

CORA Task 001
Phase Two: Data Analysis Plan for the Naval Exercise And
Experimentation Support Framework

PWGSC Contract Number: W7714-156105/001/SV TASK 001
ISR Report 6066-01-02
Version 2.0
16 January 2017

Presented to:
Dr. Peter Dobias
Ms. Cheryl Eisler
MARPAAC N02OR Team
Defence Research and Development Canada – Centre for Operational Research and
Analysis
PO Box 1700 STN Forces
Victoria, BC V9A 7N2

Prepared by:



International Safety Research
38 Colonnade Road North
Ottawa, Ontario
Canada K2E 7J6

Disclaimer: The scientific or technical validity of this Contract Report is entirely the responsibility of the Contractor and the contents do not necessarily have the approval or endorsement of the Department of National Defence of Canada.

© Her Majesty the Queen in Right of Canada, as represented by the Minister of National Defence, 2016
© Sa Majesté la Reine (en droit du Canada), telle que représentée par le ministre de la Défense nationale, 2016

DRDC-RDDC-2017-C016
Contract Report
December 2016

Abstract

The Maritime Forces Pacific (MARPAF) Operational Research Team (ORT) supports MARPAF/Joint Task Force Pacific (JTFP) exercises (EX), including tabletop and command post exercises (TTX/CPX) and war games; these can be at various levels up to and including the Naval Task Group (TG) level. [1] Until now, each ORT project was treated as a separate entity and each time an exercise support requirement arose, a new data collection and analysis plan was developed. The ORT commenced a multi-phased project to align MARPAF/JTFP exercises and experiments with the ORT role in supporting defence priorities, assigned missions, and tasks.

Phase One of the project [2] generated the six-component Exercise and Experimentation Support Framework (the 'Framework') to support exercises and experiments conducted at the Regional Joint Task Force, Naval TG, and individual ship or sub-unit level. The Framework is based on existing high level Canadian and United States (U.S.) military doctrine and existing bodies of knowledge within the civil emergency management domain. [3][4][5][3][6][7][8][9][10][11] It provides a method for task decomposition based on scenarios and training objectives which reflect current and future Royal Canadian Navy (RCN) joint priorities.

Phase Two-Part A proposed a data collection plan [12]. The development of this plan was based on the operational task requirements and their corresponding measurements that were associated with two complex scenarios (i.e., a multi-national exercise involving a Canadian Naval TG and an inter-agency, joint exercise conducted by coastal Joint Task Force). A set of eight data collection tools/templates were developed. These templates can be quickly and easily modified by MARPAF ORT to facilitate the data collection process. Phase Two also addressed the requirements for adhering to the Defence Research and Development Canada (DRDC) Human Research Ethics Committee (HREC) policies. [13]

Phase Two-Part B (i.e., this report) documents the data analysis component of the Framework. It was developed to guide MARPAF ORT researchers through the selection of appropriate analytical methods and the use of online statistical and open source tools that can be used to analyze the data gathered during the research activities. Due to limitations in project resources and time, the study team focused on the use of univariate inferential statistics to ensure that the level of effort needed to perform the analyses could be completed within a two to four-week timeline. This post-research activity timeline is a constraint that was identified by the Technical Authority (TA) and that must be managed to support the MARPAF/JTFP stakeholders' decision-making requirements.

The current work completes the Framework. The data collection plan and the data analysis plan are linked to major milestones in military and civilian exercise and experimentation cycles. These plans can be used to optimize the decision-making support MARPAF ORT provides to MARPAF/JTFP stakeholders.

QUALITY ASSURANCE AND VERSION TRACKING

Authorization

Title	Naval Exercise & Experimentation Support Framework Draft Report	
Report number	6066-01-02	
Version	2.0	Signature
Prepared by	K. Forbes	
Reviewed by	R Smith	
	S Lavigne	
Approved by	I Becking	
Approved for Corporate Release by	M. McCall	

Version Tracking

Ver.	Action	By	Date
1.0	Release to Client	M. McCall	20-Dec-16
2.0	Release to Client	I. Becking	16-Jan-17

TABLE OF CONTENTS

1. BACKGROUND	5
1.1 Purpose	6
1.2 Overview of the report.....	6
2. OVERVIEW OF DATA COLLECTION TOOLS	7
2.1 Levels of Measurement.....	8
2.1.1 Nominal and ordinal data sets.....	9
2.1.2 Interval and ratio data sets.....	10
3. DATA ANALYSIS PLAN	12
3.1 Procedure for implementing data analysis plan.....	12
3.1.1 Step 1: Data screening.....	12
3.1.2 Step 2: Descriptive analyses.....	13
3.1.3 Step 3: Hypothesis testing.....	13
3.1.4 Step 4: Post-Tests.....	16
3.2 Statistical analysis tools	16
3.2.1 Online statistical calculators	17
3.2.2 Downloadable statistical tools	18
3.2.3 Open source statistical tools	18
4. SUMMARY	19
REFERENCES	20
ACRONYMS	26
Annex A. PROCESS CHART DEPICTING DATA ANALYSIS PLAN	27
Annex B. DESCRIPTION OF Basic ANALYTICAL TESTS AND TOOLS	28
B.1 General description of basic analytical tests.....	28
B.2 Online statistical tools	31
B.3 Link between online statistical tools and analytical tests.....	35

LIST OF FIGURES

Figure 1: Process chart depicting data analysis plan.....	27
---	----

LIST OF TABLES

Table 1: Overview of data collection tools	7
Table 2: Linkage between templates and measurement levels.....	11
Table 3: Analyses associated with types of data	15
Table 4: Description of analytical tests	28
Table 5: Description of examples of online statistical tools	31
Table 6: Link between online statistical tools and analytical tests.....	35

1. BACKGROUND

The Maritime Forces Pacific (MARPAF) Operational Research Team (ORT) supports the Royal Canadian Navy (RCN) in the design and analysis of MARPAF/Joint Task Force Pacific (JTTF) exercises and war games. The ORT provides support at various levels up to and including the Naval Task Group (TG) level. [1]

For a live naval exercise (EX), the ORT supports the primary RCN objectives by formulating training objectives, defining Measures of Performance (MOP) and Measures of Effectiveness (MOE) as well as developing data collection (e.g., surveys and questionnaires) and analysis plans. Similarly, the ORT supports the design, conduct, and analysis of war games and experiments for hypothesis testing related to operational research objectives. [1]

Until now, each ORT project was treated as a separate entity, and each time an exercise support requirement arose, a new data collection and analysis plan was developed. In order to streamline this process and align MARPAF/JTTF exercises and experiments with the ORT role in supporting defence priorities, assigned missions, and tasks, a multi-phased project was initiated. This common approach would improve exercise and experiment planning efficiency and effectiveness. [1]

In Phase One, the study team generated the Exercise and Experimentation Support Framework (hereafter referred to as the 'Framework') to support exercises and experiments conducted at the Regional Joint Task Force, Naval TG and individual ship or sub-unit level. Seven scenarios, with three training objectives per scenario, were developed. This set of scenarios and training objectives was based on the current and future RCN joint priorities that were identified through discussions with the Technical Authority (TA) and key naval stakeholder consultations. [2] The Framework is based on existing high level Canadian and United States (U.S.) military doctrine and existing bodies of knowledge within the civil emergency management domain. [3][4][5][6][7][8][9][10][11] This approach ensured that a link was made between the high-level strategic guidance related to the naval and joint domains and the tasks and measures that were selected for inclusion in the data collection plan. [12]

Best practices related to general design and implementation of exercise cycles within the defence and emergency management domains were used to develop the Framework [5][9][10][11] along with guidance provided by military standards and handbooks pertaining to the design and conduct of experiments, demonstrations, tests and studies. [14][15][16][17] The guidance provided by the Framework can be tailored to meet the specific requirements for each program.

In Phase Two-Part A, the study team extended the Framework by generating a data collection plan. [12] To facilitate the development of the data collection plan, the study team selected two complex exercise scenarios that addressed the current and future RCN priorities. One scenario was based on a multinational exercise involving a Canadian Naval TG for the TG exercise (TGEX) and the other scenario was based on an inter-agency joint exercise related to an earthquake emergency response that was conducted by the coastal Joint Task Force for the TTX/CPX. These scenarios were generated with consideration given to existing scenarios and plans (e.g., Contingency

Plan (CONPLAN) LENTUS [18] and CONPLAN PANORAMA [19]) and provided the foundation for generating two separate task lists to depict the common operational requirements for the RCN. Based on these task lists, a set of eight data collection tools/templates were developed for use during exercises and experiments. These templates can be easily and quickly modified to meet the specific research requirements for any TGEX or TTX/CPX that is consistent with the scenarios. Phase Two also addressed the requirements for adhering to the Defence Research and Development Canada (DRDC) Human Research Ethics Committee (HREC) policies. [13]

This report, Phase Two-Part B describes the data analysis plan which provides guidance to the MARPAC ORT researchers regarding the types of analyses and tools that can be used to analyze the data.

1.1 Purpose

The purpose of this report is to complete the Framework that was created in Phase One [2] and Phase Two Part A [12] with the data analysis plan.

1.2 Overview of the report

This report is organized into the following sections:

- **Section 1: Background**
This section presents the project background and objectives. It also provides an overview (Section 1.2) of the material presented in the report;
- **Section 2: Overview of data collection tools**
This section summarizes the basic data collection tools that were identified in the data collection plan; [12]
- **Section 3: Methodology of data analysis**
The third section contains the methodology for conducting the analysis of the data/information gathered using the data collection tools identified in the data collection plan. [12] This section is supported by Annex A and Annex B.
- **Section 4: Summary**
This section provides a summary of the work performed during the project;
- **Annex A: Process chart depicting data analysis plan**
This Annex depicts the processes that were involved in generating the data analysis plan; and
- **Annex B: Description of analytical tests and tools**
This Annex presents a description of analytical tests and online statistical tools and open source software tools that can be used to analyze the data gathered from using the data collection tools.

2. OVERVIEW OF DATA COLLECTION TOOLS

This section includes an overview of the eight data collection tools/templates identified in the data collection plan. [12] The study team identified the type of data that could be gathered using each of the templates. The overall set of templates represents a range of basic data collection techniques for the data identified during the development of the data collection plan. It is understood that the ORT researchers would modify the templates to meet the specific requirements for each research activity. For example, parameters may be added or existing parameters adapted to meet specific research objectives. Similarly, the measures that are represented in these templates are only examples of how the measures identified in the data collection plan can be used with the templates; however, these measures can be used in various ways and should not be limited to the manner in which they are currently presented in the templates. A summary of these templates which depicts the type of data, the personnel roles involved in completing the tools, and the schedule used to administer the tools is in Table 1.

Table 1: Overview of data collection tools

Type of template	Intended objective	Type of data yielded				Roles involved		Administration schedule		
		Quantitative		Qualitative		Observers and Team	Participants	Before	During	After
		Categorical/Discrete	Dichotomou	Continuous	Open-Ended					
Record of observation	To guide the unobtrusive recording of observations.				x	x			x	
Checklist	To monitor activities that are relevant to the research objectives.	x	x			x	x		x	x
Dichotomous choice (yes/no)	To monitor activities that are relevant to the research objectives.		x			x	x	x	x	x
Likert rating scale	To gather feedback related to a specific aspect of a participant's knowledge, skill, ability, or experience.	x		x*			x	x	x	x

Type of template	Intended objective	Type of data yielded				Roles involved		Administration schedule		
		Quantitative		Qualitative		Observers and Team	Participants	Before	During	After
		Categorical/ Discrete	Dichotomou	Continuous	Open-Ended					
Open-ended questions	To gather qualitative feedback from the participants.						x	x	x	x
Time/data stamp	To monitor activities that are relevant to the research objectives.			x		x	x		x	
Data sample	To monitor activities that are relevant to the research objectives.			x		x	x		x	
Ranking	To order or prioritize according to assigned criteria that are relevant to the research objectives.	x		x		x	x	x	x	x

*Under specific conditions as described in Section 2.1.2.

2.1 Levels of Measurement

Four levels of measurement are used to categorize the type of data yielded. These levels include nominal, ordinal, interval and/or ratio scales. [20][21]21[20][22][23][24] The levels represent a hierarchy of measurement whereby each successive level possesses the characteristics of the preceding level(s) and has additional capabilities. The lowest level of measurement (i.e., nominal) allows simply for the classification of data and the highest level of measurement (i.e., ratio) allows for exact analyses of data. These levels are described below: [20]

- **Nominal:** Nominal data are assigned to independent categories. These categories are associated with general characteristics. There is no order imposed by the categorization; however, nominal scales are used to organize the data into categories that can be used to determine the frequency of occurrence (i.e., how many are observed within each category);
- **Ordinal:** Ordinal data are also assigned to a category; however, the data are ranked to reflect the amount associated with a given characteristic. The data are

counted within each category to determine the frequency of occurrence rather than measured according to a unit. Therefore, the differences between the ranks cannot be determined when using an ordinal scale;

- **Interval:** Interval data are measured and assigned to a scale. Since the assignment to the scale is based on actual measurements, the intervals between the points on the scale are consistent. The interval can then be used to measure meaningful differences between points on the scale. For this reason, interval data is referred to as continuous data. An interval scale has an arbitrary zero. Most behavioral science research projects are concerned with interval data; and
- **Ratio:** Ratio data are assigned to a scale which may have an absolute zero (i.e., there is a point on the scale that can indicate the absence of the parameter that is being measured). Ratios are “normalized measurements” that express relationship between two quantities. This means that valid ratio comparisons can be made based on the data.

In keeping with this standard approach to organizing data, the study team categorized the types of data that would be gathered from the templates by using these levels of measurement. This provided the foundation for developing the data analysis plan since the selection of the appropriate analyses is based on the respective level of measurement associated with the data set. The linkage between the data collection templates and the four data measurement levels is shown in Table 2. The following sub-sections describe the relationship between these four levels of measurement and the data sets gathered from using the templates.

2.1.1 Nominal and ordinal data sets

Categorical, discrete measures and open-ended questions that yield categorical and/or discrete data were identified in the data collection plan. [20][21][22][23][24][25] These types of data are consistent with nominal and ordinal data sets which are used to describe the general characteristics of the data set. The rationale for associating these measures with nominal and ordinal data sets is described below:

- **Categorical:** Data points can be assigned to one of multiple categories. Data are referred to as categorical or ‘nominal’ when there is no basis for ordering the data in a particular way. Categorical data can be assigned to groups based on a characteristic (e.g., gender and hair colour) and does not have to be numeric;
- **Dichotomous:** Data points that are assigned to one of two unique categories (e.g., yes/no) are a special type of categorical data and are commonly referred to as dichotomous. Dichotomous data do not have to be numeric. These data points are treated as nominal data;
- **Open-ended questions:** Qualitative data are frequently gathered from research participants and/or observers using open-ended questions. Although not specifically addressed in the data collection plan, the recording and/or translation of human behaviours for subsequent observational purposes, or of participant comments, would also be considered qualitative in nature. [25] Qualitative data sets are analyzed by assigning codes (based on categorical and/or discrete

- variables) to the participant feedback that reflect the research objectives. These data sets are treated as nominal data; and
- **Discrete:** Data that can be categorized using a finite set of numbers which reflect a sequence or an order that reflects a pre-defined parameter or characteristic. Each data point in the data set is assigned to a numeric value on the scale. There is meaning associated with an increasing and/or decreasing numerical value on the scale. However, the interpretation of the scale is typically limited to describing the magnitude of the parameter (i.e., more or less of a parameter). Data associated with these finite scales can be ranked to evaluate the overall order of the data. Differences between the observed rank orders can be subjected to analyses using non-parametric statistics (e.g., Friedman Test). [27]

Schools of Thought Analysis (SOTA) have been used to support decision making in the Canadian Department of National Defence. [26][27] These analyses cluster participant responses gathered using ordinal rating scales based on certain values. With data sets that are limited in size, this approach allows the researcher to examine the variations in the data and to identify similarities and differences within a data set. The hierarchy associated with the clustering is represented numerically using tree diagrams (also referred to as dendrograms), distributions and tables. Similarly, multi-criteria decision analyses (MCDA) are used to support decision-making within complex problem spaces involving several factors. The MCDA provides a process which can be used to structure and organize information. The output of MCDA is intended to rank and classify options based on assessments of criteria that are made by knowledgeable participants and decision-makers. [27]

2.1.2 Interval and ratio data sets

Continuous measures and some discrete measures, which can be treated as continuous measures under special conditions, were identified in the data collection plan. [20][22][23][24][25] These types of data are consistent with interval and ratio data sets. Interval and ratio data sets are summarized to describe the general characteristics of the data set (i.e., what the data looks like) prior to performing hypothesis testing. In addition, these data sets can be subjected to parametric and/or non-parametric statistical analyses to perform hypothesis testing and predictive analyses. Parametric tests can be selected when assumptions related to normality are met; otherwise, non-parametric tests are required. [28][29] There are several types of statistical tests which can be selected based on the specific research questions that are being addressed and the circumstances associated with the data collection. The rationale for associating these measures with interval and ratio data sets is described below:

- **Discrete (under special conditions):** When appropriate design considerations are made (e.g., labelling and anchors assigned to scales imply equal spacing between discrete values) discrete data can be treated as interval data. When discrete data points are interpreted as interval data, it can be used to assess the magnitude of participants' responses to criteria such as satisfaction, knowledge, and agreement. It is noted that there is an on-going debate with respect to whether it is appropriate under any circumstance to associate discrete data with the interval level of measurement; [30][31][32][33] and

- Continuous: Data that are measured and that have a meaningful scale with quantifiable and equal differences between the units are identified as interval data. However, interval data does not accommodate a value of zero. In contrast, the highest level of measurement, ratio data, does accommodate measurements with a meaningful value of zero and the rankings that are assigned to data based on their actual value. [20]

Table 2: Linkage between templates and measurement levels

	Types of data scales					
	Nominal			Ordinal	Interval	Ratio
Name of Template	Categorical	Discrete	Dichotomous	Rank order/ scale	Continuous	Continuous
Record of observation	x					
Checklist		x	x*			
Dichotomous choice (yes/no)			x			
Likert rating scale				x	x	
Open-ended questions	x					
Time/data stamp					x	x
Data sample						x
Ranking		x		x		

* Some exceptions may apply

3. DATA ANALYSIS PLAN

This section presents the data analysis plan for the Framework. This plan guides the analysis of the data that will be gathered by using the templates to evaluate exercises and test experimental hypotheses. [12]

Two assumptions were made when developing this data analysis plan. First, it was assumed that the research questions were identified during the Conceive phase of the Framework. Second, it was assumed that the data and data collection tools that will be needed to address the research questions will be identified during the Design and Planning phases of the Framework. [2] With this knowledge, the ORT researchers should be in a position to use the data analysis plan to identify the measurement level and select the appropriate analytical test once the complete data set is available.

The data analysis plan addressed a set of criteria that were provided by the TA at the kick-off meeting. These criteria are listed below:

- The analyses can be completed within a timeline, typically ranging from two to four weeks post-exercise/experiment, that will support the MARPAC/JTFP stakeholders' decision making requirements; and
- The analyses will be simple but sufficiently comprehensive to address the objectives of the research activity in a meaningful manner.

The study team selected a univariate analyses as the foundation for the data analysis plan. While the study team acknowledges that ORT researchers frequently use multivariate analyses, the study team limited the focus of the data analysis plan to descriptive and univariate inferential statistics.¹

3.1 Procedure for implementing data analysis plan

The data analysis plan adheres to the general approach that is taken to perform analyses within the domain of behavioral sciences. [34] The data analysis plan is comprised of four steps. These steps are performed sequentially starting with preparing (screening), characterizing (descriptive analyses) and evaluating the data for meaningful differences (inferential analyses). [28][35] These steps are graphically depicted in a process chart presented in Annex A and are described in the sub-sections below.

3.1.1 Step 1: Data screening

Prior to commencing a data analysis, the entire data set should be screened to ensure that the data are properly prepared for analysis and any shortcomings can be adequately addressed. [35] The screening methods are determined by the type of data and the intent of the analyses that will be performed. In general, data screening is

¹ The SOW referenced a requirement to address multi-criteria decision analysis approaches however, this was not directly addressed because it is the study team's opinion that it is a multivariate technique. Also, the data collection tools [12] do not directly address this type of analyses given the volume of data that would be expected in order for this analysis to be appropriate.

intended to ensure the reliability and accuracy of the data sets. Data screening practices should include the following components:

- Inspect data set for accuracy of input (error identification);
- Evaluate amount and distribution of missing data; and
- Identify and address outliers.

3.1.2 Step 2: Descriptive analyses

The study team identified a set of descriptive analyses that are appropriate for each type of data (Table 3). Two categories of descriptive tests are used to characterize data sets. Researchers typically describe distributions using measures from each of these categories. The appropriate usage of these tests is determined by the type of data that is being analyzed. The researcher should complete the following steps to provide a description of the data set(s). These categories are described below:

- Measures related to central tendency: These tests are used to describe the similarities that are observed in the data. [36][37] In a normally distributed population all of these measures are equal to each other. The most commonly used measures of central tendency are identified and described in Table 4 which is described in Annex B.1 [38]. In addition, Confidence Intervals (CI) and skewness were included based on guidance from the TA; and
- Measures of dispersion: These tests are used to describe the variations that are observed within the data sets. The most commonly used measures of dispersion are identified and described in Table 4, which is described in Annex B.1. [37][39][40]

3.1.3 Step 3: Hypothesis testing

The study team identified a set of inferential (univariate) analyses that are appropriate for analyzing each type of data (Table 3). [20][34] Data sets can be appropriately analyzed using various analytical approaches when the assumptions underlying the selected tests are satisfied. The ORT researchers intending to use parametric tests should determine whether the data sets are normally distributed. Some parametric tests can tolerate violations in the normality and there are instances in which parametric tests can be used when the distribution is not normally allocated (e.g., appropriate sample sizes are available, groups have unequal variances, greater statistical power is needed). [28] The likelihood that a data set will be normally distributed can be improved through the selection of dependent variables that are typically associated with a normal distribution. [22] If parametric tests are not appropriate then the ORT researcher can select the equivalent non-parametric version of the test. In addition to determining which test to use, there may be circumstances related to the data collection and/or data set (e.g., small sample sizes) which require the ORT research to select specialized versions of these tests to optimize the outputs of the tests.

Researchers may be required to work with small data sets for various reasons including limited access to relevant participants or a loss of data. In some cases, researchers may be limited to the use of descriptive analyses which characterize the overall response. However, small data sets that are intended to address experimental research

questions may be treated with more advanced techniques. First, non-parametric analyses can be selected, providing the underlying assumptions are met, to analyze small data sets because these analyses do not assume a normal distribution of data. Second, researchers can model the behaviour of the parameter that is being studied to conduct an analysis. [41] For example, a data set involving one parameter that is known to have a constant failure rate or does not have any failures can be modelled using an exponential distribution. In another example, the behaviour of a single parameter that is known to have an increasing, constant or decreasing failure rate can be modelled using the One-Parameter Weibull Distribution. Further still, small data sets can be treated with a Bayesian analysis, which models behaviour by incorporating both prior knowledge and observed data into the analysis. [41]

Due to the constraints that are imposed by the statistical analyses, the ORT researchers should ensure that there is a clear link between the research objectives and the data collection tools that are used to gather the necessary data prior to the data collection phase.

The researcher should follow five steps to perform hypothesis testing: [35]

- State the null hypothesis for the selected test;
- Identify the level of statistical significance. Typically, a probability level that is less than or equal to .05 or .01 is used;
- Calculate the degrees of freedom. This calculation will be necessary if the researcher is performing the analyses by hand. When using software tools to perform the analysis the degrees of freedom will be included in the output;
- Perform the selected statistical test. If the researcher performs the test by hand then it will be necessary to compare the calculated value to the table value. When using software tools to perform the analysis the output will compare the calculated value to the table value and provide an indication of the statistical significance; and
- State whether the null hypothesis is accepted or rejected.

The study team identified a set of parametric and non-parametric analyses that are appropriate for conducting hypothesis testing. [41] These tests are identified and linked to the data measurement levels in Table 3. Each analytical test is described and linked to examples of software tools that can be used to perform the analysis in Annex B.

Table 3: Analyses associated with types of data

Type of scale	Descriptive Analyses					Hypothesis Testing Univariate Inferential Statistics																	
	Related to Central Tendency			Dispersion		Parametric										Non-Parametric							
	Mode	Median	Mean, CI, Skewness	Range (minimum and maximum scores)	SD	Frequency Counts	Correlation	Contingency Tables	One sample t-test	Dependent t-test	Independent t-test	ANOVA	ANOVA with repeated measures	Linear regression	Post-Test	Chi Square	Spearman Rank-Order Correlation	Wilcoxon Test	Mann-Whitney Test	Kruskal-Wallis Test	Friedman Test	Post-Test	
Nominal	X			X		X										X							
Ordinal	X	X		X		X	X	X									X						
Interval	X	X	X	X	X	X			X	X	X	X	X	X	X			X	X	X	X	X	X
Ratio	X	X	X	X	X	X			X	X	X	X	X	X	X			X	X	X	X	X	X

3.1.4 Step 4: Post-Tests

Post-tests must be performed to identify the source of the statistically significant differences when there are three or more groups involved in the analysis. [43] Two broad categories of tests (planned and post hoc comparisons) are used to make comparisons between the group means, which represent the levels of a variable. Examples of applicable tests are provided below. Researchers are cautioned that the determination of the planned and post hoc comparisons is heavily dependent upon the research objectives and the predictions that are made at the outset of the research.

- Planned comparisons: Independent comparisons between levels of a variable that are defined as a priori. Planned comparisons can be performed using weighted means to compare the group means that are predicted to be significantly different. A modified Bonferroni test can also be used; and
- Post hoc comparisons: Comparison that are made between levels of a variable that are performed after a statistically significant effect is observed. Examples include the following tests: [43]
 - Scheffe's Test – Controls for violations of assumptions underlying the test;
 - Dunn's Test or modified Bonferroni Test: Not concerned with violations of assumptions and performing a comparison of less than half of the compound and pairwise tests;
 - Dunnett's Test: Not concerned with violations of assumptions and comparing the treatment to the control group; and
 - Tukey's Test: Not concerned with violations of assumptions but are testing more than half of the possible pairwise comparisons.

There are numerous planned and post hoc comparisons that can be used to identify the source of an observed significant difference. Where possible, planned comparisons are preferred because these tests are more powerful than post hoc comparisons. They may be used if predictions were made at the outset of the research regarding expected differences between the levels of a variable. If the researchers determine that planned comparisons are inappropriate then post hoc comparisons should be used. In this case, the researchers should consider the impact of the post hoc comparisons on Type I (i.e., false positive because Type I error rejects the null hypothesis when it is in fact true) and Type II (false negative because Type II error fails to reject the null hypothesis when an alternative hypothesis is in fact true). [44]

3.2 Statistical analysis tools

The study team performed an internet-based search to identify basic open source analytical tools that are applicable to the analyses identified in Table 3. The study team used criteria, provided by the TA, as guidance for performing a review of the tools that were identified from the search. This search identified basic web-based statistical calculators as well as statistical tools that can be freely downloaded for use. These categories are described in the sub-sections below.

3.2.1 Online statistical calculators

The study team identified a category of basic analytical tools, referred to as online statistical calculators that can be freely accessed by the public. [45][46][47][48][49][50][51] A review of these calculators indicated that there is a wide range of capabilities available online. Some online calculators can address several types of descriptive and inferential (parametric and non-parametric) analyses and post-tests whereas other online calculators are limited to addressing just a few analytical tests. The study team compiled a set of examples that are briefly described in Table 5 which is presented in Annex B.2. These online calculators are linked to the analyses that were identified in the data analysis plan in Table 6 which is presented in Annex B.3. These examples are not intended to represent an exhaustive list of free online statistical calculators but rather to identify tools that can be easily accessed (providing internet access is available) and considered by the ORT researchers.

It is important to note that the reliability of these tool sets has not been confirmed by the study team and should be verified by the ORT researchers prior to documenting the analytical output in a report.

The characteristics associated with the free online statistical calculators are summarized below:

- Online statistical calculators are compatible with both Windows and Macintosh operating systems;
- They do not require the users to obtain a software licence or setup a user account to access the tools. They are not add-ons to existing software packages;
- Online statistical calculators are browser-based and can be accessed directly from the websites. They do not have to be (and cannot be) downloaded or installed on the user's computer. One exception to this was noted which provided access to online calculators that could be downloaded for a fee to obtain a desktop version or the script;
- Online statistical calculators provide access to a wide range of statistical tests; however, none provide access to the full range of analyses that were identified in the data analysis plan;
- Users can choose to enter data manually or by copying and pasting data from other software applications (e.g., MS Excel);
- The output of information can be retained by using capture techniques that are commonly used in a web environment (e.g., print screen, save or print to pdf). If there are graphics that are created as part of the output then they will be captured using these standard methods of data capture;
- These tool sets provide instructions and/or tutorials to guide the users' selection for the data sets. It is noted that some online calculators provide extensive guidance whereas, others provide very little guidance. The instructions provided by the online calculators are useful however, ORT researchers will require a general, and in some cases a detailed, understanding of the analytical tests to ensure that the appropriate test(s) are selected for the data sets; and
- An assessment of user-friendliness was beyond the scope of the current project; however, the methods of interacting with these tools will be readily understood by users who are already familiar with using web-based applications.

3.2.2 Downloadable statistical tools

The study team also identified one software tool that can be downloaded freely to obtain access to statistical tools. [52] This example is briefly described (Table 5 presented in Annex B.2) and linked to the analyses that were identified in the data analysis plan (Table 6 presented in Annex B.3). These examples are not intended to represent an exhaustive list of free downloadable tools; however, the search conducted by the study team only yielded these two examples. One of these examples was reported in the preceding section as the website also provided access to online statistical calculators.

The study team was unable to download these software tools due to the team's corporate policy restrictions, which limit downloading non-proprietary software.

It is important to note that the reliability of these tool sets has not been confirmed by the study team and should be verified by the ORT researchers prior to documenting the analytical output in a report.

3.2.3 Open source statistical tools

In accordance with the SOW, this review required the study team to review the open source software packages identified on gnu.org. [53] This organization provides access to free software packages and manuals that are licensed under the GNU General Public License (GNU GPL). Data source code can be downloaded for GNU software packages, referred to as GNU packages. The programs can be redistributed and/or modified under the terms indicated in the GNU GPL. The programs can build on GNU/Linux and UNIX platforms. The website provides instructions for downloading and building the programs.

Three software packages were identified as potential options for downloading open source tools for performing statistical analyses. [54][55][56] These software packages are briefly described (Table 5 presented in Annex B.2) and linked to the analyses that were identified in the data analysis plan (Table 6 presented in Annex B.3).

The study team did not download these software packages; this was not possible due to the team's corporate policy restrictions which limit the use of open source software. The reliability of these open source software packages (GNU packages) has not been confirmed by the study team.

4. SUMMARY

The Framework is designed to standardize the processes by which the MARPAC ORT provides decision-making support to MARPAC/JTFP exercises and experiments. It provides guidance related to the types of data that can be gathered during these research activities and how the data can be gathered using a wide range of data collection tools. The Framework and the data collection plan are described separately. [2][12] This report defines the data analysis plan which supports the data collection plan. Its purpose is to guide MARPAC ORT researchers in selecting appropriate analytical tests and tools to analyze the findings.

The data analysis plan focuses on the use of descriptive and univariate inferential statistics to ensure that the level of effort needed to perform the analyses can be completed within a two to four-week time frame. This period is a constraint that was identified by the TA and ensures timely support to the MARPAC/JTFP stakeholders' decision-making processes.

The data analysis plan is linked to the phases of the Framework. The research objectives for an exercise or experiment should be identified during the Conceive phase. These objectives are used in determining the type of measures that will be implemented to perform an evaluation. The data collection tools should be identified during the Design and Planning phases. With this knowledge, the ORT researchers can use the data analysis plan to identify the measurement level and select the appropriate test for analyzing the data. It is noted that the Analysis phase will be facilitated by the ORT researchers when the link between the research objectives and the type of data (and templates) that will be gathered to conduct the evaluation is clearly articulated. This link should be established as early as possible during the exercise and experimentation cycle.

The data analysis plan defined herein is consistent with the standard approach used within the behavioural sciences that associates the type of data with one of the measurement levels (nominal, ordinal, interval, and ratio). [20][21][34][35] This categorization is used as a guide for selecting an appropriate analytical approach. The plan provides guidance with respect to the use of descriptive and inferential (univariate) tests for both parametric distributions and non-parametric distributions as well as for the use of post-tests (planned and post hoc comparisons). [28][43] The unique circumstances associated with the data collection phase (e.g., sample size) and the characteristics of the resulting data set(s) should be considered when the analytical tests are selected.

In conclusion, these findings represent the final step in the completion of the Framework. Overall, the Framework provides the ORT with two task lists that reflect common operational requirements for expected current and future RCN priorities. [3] Performance can now be evaluated using a set of task-relevant measures. Finally, guidance has been developed to support the ORT researchers in their determination of the appropriate set of analyses so that the analyses can be completed with the two to four-week time constraints.

REFERENCES

- [1] Defense Research and Development, Centre for Operational Research and Analysis, "Naval exercise and experimentation support framework development: Phase 2 data collection and analysis" 2016. Task Authorization W7714-15-6105, Task-001, Annex A: Statement of Work.
- [2] K. Salchert, "Naval exercise and experimentation support framework," 2016. [Online]. Available: http://cradpdf.drdc-rddc.gc.ca/PDFS/unc223/p803538_A1b.pdf. Accessed: Jul. 12, 2016.
- [3] Department of National Defence, Royal Canadian Navy, "LEADMARK 2050 Canada in a new maritime world," 2016. Available: <http://navalassoc.ca/wp-content/uploads/2016/05/Leadmark-2050-13-May-2016.pdf>. Accessed: September 30, 2016.
- [4] Department of National Defence, "Commander Royal Canadian Navy Standing Direction and Guidance," (CRCN SD&G) 3371-3250-1(DGNSR/RDIMS 366208), 2015.
- [5] Department of National Defence, Royal Canadian Navy, "CAF JTL 2016 Overview by operational function," 2016.
- [6] C. Eisler, A. Bourque, and W. Reive, "Fleet mix study: Capability supply and demand requirements for Iteration II," 2009. DRDC CORA TM 2009-040.
- [7] Department of National Defence, "CFCD 129 Royal Canadian Navy Readiness and Sustainment Direction," C-03-005-001/AM-01, 2009.
- [8] Department of National Defence, "CFCD 102(L) Royal Canadian Navy Combat Readiness/Training Requirements," B-GN-002-000/RQ-001.
- [9] Chief of Naval Operations & Commandant of the United States Marine Corps & Commandant Headquarters United States Coast Guard, "Universal Naval Task List (UNTL)," 2007.
- [10] U.S. Department of Homeland Security, "Target Capabilities List: A companion to the National Preparedness Guidelines," 2007.
- [11] U.S. Department of Homeland Security, "National Planning Scenarios," 2006.
- [12] K. Forbes and K. Salchert, "Phase Two: Data Collection Plan for the Naval Exercise and Experimentation Support Framework," 2016. DRDC-RDDC-2016-C326.
- [13] "TCPS Chapter 2 –Scope and approach: The interagency advisory panel on research ethics (PRE)," 2014. Article 2.1 [Online]. Available: <http://www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/chapter2-chapitre2/#toc02-1a>. Accessed: Aug 15, 2016.

- [14] MIL-STD-46855A, "Department of Defense Standard Practice: Human engineering requirements for military systems, equipment, and facilities," 24 May 2011.
- [15] Department of National Defence, "CFCD 124 Maritime Operational Test and Evaluation Guide," Revision 2, 2013.
- [16] Department of National Defence, "CFCD 101 Maritime Operational Analysis Guide," CFMWC 3350-1, 2009.
- [17] J.R. Wilson and N. Corlett, "Evaluation of human work," 3rd Edition, CRC Press, 2005.
- [18] Canadian Joint Operations Command, "CONPLAN LENTUS – Canadian Armed Forces Assistance to Federal/ Provincial/Territorial Disaster Relief Operations," 17 July 2014.
- [19] Joint Task Force Pacific, "CONPLAN PANORAMA – Response to a Catastrophic Earthquake in British Columbia," 12 February 2014.
- [20] R.C. Sprinthal, "Basic statistical analysis," 4th Ed., Allyn and Bacon, 1994.
- [21] Statistics Solutions, "Data levels of measurement." [Online]. Available: <http://www.statisticssolutions.com/data-levels-of-measurement/>. Accessed: Dec. 4, 2016.
- [22] "What is the difference between categorical, ordinal and interval variables?" [Online]. Available: http://www.ats.ucla.edu/stat/mult_pkg/whatstat/nominal_ordinal_interval.htm. Accessed: Dec. 2, 2016.
- [23] Lund Research Ltd, "Types of variable," 2013. [Online]. Available: <https://statistics.laerd.com/statistical-guides/types-of-variable.php>. Accessed: Dec. 2, 2016.
- [24] Measuring Usability LLC, "UsableStats: Nominal, Ordinal, interval and ratio," 2004. [Online]. Available: <http://www.usablestats.com/lessons/noir>. Accessed: Dec. 2, 2016.
- [25] "Chapter 37. Operations in evaluating community interventions," 2016. [Online]. Available: <http://ctb.ku.edu/en/table-of-contents/evaluate/evaluate-community-interventions/collect-analyze-data/main>. Accessed: Dec. 8, 2016.
- [26] P. Dobias et al., "STO-TM-SAS-094 Employment of non-lethal capabilities for visit, board, search, and seizure operations: Naval postgraduate school wargame," 2015. [Online] Available: http://cradpdf.drdc-rddc.gc.ca/PDFS/unc195/p801883_A1b.pdf. Accessed: Jul. 27, 2016.
- [27] F. Cameron and G. Pond, "Military decision making using schools of thought analysis - A soft operational research technique with numbers," Proceedings from the 27th International Symposium on Military Operational Research, 2010. [Online]. Available: <http://ismor.cds.cranfield.ac.uk/27th-symposium-2010>. Accessed: Jan.

11, 2017.

- [28] J. Frost, "Choosing between a Nonparametric test and a parametric test," 2015. [Online]. Available: <http://blog.minitab.com/blog/adventures-in-statistics/choosing-between-a-nonparametric-test-and-a-parametric-test>. Accessed: Dec. 9, 2016.
- [29] "Chi Square." [Online]. Available: <http://www.okstate.edu/ag/agedcm4h/academic/aged5980a/5980/newpage28.htm>. Accessed: Dec. 7, 2016.
- [30] J.D. Brown, "Likert items and scales of measurement?" [Online]. Available: <https://jalt.org/test/PDF/Brown34.pdf>. Accessed: Dec. 2, 2016.
- [31] G.M. Sullivan and A.R. Artino, "Analyzing and interpreting data from Likert-Type scales," vol. 5, no. 4, Dec. 2013. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3886444/>. Accessed: Aug. 5, 2016.
- [32] CDC, "CDC coffee break: Using Likert scales in evaluation survey work," 2012. [Online]. Available: http://www.cdc.gov/dhdsdp/pubs/docs/CB_February_14_2012.pdf. Accessed: Aug. 7, 2016.
- [33] K. Grace-Martin, "Can Likert scale data ever be continuous?," 2008. [Online]. Available: <http://www.theanalysisfactor.com/can-likert-scale-data-ever-be-continuous/>. Accessed: Jan. 11, 2017.
- [34] G. Keppel and J. Zedeck, "Data analysis for research designs," Worth Publishers, 1989.
- [35] B.G. Tabachnick and L.S. Fidell, "Using multivariate statistics," 2nd edition, HarperCollins Publishers, Inc, 1989.
- [36] M. Rouse, "What is statistical mean, median, mode and range? - definition from WhatIs.com," SearchDataCenter, 2014. [Online]. Available: <http://searchdatacenter.techtarget.com/definition/statistical-mean-median-mode-and-range>. Accessed: Dec. 4, 2016.
- [37] "Descriptive vs. Inferential statistics difference," in Data Analysis, My Market Research Methods, 2011. [Online]. Available: <http://www.mymarketresearchmethods.com/descriptive-inferential-statistics-difference/>. Accessed: Dec. 5, 2016.
- [38] D.M. Lane and H. Ziemer, "What is central tendency." [Online]. Available: http://onlinestatbook.com/2/summarizing_distributions/what_is_ct.html. Accessed: Dec. 4, 2016.
- [39] "Measures of dispersion," 2016. [Online]. Available: <http://learntech.uwe.ac.uk/da/Default.aspx?pageid=1432>. Accessed: Dec. 8, 2016.

- [40] StatTrek.com, "Standard deviation: Definition," 2016. [Online]. Available: http://stattrek.com/statistics/dictionary.aspx?definition=Standard_deviation. Accessed: Dec. 5, 2016.
- [41] Weibull.com, "You have a small data set: What do you do?" [Online]. Available: <http://www.weibull.com/hotwire/issue72/hottopics72.htm>. Accessed: Jan 11, 2017.
- [42] V. Mabin and M. Beattie, "A practical guide to multi-criteria decision analysis." [Online] Available: http://www.victoria.ac.nz/som/researchprojects/publications/Multi-Criteria_Decision_Analysis.pdf. Accessed Jan 11, 2017.
- [43] "Planned comparisons and post hoc tests." [Online]. Available: http://homepages.inf.ed.ac.uk/bwebb/statistics/Planned-comparison_post-hoc.pdf. Accessed: Dec. 8, 2016.
- [44] "Type I and type II errors understanding type I and type II errors." [Online]. Available: <http://www.stat.berkeley.edu/~hhuang/STAT141/Lecture-FDR.pdf>. Accessed: Dec. 8, 2016.
- [45] R. Lowry, "VassarStats: Website for statistical computation," 2016. [Online]. Available: <http://vassarstats.net/index.html>. Accessed: Dec. 12, 2016.
- [46] GraphPad Software, Inc., "GraphPad Software," 2016. [Online]. Available: <http://graphpad.com/quickcalcs/>. Accessed: Dec. 12, 2016.
- [47] "Online statistics calculator, basic math statistics, Tutorial." [Online]. Available: <https://www.easycalculation.com/statistics/statistics.php>. Accessed: Dec. 12, 2016.
- [48] "Free online statistics calculators." [Online]. Available: <http://www.mathportal.org/calculators/statistics-calculator/>. Accessed: Dec. 13, 2016.
- [49] "Statistical test calculators," 2016. [Online]. Available: <http://www.socscistatistics.com/tests/Default.aspx>. Accessed: Dec. 13, 2016.
- [50] 2016 calculator, "Math calculators," 2008. [Online]. Available: <http://www.calculator.net/math-calculator.html>. Accessed: Dec. 14, 2016.
- [51] CalculatorSoup, "Statistics calculators," 2006. [Online]. Available: <http://www.calculatorsoup.com/calculators/statistics/>. Accessed: Dec. 13, 2016.
- [52] StatPac Inc, "Free statistics calculator for statistical analysis significance tests," 2014. [Online]. Available: <https://statpac.com/statistics-calculator/index.htm>. Accessed: Dec. 13, 2016.
- [53] Free Software Foundation, "Various Licenses and Comments about Them," 2014. [Online]. Available: <https://www.gnu.org/licenses/license-list.html#Introduction>. Accessed: Aug. 15, 2016.

- [54] Free Software Foundation, "Dap - GNU project - free software foundation (FSF)," 2001. [Online]. Available: <https://www.gnu.org/software/dap/>. Accessed: Dec. 13, 2016.
- [55] Free Software Foundation, "GNU Datamash," 2014. [Online]. Available: <https://www.gnu.org/software/datamash/>. Accessed: Dec. 13, 2016.
- [56] Free Software Foundation, "GNU PSPP," 2013. [Online]. Available: <https://www.gnu.org/software/pspp/>. Accessed: Dec. 13, 2016.
- [57] "7.1.4. What are confidence intervals?" [Online]. Available: <http://www.itl.nist.gov/div898/handbook/prc/section1/prc14.htm>. Accessed: Jan 11, 2017.
- [58] StatTrek.com, "Relative frequency: Definition," 2016. [Online]. Available: <http://stattrek.com/statistics/dictionary.aspx?definition=Relative%20frequency>. Accessed: Dec. 8, 2016.
- [59] P.F. Smith, "Types of data, descriptive statistics, and statistical tests for nominal data." [Online]. Available: <https://www.accp.com/docs/bookstore/biosampl.pdf>. Accessed: Dec. 13, 2016.
- [60] Lund Research Ltd, "Pearson product-moment correlation," 2013. [Online]. Available: <https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php>. Accessed: Dec. 12, 2016.
- [61] Lund Research Ltd, "One-Sample t-test using SPSS statistics," 2013. [Online]. Available: <https://statistics.laerd.com/spss-tutorials/one-sample-t-test-using-spss-statistics.php>. Accessed: Dec. 8, 2016.
- [62] Lund Research Ltd, "Dependent t-test using SPSS statistics," 2013. [Online]. Available: <https://statistics.laerd.com/spss-tutorials/dependent-t-test-using-spss-statistics.php>. Accessed: Dec. 8, 2016.
- [63] Lund Research Ltd, "Independent t-test for two samples," 2013. [Online]. Available: <https://statistics.laerd.com/statistical-guides/independent-t-test-statistical-guide.php>. Accessed: Dec. 8, 2016.
- [64] Statistics Solutions, "ANOVA." [Online]. Available: <http://www.statisticssolutions.com/manova-analysis-anova/>. Accessed: Dec. 8, 2016.
- [65] Lund Research Ltd, "Repeated measures ANOVA," 2013. [Online]. Available: <https://statistics.laerd.com/statistical-guides/repeated-measures-anova-statistical-guide.php>. Accessed: Dec. 12, 2016
- [66] Lund Research Ltd, "Linear regression analysis using SPSS statistics," 2013. [Online]. Available: <https://statistics.laerd.com/spss-tutorials/linear-regression-using-spss-statistics.php>. Accessed: Dec. 12, 2016.

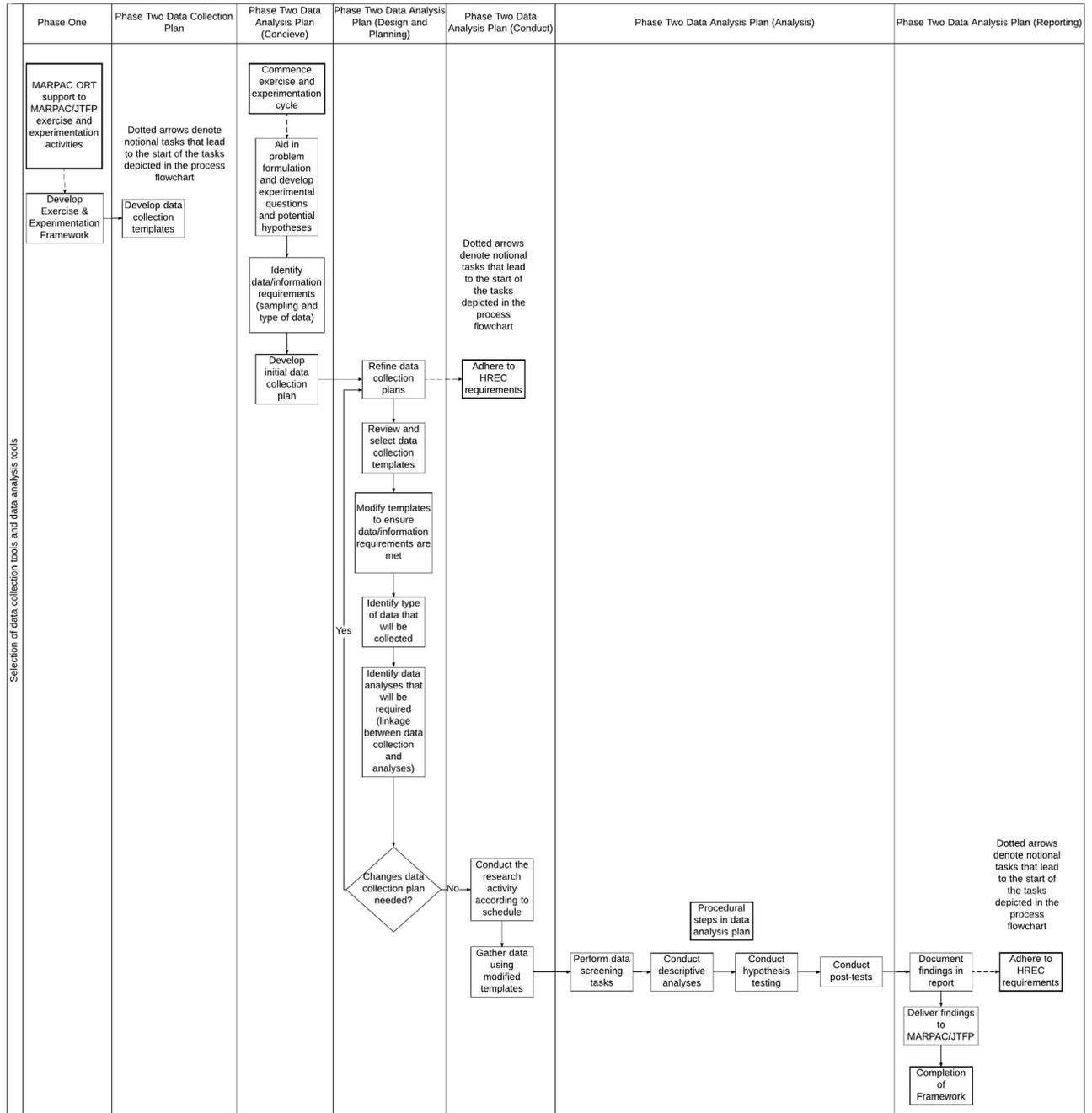
- [67] Lund Research Ltd, "Spearman's rank-order correlation using SPSS statistics," 2013. [Online]. Available: <https://statistics.laerd.com/spss-tutorials/spearman-rank-order-correlation-using-spss-statistics.php>. Accessed: Dec. 12, 2016.
- [68] Lund Research Ltd, "Wilcoxon signed-rank test using SPSS statistics," 2013. [Online]. Available: <https://statistics.laerd.com/spss-tutorials/wilcoxon-signed-rank-test-using-spss-statistics.php>. Accessed: Dec. 8, 2016.
- [69] Lund Research Ltd, "Mann-Whitney U test using SPSS statistics," 2013. [Online]. Available: <https://statistics.laerd.com/spss-tutorials/mann-whitney-u-test-using-spss-statistics.php>. Accessed: Dec. 8, 2016.
- [70] Lund Research Ltd, "Kruskal-Wallis H test using SPSS statistics," 2013. [Online]. Available: <https://statistics.laerd.com/spss-tutorials/kruskal-wallis-h-test-using-spss-statistics.php>. Accessed: Dec. 8, 2016.
- [71] The R Foundation, "What is R?" [Online]. Available: <https://www.r-project.org/about.html>. Accessed: Jan. 11, 2017.
- [72] Python.org, "Welcome to python.org." [Online]. Available: <https://www.python.org/>. Accessed: Jan. 11, 2017.
- [73] Python.org, "Python tutorial." [Online]. Available: <http://www.tutorialspoint.com/python/>. Accessed: Jan. 11, 2017.
- [74] Python.org, "FrontPage - python Wiki". [Online]. Available: <http://wiki.python.org/moin/FrontPage>. Accessed: Jan. 11, 2017.
- [75] Palisade Corporation, "StatTools: Forecasting and statistical analysis software for excel." [Online]. Available: <http://www.palisade.com/stattools/>. Accessed: Jan. 11, 2017.
- [76] Addinsoft, "XLSTAT-Base." [Online]. Available: <https://www.xlstat.com/en/solutions/base>. Accessed: Jan. 11, 2017.
- [77] Addinsoft, "XLSTAT-Premium." [Online]. Available: [Online]. Available: <https://www.xlstat.com/en/solutions/premium>. Accessed: Jan. 11, 2017.
- [78] Lund Research Ltd, "Friedman test in SPSS statistics," 2013. [Online]. Available: <https://statistics.laerd.com/spss-tutorials/friedman-test-using-spss-statistics.php>. Accessed: Dec. 12, 2016.

ACRONYMS

ANOVA	Analysis of Variance
CI	Confidence Interval
CONPLAN	Contingency Plan
CPX	Command Post Exercise
DRDC	Defence Research and Development Canada
EX	Exercise
JTFP	Joint Task Force Pacific
GNU GPL	GNU General Public License
HREC	Human Research Ethics Committee
MARPAC	Maritime Forces Pacific
MCDA	Multi-criteria Decision Analyses
MOE	Measure of Effectiveness
MOP	Measure of Performance
ORT	Operational Research Team
RCN	Royal Canadian Navy
SOTA	School of Thought Analysis
TA	Technical Authority
TG	Task Group
TGEX	Task Group Exercise
TTX	Table Top Exercise
U.S.	United States

Annex A. PROCESS CHART DEPICTING DATA ANALYSIS PLAN

Figure 1: Process chart depicting data analysis plan



Annex B. DESCRIPTION OF Basic ANALYTICAL TESTS AND TOOLS

B.1 GENERAL DESCRIPTION OF BASIC ANALYTICAL TESTS

The basic analytical tools that were selected for inclusion in the data analysis plan and identified in Table 3 are described in Table 4 below.

Table 4: Description of analytical tests

Type of Analysis	Description
Descriptive Analyses	
Related to Central Tendency	
Mean	The average score in a data set. [38]
Median	The midpoint in a sequence of numbers in a data set. [38]
Mode	The most commonly occurring number in a data set. [38]
Skewness	Skewness is used to assess the normality of the distribution. When a data set is normally distributed the value of skewness is zero. A distribution is positively skewed when there are extreme scores that are concentrated on the left and the right tail is long. A distribution is negatively skewed when there are extreme scores on the right and the left tail is long. Significance test for skewness compare the observed value to a null hypothesis of zero. When a distribution is skewed the median is the preferred measure of central tendency. [20][35]
Confidence Interval	The range of predicted values within which the true population parameter is expected, with a certain level of confidence to occur. The confidence interval is related to the sampling from the population. Confidence levels can be one-sided (either a lower or an upper threshold) or two-sided (both upper and lower thresholds). Typically associated with a probability value of .95 or .99. [20][57]
Dispersion	
Range	The difference between the highest and lowest numbers in a data set. [36]
Standard deviation SD	The calculated value that indicates the variation between the scores in the data set. [40]
Inferential (Univariate) Analyses	
Parametric	
Frequency count, relative frequency	A frequency count refers to the number of times an event is observed. Frequency is often reported as a measure of relative frequency where by the count of the subgroup is reported as a percentage or proportion of the total population that was counted. [58]

Type of Analysis	Description
Contingency tables	Contingency tables depict the relationship between two qualitative (categorical) variables. [59]
Correlation	A correlation depicts a measure of the relationship that exists between two variables. The strength of the relationship is inferred by the size of the correlation. [60]
One sample t-test	A one-sample t-test is performed to determine whether a sample comes from a population that has a specific mean. The mean of the population can be hypothesized if it is unknown. There are four assumptions which must be met in order to perform this test. [61]
Dependent (paired-samples) t-test	The dependent t-test compares the means between two related groups with respect to the same continuous variable. [62]
Independent samples (two sample) t-test	The independent samples t-test is an inferential test that is used to compare means derived from two samples to determine whether they are from different populations. [63]
Analysis of Variance (ANOVA)	The ANOVA is an extension of the t-test (and the z-test) which is used when there are more than two independent groups. A one-way ANOVA is used to compare two groups based on one independent variable. A two-way (factorial) ANOVA is used to compare groups using two independent variables; a two-way ANOVA also allows the researchers to examine possible interactions between the independent variables. [64]
ANOVA with repeated measures	The ANOVA with repeated measures is used to test dependent groups (e.g., pre-test and post-test). This ANOVA is used to test one categorical independent variable and one continuous dependent variable. [65]
Linear regression	A simple linear regression is an extension of correlation. It is used to predict the value of a variable. The predictions are based on the value of a different variable. [66]
Non-Parametric Tests	
Chi-square	A Chi-square is performed to determine whether there is a relationship between two nominal variables. In the cases where comparisons are made between two variables, the presence of a significant relationship is often interpreted as a significant difference between the variables. [29]
Spearman's Rank-Order Correlation	<p>The Spearman rank-order correlation is a non-parametric alternative to the Pearson product-moment correlation. It is used to analyze data that is measured at least at the ordinal level.</p> <p>A rank order assigns a level of priority based on a particular criterion. Ranking analyses were included in this section related to descriptive analyses because they are performed on ordinal data. However, ranking analyses can be performed on data sets that are associated with interval and ratio data providing the data have been transformed into ordinal data sets. [67]</p>
Wilcoxon Test	The Wilcoxon test is the non-parametric equivalent to the paired samples t-test. This test is used to determine whether there is a difference observed between the population distributions. The means are not used in this test because the data are not normally distributed. [68]

Type of Analysis	Description
Mann-Whitney Test	The Mann-Whitney test is the non-parametric equivalent to the independent samples t-test. This test is used to determine whether there is a difference observed between the medians. The means are not used in this test because the data are not normally distributed. [69]
Kruskall-Wallis Test	The Kruskal-Wallis test is the rank-based non-parametric alternative to a one-way ANOVA. It is also considered to be an extension of the Mann-Whitney U test. This test has four assumptions that must be met in order to ensure that the findings are reliable. [70]
Friedman Test	The Friedman test is non-parametric alternative to the one-way ANOVA. with repeated measures. It is typically used to compare groups when the dependent variable is ordinal but can also be used to analyze continuous data when the assumptions underlying one-way ANOVA with repeated measures are violated. [78] The Friedman test can also be used to determine whether there is a strong consensus between group rankings when the SOTA is performed. [27]

B.2 ONLINE STATISTICAL TOOLS

Table 5: Description of examples of online statistical tools

Tool	Description
Open source online statistical calculators	
VassarStats	VassarStats is a web-based tool for performing statistics. It claims to be user-friendly. It works best with Firefox, Safari and Chrome however, Internet Explorer is not recommended. It can be used on Windows and Macintosh platforms. It contains a simple graph maker. There are also pdf downloads (containing tables of critical values) and Visual Basic macros that can be downloaded for both Windows and Macintosh to work within MS Excel (One way Analysis of Covariance, linear correlation and regression). [45]
GraphPad Software QuickCalcs	GraphPad QuickCalcs is a web-based tool for performing statistics. It is used directly via the website and there is no requirement (or means) to download any files. A user profile is not required. Instructions for using the statistical tests are provided. There is no information regarding the use of graphics. The output can be captured using common data capture methods for web applications. This online statistical calculator is associated with a statistical program (GraphPad Prism) that is available for purchase. [46][46]
EasyCalculation	EasyCalculation is a web-based tool that contains several different categories of calculators including statistics. These statistical calculators can be used directly via the website however, there is minimal guidance provided for using the tests. A user profile is not required. There is no requirement to download any files however there is a fee-based option to download the calculators. There is no information regarding the use of graphics. The output can be captured using common data capture methods for web applications. [47]
Mathportal	Mathportal is a web-based tool for performing statistics that was created by a statistician. There are several categories of calculators including statistics. It is used directly via the website and there is no requirement to download any files. A user profile is not required. Instructions for using the statistical tests are provided and an option is provided for the users to ask the author statistics questions. There is no information regarding the use of graphics. The output can be captured using common data capture methods for web applications, shared via a link that is generated from the output page. [48]

Tool	Description
Social Science Statistics	Social Science Statistics is a web-based tool for performing statistics for hypothesis testing, p-value calculators and descriptive statistics. It was designed to optimize ease of use and clarity of presentation. It is used directly via the website and there is no requirement (or means) to download any files. A user profile is not required. Instructions for using the statistical tests are provided and the website also contains quizzes to improve understanding. There is no information regarding the use of graphics. The output can be captured using common data capture methods for web applications. [49]
Calculator.net	Calculator.net is a web-based tool that contains several different categories of calculators including statistics. These statistical calculators can be used directly via the website. A user profile is not required. There is no requirement to download any files. Instructions for using the statistical tests are provided. There is no information regarding the use of graphics. The output can be captured using common data capture methods for web applications. [50]
CalculatorSoup Online Calculator Resource	CalculatorSoup is a web-based tool that contains several different categories of calculators including statistics. These statistical calculators can be used directly via the website. A user profile is not required. There is no requirement to download any files. Instructions for using the statistical tests are provided. There is no information regarding the use of graphics. The output can be captured using common data capture methods for web applications. [51]
Open source downloadable statistical calculators	
The Statistics Calculator	The Statistics Calculator is a free downloadable version of a tool for performing statistics. Information about the tests is provided on the website however, the tool must be downloaded to the users' computer. [52]
Open source downloadable Gnu.org software packages	
DAP Stats and graphics	Dap is a GNU program. There is an existing manual that can be downloaded. It is a statistics and graphics package based on C. It can build on GNU/Linux and Unix platforms. It can be used to perform some univariate statistics, correlations, regression, ANOVA, categorical data analysis, logistic regression and non-parametric analyses. The specific categorical and non-parametric tests are unspecified in the information provided on the website. The graphics are not user friendly. [54]

Tool	Description
Datamash	Datamash is a GNU program. There is an existing manual that can be downloaded. It can run on various platforms including Unix platforms, Windows and MacOS. It is a command-line program that performs basic numeric, textual, and statistical operation on input textual data files. [55]
GNU PSPP	GNU PSPP is a statistical analysis program that works best on the GNU/Linux platform. There is an existing manual that can be downloaded. The program can be used via a graphical interface or by using the traditional syntax commands. It can create graphics including scatterplots, histograms and bar charts. The program is interoperable with other free software and can import data from spreadsheets, text files and database sources. There are multiple output formats including text, postscript, pdf, open document and html. [56]
R Programming Language	R is a free software that can be downloaded for statistical computing and graphics. It is an environment which enables a wide range of statistics to be performed. It can be downloaded for use on multiple platforms including Windows, Mac, Linux/UNIX. It is a GNU project and is available under the GNU GPL. The environment is written in the R dialect of S and it can be linked and called at run time with C, C++ and Fortran code. Online tutorials and courses can be accessed by the user community. [71]
Python	Python is a general-purpose interpreted, interactive, object-oriented and high-level programming language. It can be downloaded for use on multiple platforms including Windows, Mac, Linux/UNIX. It is available from python.org and under the GNU GPL. It is open source software and can therefore be modified and distributed. Online tutorials, user guides and examples of code that can be accessed by the user community. [72][73][74]
Proprietary online statistical calculators	
StatTools	StatTools is an add-on to Microsoft Excel which is downloaded for use within the MS Excel spreadsheet. The functions provided by the StatTools functions are not macro calculations; rather, they have been optimized using C++ and .DLLs. This add-on is equipped to perform over 30 different types of analyses. The user can define data sets and variables which are located within a single workbook or across different workbooks. Statistical tests can be selected from menus or users can write custom procedures using the Excel Developer Kit and Excels built-in Visual Basic for Applications (VBA) programming language. StatTools is equipped with data management, charting and reporting tools. This software download must be purchased, however, there is a free trial version that can be downloaded for a specific period of time. [75]

Tool	Description
XLStat	XLStat is an add-on to Microsoft Excel. There are two packages available; XLStat – Base contains the top 100 essential statistical functionalities whereas, XLStat – Premium contains over 200 XLStat features (including all of the XLStat – Base features). These packages provide access to descriptive analyses, univariate (parametric, non-parametric) and some multivariate analyses. The results can be visualized using the capabilities contained within the package. These software packages must be purchased, however, there are free trial versions that can be downloaded for a specific period of time.[76][77]

B.3 LINK BETWEEN ONLINE STATISTICAL TOOLS AND ANALYTICAL TESTS

Table 6 presents the link between the tests identified in the data analysis plan (Table 3) and the online statistical tools identified in Table 5.

Table 6: Link between online statistical tools and analytical tests

Tool	Descriptive Analyses					Hypothesis Testing Univariate Inferential Statistics																	
	Central Tendency			Disper- sion		Parametric										Non-Parametric							
	Mode	Median	Mean, CI, Skewness	Range (minimum and maximum scores)	SD	Frequency Counts	Correlation	Contingency Tables	One sample t-test	Dependent t-test	Independent t-test	ANOVA	ANOVA with repeated measures)	Linear regression	Post -Test	Chi Square	Spearman Rank-Order Correlation	Wilcoxon Test	Mann-Whitney Test	Kruskall-Wallis Test	Friedman Test	Post -Test	
Open source online statistical calculators																							
VassarStats			X		X		X	X	X	X	X	X	X	X		X	X		X	X	X		
GraphPad QuickCalcs			X		X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Tool	Descriptive Analyses					Hypothesis Testing Univariate Inferential Statistics																	
	Central Tendency			Disper-sion		Parametric									Non-Parametric								
	Mode	Median	Mean, CI, Skewness	Range (minimum and maximum scores)	SD	Frequency Counts	Correlation	Contingency Tables	One sample t-test	Dependent t-test	Independent t-test	ANOVA	ANOVA with repeated measures)	Linear regression	Post -Test	Chi Square	Spearman Rank-Order Correlation	Wilcoxon Test	Mann-Whitney Test	Kruskall-Wallis Test	Friedman Test	Post -Test	
EasyCalculation	X	X	X	X	X		X				X	X*		X	X	X	X						X
Mathportal	X	X	X	X	X		X		X	X	X			X					X				
Social Science Statistics							X	X	X	X	X	X*				X	X	X					
Calculator.net	X	X	X	X	X																		
Calculator Soup Online Calculator Resource	X	X	X	X	X																		

Tool	Descriptive Analyses					Hypothesis Testing Univariate Inferential Statistics																	
	Central Tendency			Disper-sion		Parametric										Non-Parametric							
	Mode	Median	Mean, CI, Skewness	Range (minimum and maximum scores	SD	Frequency Counts	Correlation	Contingency Tables	One sample t-test	Dependent t-test	Independent t-test	ANOVA	ANOVA with repeated measures)	Linear regression	Post -Test	Chi Square	Spearman Rank-Order Correlation	Wilcoxon Test	Mann-Whitney Test	Kruskall-Wallis Test	Friedman Test	Post -Test	
Open source downloadable statistical calculators																							
The Statistics Calculator			X		X	X				X	X	X			X	X							
Open source downloadable Gnu.org software																							
DAP Stats and graphics												X	X	X									
Datamash		X		X		X																	
PSPP	X	X	X	X	X	X	X	X	X	X	X			X	X**	X	X	X	X	X	X	X	X**

Tool	Descriptive Analyses					Hypothesis Testing Univariate Inferential Statistics																	
	Central Tendency			Disper-sion		Parametric										Non-Parametric							
	Mode	Median	Mean, CI, Skewness	Range (minimum and maximum scores)	SD	Frequency Counts	Correlation	Contingency Tables	One sample t-test	Dependent t-test	Independent t-test	ANOVA	ANOVA with repeated measures)	Linear regression	Post -Test	Chi Square	Spearman Rank-Order Correlation	Wilcoxon Test	Mann-Whitney Test	Kruskall-Wallis Test	Friedman Test	Post -Test	
Proprietary Software																							
StatTools**	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
XLStat**	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

*Only one-way ANOVA not two-way ANOVA; **Post hoc comparisons only (not planned comparisons)

** Based on the summaries and Excel capabilities (validation was not completed)