Principal Investigators:
Professor David Playford
dplayford@neda.net.au
Professor Geoff Strange
gstrange@neda.net.au

Statistical Analysis Plan v1.2 (May 2018)

Prepared by Professor Simon Stewart,

Professor Geoff Strange and Professor David Playford

Contents

Principal Investigators:
Professor David Playford
dplayford@neda.net.au
Professor Geoff Strange
gstrange@neda.net.au

FUNDING

NEDA has been funded by grants from Actelion Pharmaceuticals Australia, Bayer Pharmaceuticals and Glaxo Smith Kline Pharmaceuticals Australia. NEDA is part of funding for two Partnership Grants via the NH&MRC

STUDY REGISTRATION

NEDA is registered with the publically accessible Australian New Zealand Clinical Trials Registry (ANZCTR): ACTRN12617001387314

https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=373535

1. BRIEF BACKGROUND

Echocardiography is the most commonly used cardiac imaging tool worldwide, providing systematic evaluation of cardiac structure and function. Recent advances in echocardiography imaging, reporting and archiving has made measurement and report information more amenable for extraction of entire data sets via a "big data" approach.

The National Echocardiography Database Australia (**NEDA**) is a new echocardiography database across Australia intended to collate all digital echocardiography measurement and report data from centres (inclusive of public and private echocardiography laboratories) to produce a master database comprising standardised imaging findings on millions of patients. As an initial step, this database will be linked with the National Deaths Index (NDI) in Australia

provided by Australian Institute for Health and Welfare (AIHW) to obtain mortality data for each included individual and to explore (including artificial intelligence modelling) the prognostic implications of a broad range of cardiac abnormalities identified by echocardiography. In the future, other linked health outcomes will broaden the impact of this unique and invaluable resource.

Principal Investigators:
Professor David Playford
dplayford@neda.net.au
Professor Geoff Strange
gstrange@neda.net.au

2. STUDY AIMS & OBJECTIVES

The primary aim of NEDA is to combine large-scale, real-world echocardiographic data with linked data wherever possible (including mortality data) to investigate the characteristics and prognostic implications of a wide variety of cardiovascular abnormalities quantified by echocardiography performed in centres across Australia. Based on clinical and scientific priorities (determined by the NEDA Steering Committee and external scientific requests to access and analyse NEDA data) the NEDA Investigators aim to produce a range of high impact, high value research reports that appropriately exploits the scientific strengths of a large source of standardised echocardiographic data; particularly when combined with individually linked health outcome data (e.g. all-cause mortality, cause of death and hospital activity).

In creating a large and contiguous source of echocardiographic data from multiple sources, with the capacity to combined findings with individually linked health outcome data (through data linkage) and apply innovative analyses (e.g. through the application of artificial intelligence algorithms), the NEDA Investigators seek to produce definitive reports that positively influence clinical practice and the interpretation of echocardiographic data resulting in better health outcomes for patients investigated with echocardiography.

3. STUDY DESIGN

NEDA is a hybrid observational study that will firstly capture individual echocardiographic data (combined with basic demographic profiling) on a retrospective and prospective basis and then combine such data with other available data on clinical profile, treatment and health outcomes using data linkage techniques.

A full description of the study protocol (rationale, design and methodology) underlying NEDA has been submitted for peer-review publication and will be made available as an "open access" manuscript and via the NEDA website (https://www.neda.net.au/research-results/) once published.

NEDA conforms to the Australian National Statement on Ethical Conduct of Human Research (2015) and the Declaration of Helsinki (2013). Both currently and in the future, appropriate ethical approvals have and will be obtained from relevant Human Research Ethics Committees (HREC).

Principal Investigators:
Professor David Playford
dplayford@neda.net.au
Professor Geoff Strange
gstrange@neda.net.au

3.1 Study Setting

Australia has a hybrid public/private health care system with universal/subsidized health care for all Australians (including hospital care, primary care, investigations and treatment) via the "Medicare" system. Private health insurance is additional health cover to that provided under Medicare, to reimburse all or part of the cost of hospital and/or ancillary services incurred by an individual. According to the Australian Bureau of Statistics, in 2011-12 there were 9.7 million adult Australians with private health insurance; representing 57.1% of adult Australians. .

It is envisioned that NEDA will collect data from all States and Territories and from both public and private health facilities.

A continuously updated list of NEDA participating centres is available at - https://www.neda.net.au/participating-sites/

3.2 Study Cohort

All individuals (regardless of age and sex) who have and will undergo an echocardiographic investigation at a NEDA participating centre will comprise the study cohort: currently >350,000 individuals, 50.4% male and with a mean age of 60.8±18.6 years.

For retrospectively enrolled cases, a patient consent waiver has been authorized by each relevant HREC based on the National Health and Medical Research Council of Australia's Statement on Ethical Conduct of Human Research https://www.nhmrc.gov.au/guidelines-publications/e72.

For prospectively enrolled cases, consent is obtained by a verbal script provided to each participating laboratory (with an opt-out option provided).

The broad profile and characteristics of the NEDA cohort will be continuously updated via submitted conference abstracts and published manuscript; the details of which are available at - https://www.neda.net.au/participating-sites/

Principal Investigators:
Professor David Playford
dplayford@neda.net.au
Professor Geoff Strange
gstrange@neda.net.au

3.3 Study Data

NEDA is designed to extract data from digital echocardiography laboratories across Australia, and combine detailed cardiac measurement data into a central database via a "vendor-neutral" data extraction process that transfers every measurement for each echocardiogram performed in an entire echocardiography database into a standard NEDA data format.

A continuously updated NEDA Data Dictionary (including all primary data, those derived or imputed from primary data and subsequently derived/imputed in future analyses) is maintained

Accordingly, precise definitions for each echocardiography variable (using both a short string and a longer description) have and will be formulated. Variables with the same name as the NEDA standard are automatically matched, and those with a different name are manually matched with the NEDA standard by a Principal Investigator. Duplicate measurements with different names are combined together. Units are transformed to the single NEDA standard (e.g. centimetres per second, cm/s; milliseconds, ms; millimetres of mercury, mmHg). Repeated measurements for the same variable are converted to a single variable according to the study protocol.

Impossible variables (e.g. zero or negative ejection fraction, or gradients well above any reported range) are removed using a data range limiting tool. Duplicate echocardiograms (all variables identical including the date of study) are removed, and echocardiograms for the same patients are merged into serial studies with the same unique ID. Each individual patient within NEDA is then automatically provided with a unique identifier, and each echocardiogram for that individual is coded with the same unique patient identifier allowing serial echocardiograms to be compared over time.

Once imported into NEDA, the data populates the Master NEDA Database that forms the basis for all subsequent analyses and outputs. According pre-specified standard operating procedures, this Database will be updated (and given a new version name) following each analysis that generates new imputed or derived variables from internal or external sources including linked outcome data (e.g. Master NEDA Database v1.1_date) AND following the import of new echocardiographic data from participating centres (e.g. Master NEDA Database v2.0_date).

Variables are continuously inspected using graphical and statistical tools, and compared with existing NEDA variables to confirm all measurements have been correctly imported in format, units and range. Any anomalies can be corrected and import process repeated.

Principal Investigators:
Professor David Playford
dplayford@neda.net.au
Professor Geoff Strange
gstrange@neda.net.au

4. DATA MANAGEMENT

The NEDA Server is housed in the cloud and hosted by Azure, in Microsoft's Australia East data centre in New South Wales. All communication/transmissions are password protected, and performed using secure communication technologies (SSL encryption). SSL (Secure Sockets Layer) is an industry standard security technology for establishing an encrypted link between a validated server and a client as used by many Government and Banking systems. The client validates that it is talking to the real and correct server using the public key. All communications are then encrypted using the best protocol decided on by both the client and server, as long as it meets both the client and server's minimum acceptable security settings (controlled by the Operating Systems). This encryption ensures that only the trusted server is able to decrypt the contents of the transmission from the client.

The NEDA database within the NEDA servers are configured to use the highest level of security, with certification cross-checks to ensure current best practice is followed. Access to the NEDA database is restricted only to NEDA principal investigators and developers. Each login to the NEDA database is logged and can be traced based on originating IP address. Identified patient information is suppressed within the NEDA database, and is not visible to investigators, with only the random code being visible. All analysis of NEDA data is performed using a de-identified copy of the NEDA database – the NEDA Master Database.

5. STUDY ANALYSES

All analyses of NEDA data will be generated from the Master NEDA Database with an investigation specific database (e.g. Pulmonary Hypertension Database v1.0_date) generated for each distinct analysis and formatted according to the pre-specified NEDA Data Dictionary.

5.1 Key principles for study analyses

As per the pre-specified governance structure (https://www.neda.net.au/about-us/), all analyses require approval from the NEDA Steering Committee via a formal approval protocol. The following principles will be adhered to maintain the highest standards of scientific practice:

1. Data quality and validity will be protected via limited/authorised access to the Master NEDA Database (Principal Investigator(s) and Data Management Team)
2. Each analysis of NEDA data will require a formal research hypothesis/research question, submitted on a standard pro-forma that is considered (by the NEDA Investigators) to scientific merit and validity

Principal Investigators:
Professor David Playford
dplayford@neda.net.au
Professor Geoff Strange
gstrange@neda.net.au

3. A specific (but brief if appropriate) Statistical Analysis Plan (SAP) that is congruent with this document and the principals of data handling, analyses and interpretation will be developed to adequately address research questions/hypotheses. Each subsequent SAP will be appended to this document and used to inform future analyses.
4. Data extracted from the Master NEDA Database for the purpose of a specific analysis will be prepared in a widely accepted and validated statistical analysis program tool (e.g. SAS or SPSS) according to the NEDA Data Dictionary (with appropriate updates should new variables be generated).
5. All statistical analyses will be undertaken by appropriately qualified researchers with experience in working with equivalent data to NEDA and the research program used to derive analyses.
6. Any analysis of data will firstly involve basic descriptors according to the age and sex of cases.
7. Both clinical practice and statistical considerations (according to the normal distribution of the variable of interest) will inform the thresholds for determining potential correlations between and within echocardiographic findings and health outcomes.
8. Any statistical analyses requiring advanced multivariate analyses and/or modelling will be undertaken (or at least verified) by a qualified statistician approved and/or nominated by the Principal Investigator(s).
9. Given the nature of NEDA, all outputs derived from this study will adhere to the latest standards of the **STrengthening the Reporting of OBservational studies in Epidemiology** group – see below.
10. All statistical analysis will include the relevant syntax, saved to a file and kept within the same dropbox folder as the database.

Principal Investigators:
Professor David Playford
dplayford@neda.net.au
Professor Geoff Strange
gstrange@neda.net.au

5.2 Applying best-practice STROBE guidelines

Table 1 is a check-list of the current requirements for reporting against these standards with specific commentary relevant to the conduct and reporting of NEDA data.

https://www.strobe-statement.org/index.php?id=strobe-home (Accessed March 2018)

Table 1: STROBE Reporting Checklist Relevant to the NEDA Study

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract [This will always include the term Database plus the terms "retrospective", "prospective" and "linked data" to describe underlying NEDA methodology"]
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found [Scientific oversight from the NEDA Steering Group]
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported [Broad background provided by the NEDA protocol plus specific clinical issues added for each subsequent analysis of NEDA data]
Objectives	3	State specific objectives, including any pre-specified hypotheses [Included in the pre-specified SAP for that specific analysis/output]
Methods		
Study design	4	Present key elements of study design early in the paper [Reference to NEDA Rationale/Design paper, NEDA website and specific elements of analysis (e.g. sub-set of NEDA cohort in whom data were available) clearly stated]

Principal Investigators:
Professor David Playford
dplayford@neda.net.au
Professor Geoff Strange
gstrange@neda.net.au

Steering committee:
Professor David Celermajer
Professor Tom Marwick
Associate Professor David Prior
Professor Greg Scalia
Dr Majo Joseph
Dr Marcus Ilton
Professor Eli Gabbay
Professor Simon Stewart
Professor Max Bulsara
Professor Jim Codde

Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection [Reference to NEDA Rationale/Design paper and details that completely match those of the current NEDA Master Database and any imported data linkage (e.g. censoring point for linked National Death Index data]
Cases	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of cases. Describe methods of follow-up [Reference to NEDA Rationale/Design paper, NEDA website and specific elements of analysis (e.g. sub-set of NEDA cohort in whom data were available) clearly stated]
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable [Reference to NEDA Rationale/Design paper, NEDA website and NEDA Data Dictionary with diagnostic criteria (e.g. Pulmonary Hypertension derived from an imputed RVSP threshold] and outcome data critical to specific analyses and reporting provided included in the pre-specified SAP for that specific analysis/output]
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group [Reference to NEDA Rationale/Design paper, NEDA website and NEDA Data Dictionary with diagnostic criteria (e.g. Pulmonary Hypertension derived from an imputed RVSP threshold] and outcome data critical to specific analyses and reporting provided included in the pre-specified SAP for that specific analysis/output.

Principal Investigators:
Professor David Playford
dplayford@neda.net.au
Professor Geoff Strange
gstrange@neda.net.au

Steering committee:
Professor David Celermajer
Professor Tom Marwick
Associate Professor David Prior
Professor Greg Scalia
Dr Majo Joseph
Dr Marcus Ilton
Professor Eli Gabbay
Professor Simon Stewart
Professor Max Bulsara
Professor Jim Codde

		Any comparator groups (e.g. according to the presence absence of Pulmonary Hypertension or different thresholds of RVSP) will also be clearly identified in the pre-specified SAP for the specific analysis/output, with the number of cases with available echocardiographic and/or outcome data relative to total NEDA cases clearly identified]
Bias	9	Describe any efforts to address potential sources of bias [Reference to the NEDA website, number and relative contributions of participating centres and potential bias from – a) patients being investigated at those centres, b) the characteristics of the Australian Health Care System and patient reimbursement for investigations, c) clinical bias in reporting specific echocardiographic variables (resulting in missing data or requiring data imputation), d) purposeful exclusion of data based on pre-specified analyses and/or e) lack of available linked health outcome data in a clearly defined LIMITATIONS section in any manuscripts arising from NEDA]* *Any description of the limitations of NEDA needs to be balanced by the strength of – a) volume of data (largest known database of its kind) and b) non-selection bias of data acquisition.*
Study size	10	Explain how the study size was arrived at [Reference to the NEDA Master Database and current number of available participant and echocardiographic investigations available for analysis). Based on all available data only with no artificial exclusions permitted. The number of individual cases included in specific analyses will be outlined in the pre-specified SAP for that particular analysis]
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why [Refer to NEDA Data Dictionary for initial handling of quantitative variables. All imported (e.g. NDI data with date and coded cause of death), newly imputed or calculated variables to be clearly outlined in the pre-specified SAP for that particular analysis]

Principal Investigators:
Professor David Playford
dplayford@neda.net.au
Professor Geoff Strange
gstrange@neda.net.au

Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding [Reference to NEDA Statistical Analysis Plan – see Statistical Analyses section for broad principals of analysis plus all specific analyses clearly outlined in the pre-specified SAP for that particular analysis. Inclusion of an experienced Statistician to supervise any advanced statistical analyses]
		(*b*) Describe any methods used to examine subgroups and interactions [Reference to NEDA Statistical Analysis Plan – see Statistical Analyses section for broad principals of analysis plus all specific analyses clearly outlined in the pre-specified SAP for that particular analysis. Inclusion of an experienced Statistician to supervise any advanced statistical analyses]
		(*c*) Explain how missing data were addressed [All variables will be described according to numerator/denominators with missing data clearly identified. A priori exclusion of cases with missing data relevant to the research hypothesis/question will be identified in the pre-specified SAP for that particular analysis: as will be any critical variables that need to be imputed or calculated from available data in the Master NEDA Database (e.g. RVSP)]
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed [Limitations of linkage to external data (e.g. outcome of NDI matching and determination of death) clearly explain in the METHODS section of reports. *Case-control study*—If applicable, explain how matching of cases and controls was addressed [If case-control methods are applied these will be reviewed by an experienced Statistician) and will be clearly described in the pre-specified SAP for that particular analysis]

Principal Investigators:
Professor David Playford
dplayford@neda.net.au
Professor Geoff Strange
gstrange@neda.net.au

		Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy [If advanced sampling methods applied (beyond availability of data) these will be reviewed by an experienced Statistician) and will be clearly described in the pre-specified SAP for that particular analysis]
		(*e*) Describe any sensitivity analyses
Results		
Cases	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed [Refer to the NEDA website, NEDA Master Database. Reported numbers will be clearly described in the pre-specified SAP for that particular analysis]
		(b) Give reasons for non-participation at each stage [Refer to the NEDA website, NEDA Master Database. Reported numbers will be clearly described in the pre-specified SAP for that particular analysis]
		(c) Consider use of a flow diagram [It will be standard practice for every NEDA report to include a flow-diagram demonstrating how many NEDA cases were available for analysis and how many specific cases are being reported on]
Descriptive data	14*	(a) Give characteristics of study cases (eg demographic, clinical, social) and information on exposures and potential confounders [See Statistical Analyses section]
		(b) Indicate number of cases with missing data for each variable of interest [All participating cases will be described according to numerator/denominators with those excluded clearly identified in the study flow diagram (13.c). A priori exclusion of cases with missing data

Principal Investigators:
Professor David Playford
dplayford@neda.net.au
Professor Geoff Strange
gstrange@neda.net.au

		relevant to the research hypothesis/question will be clearly identified in the pre-specified SAP for that particular analysis]
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount) [The nature of all linked outcome data (e.g. NDI data from the time of echocardiographic assessment to death or study censor point expressed in days) will be clearly identified in the pre-specified SAP for that particular analysis]
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time
		Case-control study—Report numbers in each exposure category, or summary measures of exposure
		Cross-sectional study—Report numbers of outcome events or summary measures [All outcomes measures identified in the pre-specified SAP for that particular analysis will be reported in subsequent study reports]
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included
		(*b*) Report category boundaries when continuous variables were categorized
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period [See Statistical Analyses for reporting of unadjusted, sex-specific and age-specific categorical and continuous data. All adjusted analyses will be accompanied by 95% Confidence Intervals (CI's) where appropriate. All advanced statistics/modelling (and subsequent outputs) will be supervised by an experienced bio-statistician]
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses

Principal Investigators:
Professor David Playford
dplayford@neda.net.au
Professor Geoff Strange
gstrange@neda.net.au

Steering committee:
Professor David Celermajer
Professor Tom Marwick
Associate Professor David Prior
Professor Greg Scalia
Dr Majo Joseph
Dr Marcus Ilton
Professor Eli Gabbay
Professor Simon Stewart
Professor Max Bulsara
Professor Jim Codde

		[All reported analyses will be clearly identified in the pre-specified SAP for that particular analysis and will be reported in subsequent study reports]
Discussion		
Key results	18	Summarise key results with reference to study objectives [Reference to a priori study objectives outlined in the SAP for that particular analysis]
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias [Reference to the NEDA Rationale and Design paper]
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence [Expert overview of the NEDA Steering group]
Generalisability	21	Discuss the generalisability (external validity) of the study results [Reference to the NEDA Rationale and Design paper]
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based [Provided by the NEDA Investigators for primary funding of the NEDA Study with clear reference to any external sources relevant to particular analyses or authors]

Principal Investigators:
Professor David Playford
dplayford@neda.net.au
Professor Geoff Strange
gstrange@neda.net.au

5.3 Statistical Analyses

All individual analyses of NEDA data will follow the general statistical principles outlined in this document. As described above, all planned analyses of NEDA data to test a priori research hypotheses/questions will be accompanied by a project-specific SAP that outlines – **a**) the research question/hypothesis being tested, **b)** the sub-set of NEDA relevant to answering the research question/hypothesis, **c)** the primary health outcomes (if linked health outcome data are being examined) and echocardiographic data (pre-existing or derived from the current NEDA Master Database) of interest and **d)** specific statistical analyses used to formally test the research questions/hypothesis being examined.

Description of the baseline characteristics of all included cases will be presented by pre-specified groups and according to age (in age-groups) and sex where appropriate. Discrete variables will be summarised by frequencies and percentages (with 95% CI where appropriate). Percentages will be calculated according to the number of patients for whom data are available. Where values are missing, the denominator will be stated in the corresponding summary table, in either the body or a footnote. In some instances, additional frequencies and percentage of patients in each category will be reported as indicated below. Continuous variables will be summarised by standard measures of central tendency and dispersion, using mean and SEM, as well as median and IQR where appropriate.

Unless otherwise specified, any between group comparisons (including those according to age and sex) will be assessed by Student's t-tests, Mann Whitney U test (for non-normally distributed continuous data) and Chi-squared test (with calculation of odd ratio's [OR] and 95% confidence intervals [CI's]) where appropriate.

All-cause and cause-specific mortality data (utilising time from the index echocardiographic investigation to either death or censoring during a pre-specified follow-up period) will be analysed using Kaplan-Meir survival analysis with log rank test for any group comparisons.

Principal Investigators:
Professor David Playford
dplayford@neda.net.au
Professor Geoff Strange
gstrange@neda.net.au

Fixed (e.g. 5-year survival) and time-dependent survival analyses will be further performed with Multiple Logical Regression and a Proportional Hazards Model (where proportional hazards is confirmed) to derive adjusted OR's and hazard ratios (HR's), respectively. Where key data are not available in all cases (e.g. left ventricular ejection fraction), separate models will be constructed. Specific methods (e.g. back-ward step-wise model) and analyses of the predictive accuracy of generated models (generation of Receiver Operating Characteristics curves) will be routinely performed.

Unless otherwise specified (e.g. multiple comparisons requiring Bonferroni correction), statistical significance will be accepted at the level of $p<0.05$ (two-sided).