HHS Public Access

Author manuscript

J Pain Symptom Manage. Author manuscript; available in PMC 2024 January 01.

Published in final edited form as:

J Pain Symptom Manage. 2023 January ; 65(1): e87–e95. doi:10.1016/j.jpainsymman.2022.08.004.

A Visual Tool to Help Develop a Statistical Analysis Plan for Randomized Trials in Palliative Care

Gregory Samsa, PhD[1], **Kathryn Colborn, PhD**[2], **Maren Olsen, PhD**[1], **Gina-Maria Pomann, PhD**[1], **Steven Grambow, PhD**[1], **Megan Neely, PhD**[1], **Jesse Troy, PhD**[1]

[1]Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham NC, USA

[2]Department of Surgery, University of Colorado School of Medicine, Denver CO, USA

Abstract

Collaboration with a statistician about the design of a statistical analysis plan can be enhanced by illustrating how statisticians conceptualize their task. This conceptualization can be represented by a directed acyclic graph (DAG), which illustrates the statistician's approach and also provides an actionable tool to assist in the development of the plan.

Keywords

constructivism; randomized trials; statistical consultation; statistical analysis plan

Introduction

The Palliative Care Research Cooperative Group (PCRC) supports palliative care research in a number of ways, including regular "Intensive" short courses on the design and conduct of randomized trials (1). The format of these Intensives combines didactic instruction and experiential learning, the latter of which is accomplished by dividing the participants into teams and having each team design a trial. Faculty members accompany the teams, and a notable feature is that each team includes a statistician.

Collaboration with a statistician can be enhanced by understanding how they conceptualize questions of study design and data analysis. In the language of constructivism: by explicitly encountering their "mental maps" of these topics. One such mental map pertains to the development of a statistical analysis plan (SAP), an outline of which is one of each team's work products.

Here, we use a figure to describe how statisticians typically conceptualize SAPs, and then illustrate how this figure can be used to develop an outline of a generic SAP for a

Corresponding author: Gregory Samsa PhD, 11084 Hock Plaza, Durham NC 27710, USA, greg.samsa@duke.edu.

randomized trial of a pain coping intervention in palliative care. It is our hope that making this conceptual framework explicit can help support more productive interactions between palliative care researchers and statisticians.

Operational definition of the SAP

For the purposes of the Intensives, the SAP is simply an outline of the crucial analyses to be performed. This outline supports study design by, for example, ensuring that crucial variables are identified and collected, and that they are sufficient to support the study aims. In practice, the SAP is one of a set of interrelated documents around which the entire research project is organized. Other key documents include the study protocol, which describes in detail how the study will be executed (and includes key elements of the SAP), the data dictionary, which among others attaches names to the variables used in the SAP, the data management plan, which describes the steps which move from the raw data to the analysis files required by the SAP, and the data analysis plan, which provides a programmers-eye perspective on how the SAP will actually be executed. The data dictionary, data management plan and data analysis plan assist those responsible for statistical programming to accomplish their tasks efficiently and reproducibly.

Appendix 1 illustrates some of the components of the SAP for a randomized trial. The level of detail within the SAP varies across trials, with “applicable clinical trials” (2), which are unusual within palliative care, typically having the most specificity. SAPs can potentially be modified over time, so long as a transparent process exists for approving and recording these modifications. For example, the SAP would track important decisions made by the study team during the data analysis process, such as removers of specific confounders because they exhibited a large proportion of missing data or decisions to transform numerical data to achieve stable estimates. As another example, if during the conduct of the trial the science advances to the point where additional hypotheses are suggested, such hypotheses can be accommodated by revising the SAP.

Conceptual framework: classifying the analysis variables

Figure 1 illustrates the statistician’s conceptual framework. The boxes represent constructs, each of which will be operationalized by one or more study variables. The arrows represent causal (directional) relationships, and the width of the arrows indicates the strength of the relationship in question.

First considering the boxes, the intervention (box A) is the primary predictor. In a randomized trial the intervention is the study group to which patients are randomized.

Box B represents factors which affect the strength of the intervention effect. These are often termed “effect modifiers”. For example, if a pain coping intervention is more effective for patients with a primary diagnosis of cancer than for patients with other primary diagnoses, then primary diagnosis is an effect modifier. Within statistical models, effect modifiers are represented by “interaction terms”, and so statisticians sometimes refer to both as interactions (3).

Box C represents other factors that affect outcome (i.e., “covariates”).

Boxes D through F represent outcome variables, arranged by location on the causal pathway from intervention to outcomes. Mechanistic outcomes (box D) are directly caused by the intervention -- for example, for an intervention which focuses on pain coping techniques, an increased knowledge of pain coping techniques is a mechanistic outcome. Improved pain coping techniques might more directly affect a proximal outcome such as maximum daily pain score (box E), and might less directly affect a distal outcome such as overall quality of life (box F).

Figure 1 implicitly assumes that the only way that the intervention affects distal outcomes is through the associated proximal outcome. This is a simplifying assumption, and if for example a direct relationship between the intervention and quality of life exists then an additional arrow would be drawn. Because of the flow of causation, not only can the relationship between A and D be assessed, but also the relationship between A and E and the relationship between A and F. Finally, figure 1 contains a single line between the covariates C and the outcomes D-F, even though some covariates might be directly related to some but not all of the outcomes. Whether to use the same set of covariates for all analyses for the sake of consistency or, instead, to directly link covariates with specific outcomes, depends on context.

Conceptual framework: creating a generic SAP

Key to understanding how statisticians conceptualize SAPs is to recognize that, because their main goal is to assess the impact of the interventions being studied, essentially all RCTs can be represented by figure 1. Thus, the statistician can be counted upon to try to place all the study variables they encounter into boxes A-F.

Moreover, figure 1 contains relatively few arrows, and each of these arrows corresponds to a potential element of the SAP. For example, the arrows from A to D, from A to E (i.e., through D), and from A to F (i.e., through D and E) all represent direct assessments of the impact of the intervention, which only differ according to the choice of outcome.

The intention of an analysis can usually be described through a directional relationship (e.g., A -> E). Once such a relationship is identified, the statistician proposes a specific analytical technique based upon considerations such as scale of measurement. For example, with A binary and E continuous, the relationship could be assessed using a t-test.

To illustrate this process for a randomized trial of a pain coping intervention, the specific aims might be:

- Aim 1: (primary) Compare the pain coping intervention to usual care on the primary outcome of daily pain score at 3 months.
- Aim 2: (secondary) Determine whether the efficacy of the pain coping intervention (on daily pain score at 3 months) depends on primary diagnosis (i.e., cancer versus non-cancer).
- Aim 3: (exploratory) Compare the pain coping intervention to usual care on the exploratory outcome of quality of life at 3 months.

The SAP might include text such as the following:

- Aim 1 will be addressed using a t-test with study group as the (binary) predictor variable and daily pain score at 3 months as the (continuous) outcome. Recognizing that the t-test is a special case of a linear regression model with study group as its only predictor, this primary (unadjusted) analysis will be supplemented by a linear regression model with study group as the primary predictor, the covariates age (continuous), gender (binary) and primary diagnosis (binary) as adjustment variables, and daily pain score at 3 months as the outcome. These analyses will be preceded by an exploratory mechanistic analysis which compares the study groups on the (continuous) outcome of knowledge of pain coping skills.
- Aim 2 will use a similar linear regression model as aim 1, but with the primary predictors consisting of study group, primary diagnosis, and the interaction between study group and primary diagnosis. This interaction term tests whether the efficacy of the pain coping intervention depends on primary diagnosis, and is the focus of this aim.
- Aim 3 will use essentially the same approach as aim 1, but with quality of life as the (continuous) outcome variable.
- Various assessments of consistency will also be performed. Among these will be a regression model which assesses whether those patients with the greatest reduction in daily pain achieve the greatest improvement in quality of life. Additional analyses will, among others, assess the fidelity by which the intervention was delivered (not illustrated here).

In creating the above text, the statistician uses: (1) figure 1; (2) the study aims; and (3) the scale of measurement of the variables in the various analyses. The linkage between the DAG and the study aims generates the core of the analysis plan (A*B denotes the interaction between A and B):

- Aim 1: exploratory -- impact of study group on knowledge: A -> D
- Aim 1: primary -- impact of study group on daily pain, unadjusted: A -> E
- Aim 1: primary -- impact of study group on daily pain, adjusted: A, C -> E
- Aim 2: secondary -- consistency of intervention effect across primary diagnosis, unadjusted A, B, A*B >- E
- Aim 2: secondary -- consistency of intervention across primary diagnosis, adjusted: A, B, A*B, C -> E
- Aim 3: exploratory -- impact of study group on quality of life, unadjusted: A -> F
- Aim 3: exploratory -- impact of study group on quality of life, adjusted: A, C -> F
- Supplemental -- impact of change in daily pain on change in quality of life (among others): E -> F

Figure 2 illustrates applying this process for aim 1. The constructs mentioned in the specific aims have been operationalized as variables within boxes A-F. The “box and arrow version” of the primary unadjusted analysis results in an arrow from box A to box E, colored green. Once this arrow is identified, the scale of measurement of the variables in these boxes is used to propose a statistical technique: here, because the predictor is binary and the outcome is continuous, a t-test is proposed. To create an adjusted analysis, the orange box and arrows are added, leading to a linear regression model.

Comments on the generic SAP

The above SAP treats the unadjusted assessment of the impact of the intervention as primary and the covariate-adjusted version as supplemental. Some trialists prefer it to be the other way around, the argument being that even though randomization should ensure approximate balance between the treatment groups, including covariates in the model removes some of the noise in the outcome variable and thus yields more precise conclusions, especially if the covariates are strong predictors of outcome. In our experience, with individual-level randomization approximate balance among study groups is typically achieved, but when a cluster or clinical site is the unit of randomization achieving balance can be more problematic (i.e., the smaller the site, the less likely it is to achieve balance, and the greater the potential for bias in the estimate of treatment effect). Much of the benefit of the SAP derives from pre-specifying the analyses: following this principle it is critical to pre-specify whether the adjusted or unadjusted analysis will be considered primary, and also either (a) the set of adjustment variables; or (b) a process for identifying the set of adjustment variables which will be implemented before assessing the impact of the intervention. Publicly archiving the SAP before data analysis begins can provide additional credibility.

Most trials include mechanistic, proximal and distal outcomes, and the main question becomes which outcome to treat as primary. In general, as the causal pathway moves from left to right (a) the strength of the intervention effect decreases; and (b) the practical significance of the outcome increases. For example, the practical significance of an increase in knowledge of pain coping skills is much less than an improvement in quality of life. A small pilot trial might only have enough statistical power to address a mechanistic outcome such as knowledge of pain coping skills, the rationale being that this is a necessary (albeit not sufficient) condition for a pain coping intervention to have downstream effects. A definitive efficacy study should be powered around outcomes that matter to patients.

In general, hypotheses involving interactions (i.e., “the intervention works best for some subgroups”) require larger sample sizes than those involving main effects (i.e., “the intervention works overall”). Randomized trials will sometimes define the primary hypothesis in terms of the overall impact of the intervention, power the study accordingly, describe one or more interactions as secondary hypotheses, and note that the study is powered to detect dramatic interactions but not more subtle ones. A follow-up study might assess interactions more definitively.

Exploratory hypotheses vary from trial to trial. Some can involve non-directional relationships -- for example, within E it might be asked whether those patients with the

largest changes in daily pain score also had the largest changes in maximum pain score. Fidelity to the intervention protocol is often assessed within the intervention group only.

Creating the analysis variables

As illustrated above, each of the constructs represented within figure 1 is operationalized by one or more analysis variables. For example, the intervention in box A is operationalized by the single variable “study group”. The covariates in box C typically include multiple variables such as age, gender, cancer stage, general health status, etc. Covariates would be moved from box C to box B if they affect the efficacy of the intervention.

Defining the outcome variables in boxes D-F requires accomplishing two conceptually distinct tasks. First, the outcome variables which represent the constructs of mechanistic, proximal, and distal outcomes must be selected -- for example, which pain scale should be used, whether the analysis should focus on average daily pain or maximum daily pain, etc. Second, how the outcome variables will enter the SAP must be determined. For example, if pain is measured at baseline, 1 month, 3 months, 6 months and 12 months, perhaps the primary outcome variable will be change from baseline to 6 months, with secondary outcomes being based on change from baseline to 1 month, baseline to 3 months, and baseline to 12 months. Perhaps the ensemble of longitudinal pain scores will be summarized as a single area under the curve. Determining how to operationalize the outcome variable can be particularly complex when multiple observations are taken over time. Your statistician will be trying to assist in finding a version of the outcome variable which (a) matches the underlying construct as much as possible; (b) has good measurement properties (4); and (c) is intuitive to the non-statistician investigators. The PCRC’s instrument library is an excellent place to start.

Link to power and sample size calculations

Without devolving into a treatise on sample size calculations, we simply note that the link between Figure 1, the study design, the study hypotheses, and the sample size calculations within the SAP is formed by recognizing that (1) the sample size calculation is based on the test of the primary study hypothesis; and (2) ignoring pilot studies, this hypothesis is either A -> E or A -> F, possibly accounting for any gains in precision which are induced by including the covariates C in the statistical model.

Additional considerations in developing the SAP

So long as its assumptions are satisfied (an important caveat), a statistically sophisticated analytic framework will tend to be more efficient than a more basic one. Nevertheless, for developing the SAP we recommend starting with an approach which is accessible to all investigators and then building from there, since the better the statistical methods are understood the more likely they are to be consistent with the underlying science. It can sometimes be helpful to describe a more sophisticated approach in simpler terms. For example, the text of a grant application might include: “Longitudinal outcomes will be processed with maximal statistical efficiency using a linear mixed model with patient as a random factor, time, intervention, their interaction and various covariates as fixed factors, with the usual approaches to multiple imputation of missing values. For clarity of exposition,

we will describe some of the hypotheses which can be explored using this model in simpler terms -- for example, the primary outcome variable is described as the difference in quality of life at month 6 minus quality of life at baseline."

How sophisticated to make the analytic framework often depends on how much it improves upon the simpler alternative. No matter how the SAP is described in shorthand, its elements should be pre-specified.

What is unique to palliative care research?

The discerning reader will likely notice that the treatment to date pertains to randomized trials rather than randomized trials in palliative care. Indeed, our framework is independent of content area. However, within this framework a number of design and statistical issues can be particularly relevant for palliative care trials. Without attempting a comprehensive discussion, some of these issues include the following:

- *Respondent burden*. The choice of measures representing the constructs within Figure 1 might be constrained by limitations on respondent energy, availability and commitment.
- *Attrition*. Often, significant numbers of patients are expected to die or otherwise drop out during follow-up, which affects how longitudinally collected outcome variables should be treated within the SAP.
- *Missing data*. Respondent burden and attrition, among others, often result in a notable amount of missing data, which must be accommodated within the SAP.
- *Caregivers*. In addition to measures which can be separately analyzed among patients or their caregivers, some measures truly represent dyadic constructs, and should be conceptualized and analyzed accordingly.
- *Modest effect sizes*. Because multiple factors affect patient outcomes in addition to the intervention, effect sizes of even successful interventions tend to be modest, which from a statistical perspective suggests that sample sizes should be large, despite the acknowledged logistical difficulties. The PCRC can provide an important resource for helping to scale up a study to multiple sites.

These issues are not at all trivial, and consultation with an experienced statistician is recommended. Accessible recommendations about methodological and statistical considerations in palliative care research are included in the references (5,6).

Discussion

A fundamental premise is that collaboration between statisticians and other investigators can be enhanced if each understands how the other conceptualizes their task. As a tool for helping Intensive participants understand how statisticians think, we have focused on a figure which represents the mental map that statisticians apply to SAPs for randomized trials. In other contexts (e.g., computer science, causal inference), this figure is termed a "directed acyclic graph" and an expanding body of research describes criteria contributing toward their effectiveness (7). Our experience in the Intensives, admittedly subjective and

highly selective, is that investigators often find this figure to be illuminating, and it is our hope that it will be useful to the readers of this journal as well.

In practice, the SAP can become quite detailed, and is tailored to the study in question. Nevertheless, essentially all randomized trial SAPs are designed around figure 1, study aims, and scale of measurement of the variables. This is not intended to be a comprehensive description of SAPs, nor a tutorial in precisely how to write SAPs, but instead has the more limited purpose of describing a structure within which the outline of SAPs can be developed. Appendix 1 illustrates a table of contents for a typical SAP, and appendix 2 provides a checklist used at one of our institutions.

We have described the use of the conceptual framework underpinning the statistician's contributions toward the development of the SAP. The educational setting in which it is currently applied is PCRC-sponsored Intensives, and thus the primary target audience is palliative care researchers. By analogy with our efforts to deconstruct interactions between statisticians and other investigators (8), which are embedded within an educational program for biostatisticians, we are optimistic that this conceptual framework can also be beneficial to a statistical audience, perhaps as part of a module on the development of SAPs and effective communication around those SAPs.

Disclosure and acknowledgements

The authors have no conflicts of interest to disclose. Drs. Samsa, Colborn, and Olsen were supported by the National Institute of Nursing Research of the National Institutes of Health under award number U2CNR014637. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Appendix

Appendix

Appendix 1:

Example Table of Contents for a Statistical Analysis Plan (SAP) for a Randomized Controlled Trial

SAP Table of Contents	Notes
Background	Usually taken from the study protocol. The background and rationale are often presented in abbreviated form in the SAP with reference made to the study protocol.
Study Rationale and Scientific Hypotheses	
Aims (or Objectives)	
Primary	
Secondary	
Exploratory	
Outcomes (or Endpoints)	This section defines the variables that are to be collected for the primary analyses with traceability to the study aims. This section of the SAP will also typically define the measurement tools and scale of measurement for the variables. For example, see Zarin DA, et al. *N Engl J Med* 2011;364:852-60. The outcomes measured may be mechanistic, proximal, or distal as described in the article.
Primary	
Secondary	

SAP Table of Contents	Notes
Exploratory	
Overview of Study Design	This section will identify the study groups of interest, e.g., treatment vs. control, and how they are to be compared (parallel-arm, crossover, etc.).
Inclusion-Exclusion Criteria	Some important study variables, such as effect modifiers and covariates may be mentioned in this section where the study population is defined. For example, the article mentions primary diagnosis as an effect modifier. In that case, the diagnoses the study is interested in enrolling might appear in the inclusion criteria. Likewise, if age is an important covariate the range of ages eligible for study will likely appear here as well.
Randomization and Blinding	Assignment to study groups will typically be done within strata of variables that are considered effect modifiers or covariates. In the case of effect modifiers, the stratification is crucial to enable the desired analysis. Stratifying randomization on covariates implies one is concerned about confounding. Special plans will be detailed with respect to maintenance of the blind in trials that require this.
Sample Size Justification	The content in the article describes the minimum set of information that statisticians will require to determine the study sample size. The sample size is also directly related to the analytical plan, as described below.
Analysis Populations	Which groups of patients are relevant for which analyses? For example, a mechanistic outcome may only be relevant for patients who received at least one dose of the therapy. A proximal or distal efficacy outcome may be relevant for anyone who the investigator intended to treat with the therapy, regardless of whether the patients complied or not.
Missing Data Plan	Most trials will pre-specify the handling of missing data, e.g., due to participant dropout. The statistical methods here can get quite tricky.
Data Monitoring Plan	The plan for periodic assessment of data quality, sometimes described in a separate document but sometimes also included in the SAP. For example, in cases where independent data monitoring committees are used the SAP may specify certain periodic analyses.
Stopping Rules	Clinical trials of drugs, biologics, and medical devices often have complicated statistical safety monitoring rules. The purpose of these monitoring rules is, for example, to raise alarm if the rate of known medical complications exceeds what is expected for the population under study. This could be an indication that the experimental treatment or other protocol-related procedures is putting patients at undue risk and the trial may need to be stopped.
Data Analysis Software	It is typical to specify which software will be used to perform the analyses, and even to give example program code for the primary analyses in the SAP to illustrate the intent of the statistical analyses.
Planned Analyses with Example Tables, Listings, and Figures	
Subject Disposition	This is where you would find a CONSORT diagram.
Protocol Adherence	Lack of adherence to the treatment regimen can complicate interpretation of the analysis.
Participant Characteristics	Usually baseline characteristics of all randomized participants but perhaps also baseline characteristics of those who were analyzed, if this is a sufficiently different set of people than those who were randomized, e.g., due to extensive dropout.
Efficacy Analyses	These plans describe how the outcomes will be summarized and compared between the study groups. For example, see Zarin DA, et al. *N Engl J Med* 2011;364:852-60.
Safety Analyses	The statistical methods used to incorporate effect modifiers and covariates in the analysis are included here, e.g., multivariable hierarchical linear regression model with treatment-by-subgroup interaction term, etc. The details surrounding safety analyses will usually be more expansive in trials of drugs, biologics, or medical devices.

Appendix 2:

Statistical analysis plan checklist (some items are specific to trials and others to observational studies)

Element	Description
Administrative information	
Study information	Descriptive title that matches the protocol, with SAP either as a forerunner or subtitle
	Trial registration number, protocol version number, and/or IRB number.
	CRU/Department/Division/Center/other collaborative unit that the study falls under
Roles and responsibilities	Listing of principal investigators, clinical leads, and co-authors (if known)
	Name and affiliation of SAP author(s)
	Names, affiliations, and roles of other SAP contributors (e.g. senior statistician)
SAP information	SAP version number, with date of current version and original creation date
Project information	Project folder location
	Project goals (e.g. manuscript, abstract, presentation, etc.)
	Project deadlines (of listed goals)
	Effort estimate
Investigator agreement	Confirmation that BERD Method Core's collaborative process has been reviewed, that all statistical analyses included in an abstract or manuscript should reflect the SAP, no changes should be made to the SAP without discussing with the SAP author, all biostatisticians on the SAP are co-authors on the manuscript, and that publications resulting from the SAP must cite grant number UL1TR002553 and be submitted to PubMed Central
Signatures	Signatures of SAP author, senior statistician, and principal investigator(s)
Activity log	
SAP revisions	SAP revision history with dates
	Justification for each SAP revision
	Timing of SAP revision in relation to any interim analyses or submissions
Study overview	
Background and introduction	Synopsis of scientific background and rationale for the study
Aims and hypotheses	List of all scientific aims/objectives of the study, with specifications of primary, secondary, etc.
	List of all statistical hypotheses (corresponding to the scientific aims), with specifications of primary, secondary, etc.
Variables of interest	List of all outcome/endpoint variables, with a description of their coding/units, timing, and source, corresponding to the statistical hypotheses. If any variables are defined using ICD or CPT codes, list them out.
	List of all exposure variables, with a description of their coding/units, timing, and source, corresponding to the statistical hypotheses. If any variables are defined using ICD or CPT codes, list them out.
	List of any additional variables of interest (e.g. covariates, potential confounders, effect modifiers, etc.) in the analysis
	Location of data dictionary (or provided as an appendix)
	Report category boundaries if continuous variables are collapsed into categories, and describe any other relevant data transformations

Element	Description
Causal graph (DAG)	May be helpful to include a DAG or other graph/diagram that describes the way the variables of interest are presumed to relate
Study methods	
Study plan and design	Description of the study design (e.g. parallel group randomized trial, case-control study, cohort study, etc.)
	Study setting, location, and relevant dates (e.g. periods of enrolment, exposure, follow-up, and collection)
	Description of intervention or exposure groups, with allocation ratios, and details of any matching criteria
	Details on randomization (e.g. stratification factors) and blinding procedures
	List of eligibility and/or inclusion/exclusion criteria
	Description of screening/enrolment/recruitment processes
	Description of patient flow (e.g. CONSORT diagram)
	Description of analysis population (e.g. intention to treat, per protocol, etc.)
	Definitions of adherence/compliance, protocol deviations, loss-to-follow-up, adverse events, etc.
	Time points at which outcomes are measured
	Timing of final analyses (are all outcomes analysed collectively, or will short-term outcomes be analysed separately from long-term outcomes, etc.)
Sample size	Sample size calculation or justification (either provided in full or summarized, with link to original source)
	Description of pre-planned subgroup analyses, power for these analyses, and planned multiple comparison adjustment procedures
Interim analyses	Description of what interim analyses will be conducted at which time points, and what methods used to adjust significance levels due to the interim analysis
	Details of any guidelines (e.g. safety, futility) for stopping the study early
	Details of any changes to trial design due to interim analyses (e.g. enrolling more patients)
Data	Description of data collection/acquisition process, with contact information for team member responsible
	Description of data flow/transfer from primary data collection through to creation of final analysis dataset
	Data transfer method and date
	Folder location where datasets are stored
	Description of any additional data management, quality control, or processing undertaken
	If any data are extracted from a database, a description of the database and the query used for the extraction, and whether/how it was merged with any data from outside that database. If the study involved linkage of databases, consider use of a flow diagram to demonstrate the data linkage process, including the number of individuals with linked data at each stage.
	Description of any other data sources incorporated in the analysis
Missing data	Description of sources and magnitudes of missing data
	Description of how missing data patterns will be presented/summarized (may be helpful to have a table shell or draft CONSORT-style diagram)
	Description of contingency plans for handling missing data in analysis
Simulations	If conducting a simulation, a description of the purpose of the simulation and its design (e.g. fully factorial, partially factorial, grid search, etc.)
	Define the fixed and variable factors or parameters in the simulation, the estimands/targets of the simulation, and the performance measures to be estimated (with justifications of their relevance to the estimands/targets)

Element	Description
	Description of the tabular and graphical presentations of simulation results and their interpretation
Statistical analysis plan	
Statistical significance	Hypothesis testing framework (e.g. superiority, equivalence, non-inferiority), or description of alternative analytic framework (e.g. evaluation of a posterior in a Bayesian analysis, etc.)
	Level of significance for primary hypotheses, including a description and rationale for any multiple comparisons adjustment or Type I error control procedures
	Description of any decision-making rules based on confidence intervals, credible intervals, prediction intervals, Bayes' factors, or other alternative inferential methods
	Description of how the results of any hypothesis tests (or alternative inferential methods) will be interpreted with respect to both the statistical hypotheses and scientific aims/objectives of the study
Descriptive statistics	List of characteristics (e.g. demographic, clinical) to be summarized descriptively (e.g. "Table 1")
	Description of how these characteristics will be summarized descriptively (e.g. means/medians vs. N (%), tabular displays, graphical displays, etc.)
	Summarize follow-up time (e.g. average and total amount) and number of events
Analysis methods	For each aim/hypothesis (see items 9a/9b), a description of what analysis method will be used and how the results from this method will be reported and interpreted
	Description of any transformations, standardizations, covariate or confounder adjustments, weighting, or stratification methods to be used and why.
	For each analytic method proposed, a description of the assumptions of that method and what processes will be used to evaluate whether or not those assumptions hold
	Details of contingency plans/alternative methods to be used if the assumptions are found not to hold
	In the case of non-standard test statistics, formulas provided for the test statistic with a description of the mathematical null hypothesis, how significance is determined, and how the test statistic is interpreted
	In the case of regression models, formulas provided for the full model with a description of which parameters are to be used, how they will be interpreted, how confidence intervals will be constructed, etc.
	In the case of survey, hierarchical/nested, or clustered data, a description of what methods will be used to adjust for the data structure and why (e.g. if using a GEE, describing which correlation structure and why it was chosen, etc.)
	For non-continuous outcomes, clearly explain the effect used (e.g. risk difference, risk ratio, odds ratio, etc.), whether it is relative or absolute, and justify why that was chosen as the effect measure of interest
	Documentation of any non-standard methods used (e.g. using alternative degree of freedom calculation methods, using a non-canonical link function, etc.)
	Description of any limitations, sources of bias, internal/external validity, and other relevant discussions concerning the interpretation and generalizability of the design or methods used
Additional analysis methods	Description of any pre-planned sensitivity analyses and how they will be interpreted
	Description of pre-planned subgroup analyses, power for these analyses, and planned multiple comparison adjustment procedures
	Description of any additional post-hoc calculations or analyses (e.g. evaluating interaction/ modification effects, calculating mediation or local average treatment effects, evaluation of AUROC curves, etc.)
	If conducting any bootstrap analyses, a description of the sampling algorithm and number of iterations used
	If conducting any cross-validation procedures, a description of how the cross-validation is conducted (e.g. leave-one-out, train/validation/test, etc.)

Element	Description
Exploratory analysis	Description and justification for any pre-planned exploratory analyses and what methods will be used to conduct them
	Framework for conducting any unplanned exploratory analyses and how they will be integrated into the planned analysis
Software	List of statistical software (along with version numbers) to be used for each phase of the analysis; in the case of R or Stata, additionally list any requisite installed packages and their version numbers
Other	Description of any additional planned analyses of the data (e.g. a safety analysis looking at adverse event rates for a Data Safety Monitoring Board,
Tables and figures	
Table shells	Example tables related to any of the conducted analyses; if possible including any available preliminary data
Example figures	Example figures related to any of the conducted analyses; if possible including any available preliminary data.
References	
References	References for any non-standard statistical methods used
	References (and locations) for any relevant protocols, standard operating procedures, or other documents cited in the SAP
Additional information	
Appendices	If necessary, appendices may be included (e.g. a full data dictionary, a copy of a Case Report Form, etc.)
Addenda	Any additional analyses conducted that were not included in the SAP should be documented in an addendum, describing the purpose of the additional analysis, when it was conducted, and by whom

References

1. https://palliativecareresearch.org/ Accessed May 1, 202
2. https://grants.nih.gov/clinicaltrials_fdaaa/docs/flow_chart-act_only.pdf. Accessed May 1, 2002.
3. Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction, International Journal of Epidemiology, Volume 41, Issue 2, April 2012, Pages 514–520, 10.1093/ije/dyr218 [PubMed: 22253321]
4. Sullivan GM. “A primer on the validity of assessment instruments.” Journal of graduate medical education vol. 3,2 (2011): 119–20. doi:10.4300/JGME-D-11-00075.1 [PubMed: 22655129]
5. Higginson IJ, Evans CJ, Grande G, et al. Evaluating complex interventions in end of life care. BMC Medicine, 2013, 11;111. http://www.biomedcentral.com/1741-7015/11/111. [PubMed: 23618406]
6. Missing data in palliative and end of life care trials: Guidance on how to reduce, handle and report missing data. June 2022. https://www.mariecurie.org.uk/globalassets/media/documents/research/Missing-Guidance-Report-final-June-2022.pdf
7. Williams TC, Bach CC, Matthiesen NB, et al. Directed acyclic graphs: a tool for causal studies in paediatrics. Pediar Res 84, 487–493 (2018). 10.1038/S41390-018-0071-3.
8. Samsa GP, LeBlanc TW, Locke SC, Troy JD, Pomann GM. A model of cross-disciplinary communication for collaborative statisticians: implications for curriculum design. 2018. Journal of Curriculum and Teaching. Online August 3, 2018. DOI:10.5430.jct.v7n2p1.

Key message

A directed acyclic graph (DAG) is a useful visual tool for developing the statistical analysis plan for a randomized trial, and helps facilitate communication between statisticians and other investigators.

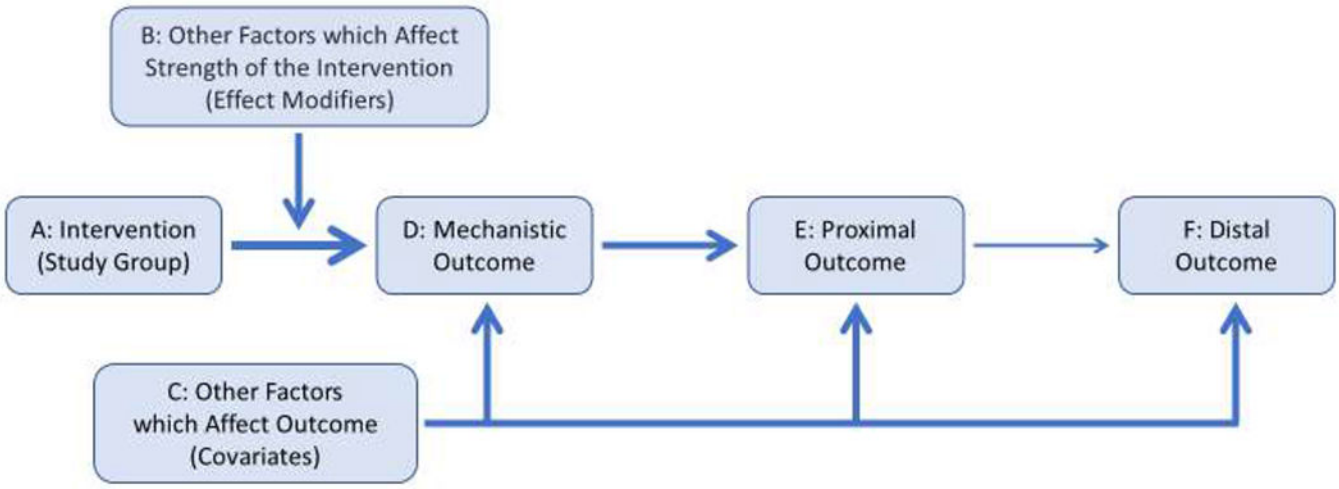

Figure 1: Conceptual model of a statistical analysis plan

Legend: The 6 boxes represent the boxes of the generic DAG. The arrows represent the flow of causality. The width of the arrows from A to D to E to F represents the strength of the relationship, which typically weakens as we proceed along the causal pathway. The impact of the covariates C on the outcomes D, E and F need not be identical.

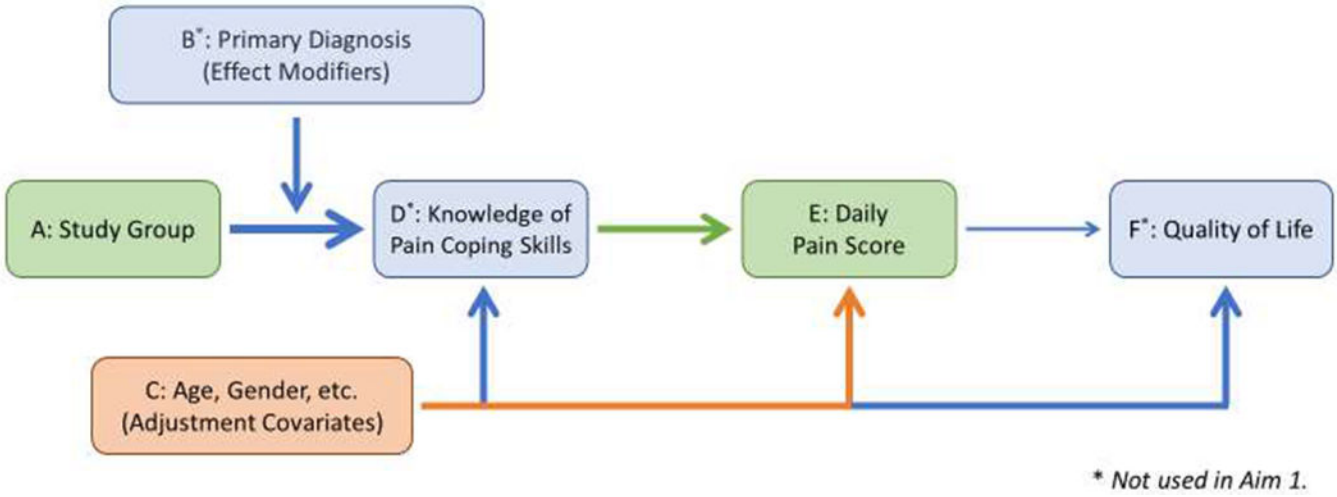

Figure 2: DAG for pain study, aim 1, unadjusted analysis

Legend: The 6 boxes represent the boxes of the generic DAG (figure 1). Blue boxes are not used in aim 1. The green boxes and arrows represent the unadjusted analysis (the green arrow extends from A to E.) The orange box and arrow would be added to create the adjusted analysis.