns of learning. It is not an exhaustive glossary but should provide context for a discussion of this report.

accreditation – The official certification that a model or simulation and its associated data are acceptable for use for a specific purpose."[150]

advanced warfighting experiment – A US term, usually meaning defense experimentation tackling complex transformational issues on a large scale.[151]

aggregated data – Aggregated data is data which is summarized, usually in a way that supports statistical analysis.[152]

analysis of variance (ANOVA) – The use of statistical models to test hypothesis about the differences (variation) between group means in a data sample. ANOVA provides a way to extend a T-Test to more than one group of data.[153]

analytic wargame – Analytic wargames typically employ command and staff officers to plan and execute a military operation, often with some form of constructive simulation adjudicating outcomes between turns (sometimes overnight).[154]

artificial intelligence (AI) – A discipline that includes many sub-disciplines, all of which are focused on enabling machines to reason or act intelligently as measured by human rationality. The discipline includes big data analytics, natural language processing, machine perception, intelligent data mining, speech generation, machine learning, automated planning, knowledge representation theory, and other problem-specific areas.[155]

---

[150] Department of Defense, *Test & Evaluation Management Guide*.

[151] Paul Labbe and Richard A. Nunes Vaz, "Guide for Understanding and Implementing Defense Experimentation (Guidex), Version 1.1," *ResearchGate* (February 2006), https://doi.org/10.13140/2.1.4937.6648, https://www.researchgate.net/publication/265315028. GuideX features the following caveat: "Although the term "warfighting experimentation" is used by all of the TTCP nations, AG-12 has found that its meaning is not consistent across the nations and it is not helpful in communicating GUIDEx's message. For example: in some countries it is taken and used to imply experimentation only in warfighting scenarios, rather than in all military operations; in some it is taken to mean only experimentation involving the presence of warfighters in their operational role; and in some it is taken to cover all empirical military analyses, not just experimentation as described in this guide."

[152] John Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*, Institute for Defense Analyses (Alexandria, VA: Institute for Defense Analyses, 1 June 2021).

[153] Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.

[154] Labbe and Nunes Vaz, "Guidex, Ver. 1.1."

[155] Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.

Bayes Rule – An equation for calculating the probability that something is true if something potentially related to it is true. One of the values of this theorem is that it allows easier updates to probability estimates when new data becomes available.[156]

Bayes Theorem – See Bayes Rule

bias value – The weighted sum of inputs provided to a neuron in a neural network. It is typically used to shift the weighted sums into a specific range (like an offset from the origin =). In this use, it is more correctly referred to as the "bias term."[157]

big data – Large, complex data or data sets of greater volume, velocity, and variety than could previously be practically managed with traditional data processing software. Big data can be used to reach conclusions that are more complete, and with greater confidence, to derive insight that may not have been otherwise possible.[158] Big data can come from a variety of sources, including (but not limited to) sensors, audiovisual products, online communications, and online user activity tracking.[159]

causation – The action of causing some specific outcome.[160]

coding – The process whereby raw data are transformed into a standardized form that is suitable for machine processing and analysis.[161]

conceptual approach – The underlying concept of operation that helps define a specific approach to problem solving or activity.[162]

confidence interval – The range of values within which a population parameter is estimate to lie. For example, a survey may show 40% of a sample favoring candidate A. Although the best estimate of the support existing among all voters would also be 40%, we would not expect it to be exactly that. Therefore, we might compute a confidence interval within which the actual percentage of the population probably lies (35% to 45%). It is necessary to specify a *confidence level* in connection with every confidence interval.[163]

---

[156] Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.

[157] Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.

[158] OCI, "Oracle Cloud Infrastructure: Big Data," Oracle, 2022, accessed 10 February, 2022, https://www.oracle.com/big-data/what-is-big-data/.

[159] Ipsos, "Big Data," in *Ipsos Encyclopedia* (Paris, France: Ipsos Group, 2022). https://www.ipsos.com/en/ipsos-encyclopedia-big-data; Anuj Mediratta, *Big Data: Terms, Definitions, and Applications*, Dell Technologies (Round Rock, TX, 2015), https://education.emc.com/content/dam/dell-emc/documents/en-us/2015KS_Mediratta-Big_Data_Terms,_Definitions_and_Applications.pdf.

[160] Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.

[161] Allen Rubin and Earl R. Babbie, *Research Methods for Social Work*, 8th ed., Brooks/Cole Empowerment Series, (Boston, MA: Cengage, 2013).

[162] Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.

[163] Rubin and Babbie, *Research Methods for Social Work*.

confidence level – The estimated probability that a population parameter lies within a given confidence interval. Thus, we might be 95% confident that between 35% and 45% of all voters favor candidate A.[164]

construct – An abstract concept identified for measurement.[165]

construct validity – Construct validity is the extent to which the measure 'behaves' in a way consistent with theoretical hypotheses and represents how well scores on the instrument are indicative of the theoretical construct;[166] one way to test the validity of a test; it's used in education, the social sciences, and psychology; it demonstrates that the test is actually measuring the construct it claims it's measuring.[167]

constructive simulation – The closed-loop force-on-force simulations employed by the modeling and simulation and military operational research communities. Once designers choose the initial parameters, start the simulation, and run it to completion, there is no human intervention in the play of the simulation. Analytic wargames sometimes use such simulations but the human intervention is essentially between runs. In some quarters, the term constructive simulation is used to describe large scale command post exercise (CPX) drivers such as JTLS. In GUIDEx the term is *not* used in this way and such tools would be considered to be HITL simulations.[168]

control variable – One can prevent the effects of a specific identifiable extraneous variable from clouding the results of an experiment by holding the value of this extraneous variable constant, e.g., all selected subjects have the same level of training, C. A variable that is thus held constant is called a control variable. Similarly, in a multiple regression equation, specific extraneous independent variables, e.g., C, can be held constant or statistically controlled in examining the impact of A on B, the dependent variable. The resulting correlation is then called a partial correlation between A and B controlling for C.[169]

correlation – This term describes the behavior of two classes of data that vary in agreement with each other; if one class of data has increasing values the correlated data increases in value as well. As an example, the price paid for a home might correlate with the home buyer's income. A correlation coefficient is a measure of how well two data sets correlate. A correlation coefficient of 1 indicates perfect correlation. Two classes of data can also

---

[164] Rubin and Babbie, *Research Methods for Social Work*.

[165] Maruyama and Ryan, *Research Methods in Social Relations*.

[166] "Construct Validity," in *International Encyclopedia of Public Health*, ed. Harald Kristian Heggenhougen (Cambridge, MA: Elsevier / Academic Press, 2008). https://www.sciencedirect.com/topics/social-sciences/construct-validity.

[167] Stephanie Glen, "Construct Validity—Simple Definition, Statistics Used," Statistics How To, updated 30 December 2014, 2022, accessed 10 February, 2022, https://www.statisticshowto.com/construct-validity/.

[168] Labbe and Nunes Vaz, "Guidex, Ver. 1.1."

[169] Labbe and Nunes Vaz, "Guidex, Ver. 1.1."

have negative correlations (one increases and the other decreases) which could be characterized by negative correlation coefficient values. A fundamental point is that correlation and causation are two different concepts and one does not necessarily imply the other.[170]

data – A representation of facts, concepts, or instructions in a formalized manner suitable for communication, interpretation, or processing by humans or by automatic means.[171]

data cleaning – A multi-step process of analyzing data to detect errors or corrupt data and then applying a variety of techniques to repair that data.[172]

data cleansing – See Data Cleaning

data conditioning – This activity is closely related and usually including the data cleansing activity that optimizes the movement and management of data in order to protect it and increase its productivity. It typically includes managing the data to store and protecting them while transforming them to ready it for intended purposes.[173]

data integration – The process of combining data from different sources into a single, unified view. Integration begins with the ingestion process, and includes steps such as cleansing, schema mapping, and transformation. This process is sometimes called Data Unification.[174]

data science – The art and science of analyzing large amounts of data to provide actionable intelligence and insights to decision makers.[175]

deductive reasoning – A method of logical reasoning based on formal systems of logic and which always results in a necessarily true conclusion when the premises (propositions) are true. This type of reasoning is used in rule-based systems. The most famous example of deductive reasoning is the syllogism (a major premise, a minor premise and the conclusion):[176]

(major premise) All men are mortal
(minor premise) Socrates is a man
(conclusion) Therefore Socrates is mortal

defense experiment/experimentation – The application of the experimental method to the solution of complex defense capability development problems, potentially across the full spectrum of conflict types, such as warfighting, peace-enforcement, humanitarian relief and peace-keeping.[177]

---

170 Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.

171 Army Modeling and Simulation Office, "Modeling and Simulation Glossary: Data," US Army, 2022, accessed 10 February, 2022, https://www.ms.army.mil/contact_us.html.

172 Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.

173 Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.

174 Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.

175 Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.

176 Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.

177 Labbe and Nunes Vaz, "Guidex, Ver. 1.1."

defense experimentation – Controlled and directed activities designed to discover new information about an idea or concept, test a hypothesis or validate a solution or choice in support of Force Development.[178]

demonstration – scripted and orchestrated activities that minimize the risk that the solution demonstrated will fail. They are primarily intended to display a solution's military utility in specific operational environments to people unfamiliar with the technology or concept or to senior leaders responsible for making decisions regarding its employment, deployment, or acquisition in order to garner support for the technology or concept."[179]

development experimentation – Designed to help develop an idea, concept or capability, to mature it to a point where it can be validated. A refining process designed to test, at an early stage, whether the idea, concept or capability will deliver against its expectations.[180]

discovery experimentation – Designed to build understanding, to inform the development of potential solutions, to introduce novel ideas, concepts and capabilities in the early stages and to help refine the question to be addressed. This type of experiment should be conducted against a broad hypothesis to ensure it has the freedom to explore but bounded for it to be achievable. (DFDB, October 2019)[181]

ecological fallacy – Erroneously drawing a conclusion about individuals based solely on the observation of groups.[182]

exercise – A military maneuver or simulated wartime operation involving planning, preparation, and execution, that is carried out for the purpose of training and evaluation;183 Current-oriented event primarily supporting commanders' training and readiness objectives for a training audience; can be leveraged as a venue for experiments.[184]

exercise exploitation (intrusive) – Exploiting a training exercise for experimental or other non-training-related purposes where there is a need for some deliberate and pre-agreed intervention into the running of the exercise.[185]

---

178 Major General Darrell Amison, *Defence Experimentation for Force Development Handbook, Version 2*, UK Ministry of Defence (Bristol, UK: UK Ministry of Defence, January 2021).

179 Office of the Secretary of Defense for Research and Engineering, *Department of Defense Experimentation Guidebook: Prototypes and Experiments*, Department of Defense (Washington, DC: Department of Defense, 18 November 2021), https://www.dau.edu/tools/Lists/DAUTools/Attachments/381/DoD%20Experimentation%20Guidebook%20v2.0%202021.pdf.

180 Amison, *Defence Experimentation for Force Development Handbook*.

181 Amison, *Defence Experimentation for Force Development Handbook*.

182 Rubin and Babbie, *Research Methods for Social Work*.

183 Joint Staff Joint Force Development, Joint Operations, JP 3.0 (Washington, DC: Department of Defense, 2018).

184 MG Lew Irwin, Joint War Gaming & Experimentation, (Washington, DC: Department of Defense, 2019).

185 Labbe and Nunes Vaz, "Guidex, Ver. 1.1."

exercise exploitation (passive) – Exploiting a training exercise for experimental or other non-training-related purposes where there is no interference in the running of the exercise and only unobtrusive, passive data collection will be performed.[186]

experiment – An empirical means of establishing cause-and-effect relationships through the manipulation of independent variables and measurement of dependent variables in a controlled environment;[187] Future-oriented testing of proposed capabilities to evaluate their ability to enable preferred ways of operating in anticipated future operating environments.[188]

external validity – The extent to which one can generalize a study's findings to settings and populations beyond the study conditions.[189]

face validity – That quality of an indicator that makes it seem a reasonable measure of some variable.[190]

factorial validity – Whether the number of constructs and the items that comprise those constructs on a measurement scale are what the research intends.[191]

field exercise – Any exercise using live simulation.[192]

field experiment – A defense experiment based on live simulation;[193] Future-oriented testing of proposed capabilities to evaluate their ability to enable preferred ways of operating in anticipated future operating environments.[194]

hand-crafted knowledge – Hand-crafted knowledge systems, primarily knowledge-based systems, expert systems, and planning systems, "enable reasoning over narrowly defined problems" but handle uncertainty poorly and have no capacity to learn.[195]

heuristics – Improved, more informed search strategies rely on heuristics to reduce the number of options that have to be considered. Heuristics are "rules of thumb" that provide a best-guess solution before evaluating a breadth of options. If heuristics are applied to a route planning system, the software would look at the transitions (roads) from the starting point and select the one that seemed to move closest to the goal by estimating the straight-line distance between the intermediate state and goal state. The search would continue to apply the heuristic of taking the transition (road) that moved closest to the goal until identifying a path from start to goal. Many types of heuristics exist in AI systems. The main point of heuristics is to reduce the problem-solving effort by using estimates to guide

---

186 Labbe and Nunes Vaz, "Guidex, Ver. 1.1."
187 Labbe and Nunes Vaz, "Guidex, Ver. 1.1."
188 Irwin, Short Joint War Gaming & Experimentation.
189 Rubin and Babbie, *Research Methods for Social Work*.
190 Rubin and Babbie, *Research Methods for Social Work*.
191 Rubin and Babbie, *Research Methods for Social Work*.
192 Labbe and Nunes Vaz, "Guidex, Ver. 1.1."
193 Labbe and Nunes Vaz, "Guidex, Ver. 1.1."
194 Irwin, Short Joint War Gaming & Experimentation.
195 Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.

choices during the search for a solution. Heuristic use can lead to incorrect solutions, but this is not always the case.[196]

inductive reasoning – A process of logically arriving at a conclusion about a member of a class from examining a number of other class members. This method of reasoning does not necessarily produce true conclusions. As an example, suppose it is known that Tom's car has 4 tires and that Bill's car has 4 tires. Inductive reasoning allows the conclusion that all cars have 4 tires. Inductive reasoning is closely related to learning and is the type of reasoning used in many neural network implementations (e.g., classification systems).[197]

integrated analysis and experimentation campaign – A planned sequence of related defense experiments, studies and/or analytical activities designed to advance the understanding of a military force development problem. Within the campaign, the key role of an experiment is to generate some linkage between cause-and-effect. Integrated analysis and experimentation campaigns can mitigate the risks associated with particular analytical techniques using the strengths inherent in other methods and thus build validity in the campaign outcomes.[198]

integrative review – A research review that summarizes empirical or theoretical literature to provide a more comprehensive understanding of a particular phenomenon. While there are commonalities among all review methods (meta-analyses, systematic reviews, qualitative reviews, integrative reviews), each has a distinct purpose, sampling frame, definition, and type of analysis.[199]

internal validity – The degree to which an effect observed in an experiment was actually produced by the experimental stimulus and not the result of other factors.[200]

joint concept – Identifies a current or future military challenge and proposes a solution to improve the joint force's ability to address that military challenge. A joint concept may also propose new ways to employ the joint force based on future technology.[201]

Joint Experimentation Network, JExNET – A knowledge sharing community of interest supporting alignment of experimentation across the Defense enterprise and web-based

---

196 Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.

197 Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.

198 Labbe and Nunes Vaz, "Guidex, Ver. 1.1."

199 Robin Whittemore and Kathleen Knafl, "The Integrative Review: Updated Methodology," *Journal of Advanced Nursing* 52, no. 5 (December 2005), https://doi.org/10.1111/j.1365-2648.2005.03621.x, https://pubmed.ncbi.nlm.nih.gov/16268861/.

200 Rubin and Babbie, *Research Methods for Social Work*.

201 LTG William C. Mayville Jr., *Guidance for Developing and Implementing Joint Concepts*, Department of Defense (Washington, DC: Department of Defense, 17 August 2016), GL-4, https://www.jcs.mil/Portals/36/Documents/Library/Instructions/CJCSI%203010.02E.pdf?ver=2018-10-26-171040-997.

knowledge platform that serves as a repository for event data and results reports;[202] used in this report as a reference to the knowledge platform.

Likert scale – A type pf composite measure developed by Rensis Likert to improve the levels of measurement in social research that uses standardized response categories in survey questionnaires. "Likert items" use such response categories as *strongly agree*, *agree*, *disagree*, and *strongly disagree*. Such items may be used to construct true Likert scales, and can also be used to construct other types of composite measures.[203]

linear regression – A technique to look for a linear relationship (that is, one where the relationship between two varying amounts can be expressed with an equation that represents a straight line on a graph) by starting with a set of data points and computing the "least squares" line: the line that has, on an x-y graph, the smallest possible sum of squared distances to the actual data point y values.[204]

live simulation – Simulations of military operations in a live environment with actual military units and real military equipment and operational prototypes, with only weapon effects being simulated.[205]

measure of effectiveness – A measure that describes the influence or benefit of a concept within its operational context.[206]

measure of performance – A measure that describes the influence or benefit of a concept in terms of its internal structure, characteristics and behavior.[207]

meta-analysis – The use of statistical techniques in a systematic review to integrate the results of included studies; a component of a systematic review;[208] a procedure for calculating the average strength of association between variables (i.e., the mean effect size) across previously completed research studies in a particular field.[209]

metadata – A type of data that provides information about related data or databases; data about data. As an example, modern digital photographs include EXIF (derived from the Exchangeable Image File Format standard) data such as date and time, even though that data is not visible in the image. Metadata is data about data.[210]

---

202 Mr. Charles W. Robinson, *Jexnet Knowledge Platform, Version 1.5*, Joint Staff J7 (Suffolk, VA: Joint STaff J7, 9 October 2020).
203 Rubin and Babbie, *Research Methods for Social Work*.
204 Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.
205 Labbe and Nunes Vaz, "Guidex, Ver. 1.1."
206 Labbe and Nunes Vaz, "Guidex, Ver. 1.1."
207 Labbe and Nunes Vaz, "Guidex, Ver. 1.1."
208 Moher et al., "The PRISMA Statement," 1.
209 Rubin and Babbie, *Research Methods for Social Work*.
210 Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.

meta-ethnography – Unlike *meta-analysis*, the underlying logic of meta-ethnography is interpretation rather than aggregation. Although meta-ethnography includes an element of aggregation, its synthesis function is interpretation. Rather than pooling findings from studies, key concepts are translated within and across studies, and the synthesis product is a new interpretation. In a meta-ethnography, the synthesis goal is to achieve a greater level of understanding or conceptual development than can be found in any individual empirical study.[211]

normalization (of input data) – The process of converting an actual range of values into a standard range of values, typically -1 to +1 or 0 to 1. It is not strictly necessary to normalize input data for training a neural network, but in practice, normalized data appears to speed the training process.[212]

observational study – An objectively-observed, practical event which does not involve the deliberate or purposeful manipulation of independent variables to establish cause-and-effect relationships. Observational studies may be used to establish associative or correlative relationships.[213]

qualitative analysis – The non-numerical examination and interpretation of observations for the purpose of discovering underlying meanings and patterns of relationships.[214]

qualitative research method – A research method that emphasizes depth f understanding and the deeper meanings of human experience and that aim to generate theoretically richer albeit more tentative observations. commonly used qualitative methods include participant observation, direct observation, and unstructured or intensive interviewing.[215]

quantitative analysis – The numerical representation and manipulation of observations for the purpose of describing and explaining the phenomena that those observations reflect.[216]

quantitative research method – A research method that typically seeks to produce precise and generalizable findings. Studies using quantitative methods typically attempt to formulate all or most of their research procedures in advance, and then try to adhere precisely to those procedures with maximum objectivity as data are collected.[217]

regression – A procedure for fitting any kind of model (e.g., line, curve) to a set of data points in a way that produces the best fit of the model with the data. “Regression” is sometimes used as a shorthand for “linear regression” although they are not identical.[218]

---

[211] Harden, “Mixed-Methods Systematic Reviews.”

[212] Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.

[213] Labbe and Nunes Vaz, “Guidex, Ver. 1.1.”

[214] Rubin and Babbie, *Research Methods for Social Work*.

[215] Rubin and Babbie, *Research Methods for Social Work*.

[216] Rubin and Babbie, *Research Methods for Social Work*.

[217] Rubin and Babbie, *Research Methods for Social Work*.

[218] Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.

reliability – The quality of a measurement method that suggests that the same data would have been collected each time in repeated observations of the same phenomenon. Not the same as *validity*.[219]

replication – Generally, the duplication of a study to expose or reduce error or the reintroduction or withdrawal of an intervention to increase the internal validity of a quasi-experiment or single-case design evaluation.[220]

scenario – A description of the area, the environment, means, objectives and events related to a conflict or a crisis during a specified time frame suited for satisfactory study objectives and the problem analysis directives.[221]

search strategies – Search strategies include many techniques to consider and weigh options when identifying and evaluating potential solutions to a problem. The most basic search strategies are known as "brute force" strategies because they consider every possible option. Search strategies often underpin planning systems when the problem can be represented as a series of states with allowed transitions between states (e.g., a state graph) and the overall solution is a specific traversal of the graph—from the start state to the goal state, with intermediate states in between. A simple example of this type of planning system is route planning. Software represents road intersections as states and the roads between the intersections as the legal transitions. At the first intersection, the software selects an option and runs until it either reaches the desired destination or dead-ends. It then tries another full route. Basic search strategies can ensure each legal transition is only considered once (e.g., a breadth-first search or a depth-first search). After running through all possible routes, the answer is identified as the "best" route in terms of time, distance, or some other parameter. Given that brute force strategies consider the full range of options, they can be slow.[222]

seminar wargame – A structured discussion between experts in several fields to elicit opinions and judgments from them, and to increase understanding. It is more structured than brainstorming (or seminars), but is not normally supported by any kind of simulation (like analytic wargames).[223]

social desirability bias – A source of systematic measurement error involving the tendency of people to say to do things that will make them or their reference group look good.[224]

statistic – A summary description of a variable in a sample.[225]

---

[219] Rubin and Babbie, *Research Methods for Social Work*.
[220] Rubin and Babbie, *Research Methods for Social Work*.
[221] Labbe and Nunes Vaz, "Guidex, Ver. 1.1."
[222] Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.
[223] Labbe and Nunes Vaz, "Guidex, Ver. 1.1."
[224] Rubin and Babbie, *Research Methods for Social Work*.
[225] Rubin and Babbie, *Research Methods for Social Work*.

systematic review – A research review that combines the evidence of multiple studies. They often include the statistical methods of meta-analysis if primary studies meet the assumptions required for meta-analyses. If primary studies cannot be combined statistically, a narrative analysis is done in conjunction with vote counting or other quasi-statistical approaches.[226]

systematic review – a review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyze data from the studies included in the review.[227]

table top exercise (TTX) – structured wargames where warfighters work through scenarios to discover and define capability gaps and their boundaries, and where initial insights into the value of proposed solutions to those gaps, across the full DOTMLPF-P spectrum, is discussed.[228]

trial – A single opportunity to observe the effect of a particular treatment in an experiment. ii. Often used to mean the same as test, as in Field Trials, Flight Trials, Sea Trials, etc.[229]

triangulation – The use of more than one imperfect data collation alternative to which each option is vulnerable to different potential sources of error.[230]

unit of analysis – The *what* or *whom* being studied.[231]

unstructured data – Data that does not conform to a specific schema or structure. Typical examples include email, video photographs, and digital audio files.[232]

validation – The process of determining the degree to which a model and its associated data accurately represent the real world from the perspective of the model's intended uses. ("Has the system been built correctly?")[233]

validity – A term to describe a measure that accurately reflects the concept that it is intended to measure. The ultimate validity of a measure can never be proven, but we can still agree to its relative validity, content validity, construct validity, internal validity, and external validity.[234]

verification – The process of determining that a model implementation and its associated data accurately represent the developer's conceptual description and specifications. ("Has the correct system been built?")[235]

---

226 Whittemore and Knafl, "The Integrative Review."
227 Moher et al., "The PRISMA Statement," 1.
228 Office of the Secretary of Defense for Research and Engineering, *Dod Experimentation Guidebook*, 9.
229 Labbe and Nunes Vaz, "Guidex, Ver. 1.1."
230 Rubin and Babbie, *Research Methods for Social Work*.
231 Rubin and Babbie, *Research Methods for Social Work*.
232 Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.
233 Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.
234 Rubin and Babbie, *Research Methods for Social Work*.
235 Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.

wargame – Representations of conflict or competition in a synthetic environment in which people make decisions and respond to the consequences of those decisions;[236] Future-oriented event employing innovative future forces and capabilities in representative future operating environments to gain insight about potential approaches and requirements against anticipated future challenges.[237]

wargaming – A synthesis of warfare with a defined ruleset, involving the multi-sided and adversarial engagement of human players. Wargames may or may not use an experimental approach as described in GUIDEx. The possible range of underlying computer simulation support is: i. none (i.e., seminar or tabletop wargames); ii. an Analytic Wargame (i.e., turn-based adjudication); or iii. a HITL simulation (e.g., Janus or JSAF) (i.e., continuous human interaction). iv. Human interaction with wargames is usually, but not necessarily, abstract, in that the real organizational structures and manning levels are not accurately represented. For example, two or three officers may represent an entire headquarters.[238]

Z-score – A Z-score is a statistical measurement of a data point's relationship to the mean value in a set of data points. Specifically, it is the number of standard deviations between (below or above) the population mean and the data point. A Z-score is also known as a standard score and it can be placed on a normal distribution curve.[239]

---

[236] Joint Staff Joint Force Development, Joint Planning, JP 5.0 (Washington, DC: Department of Defense, 2020).

[237] Irwin, Short Joint War Gaming & Experimentation.

[238] Labbe and Nunes Vaz, "Guidex, Ver. 1.1."

[239] Caldwell et al., *A Terms of Reference for Potential Us Marine Corps Big Data and Big Data Analytics*.

Appendix H: Abbreviations

CRC	concept required capability
DoD	Department of Defense
GRADE	Grading of Recommendations, Assessment, Development and Evaluation
IDA	Institute for Defense Analyses
JExNET	Joint Experimentation Network
JWED	Joint Wargaming and Experimentation Division
M&S	modelling and simulation
MLC	mission-level concept
NASIC	National Air and Space Intelligence Center
NDS	National Defense Strategy
NIPR	non-secure internet protocol router
PICO	population, intervention, comparison, outcome
POC	point of contact
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QARI	Qualitative Appraisal and Review Instrument
QUOROM	Quality of Reporting of Meta-Analyses
RFI	request for information
SWOT	strengths, weaknesses, opportunities, threats
TMT	task management tracker
U.S.	United States
VV&A	Verification, Validation, and Accreditation

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From – To)
22-06-2022	Study (Final)	

4. TITLE AND SUBTITLE	
Meta-Synthesis Framework to Understand Results across Defense Experiments, Studies, and Wargames	5a. CONTRACT NO. HQ0034-14-D-0001
	5b. GRANT NO.
	5c. PROGRAM ELEMENT NO(S).

6. AUTHOR(S)	
Yuna H. Wong, project lead; Darrin Denny, deputy project lead; Akar Bharadvaj, Burgess Laird, Carolyn A. Leonard, Neil Mithal, Robert Polk, Robert Stimets, Cameron Swann, Ana Venegas	5d. PROJECT NO.
	5e. TASK NO. < old and new if used>
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NO.
Joint Advanced Warfighting Division (JAWD) Institute for Defense Analyses (IDA) 4850 Mark Center Drive, Alexandria, VA 22311-1882	IDA Paper P-33123 Log no. H 22-000237

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR'S / MONITOR'S ACRONYM(S)
Executive Director Joint Warfighting Development and Design Joint Staff, Suffolk, VA 23435	JS J7
	11. SPONSOR'S / MONITOR'S REPORT NO(S).

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Unclassified

14. ABSTRACT

This report provides Joint Staff J7 with a meta-synthesis framework to aggregate and understand results across experiments, studies, and wargames. It includes an Analyst Notebook that walks an analyst through the steps in the framework and recommendations to the sponsor on its use. The report covers existing frameworks and approaches to confidence in a number of fields and adapts from these to create a framework for J7. There is also a brief review of Joint Experimentation Network (JExNET) and a few recommendations on knowledge management.

15. SUBJECT TERMS

Meta-synthesis, Joint concepts, Campaigns of learning, Experimentation, Organizational learning, Joint Experimentation, JExNET

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NO. OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	UU	115	Tom Irwin
U	U	U			19b. TELEPHONE NUMBER (Include Area Code) 757.203.7001

This page is intentionally blank.