SAPs are for saps: A guide to the ins and outs of SAPs in iterative, observational, or publication-oriented research
Michelle Turner, ICON Clinical Research, San Francisco, CA
Lawrence Rasouliyan, ICON Clinical Research, Barcelona, Spain

ABSTRACT
Statistical analysis plans (SAPs) can be an invaluable tool. They are intended to pre-specify the objectives and endpoints of the analysis, relevant cohorts of patients to include or exclude from specific analyses, and the statistical methodology to be employed. Often the goals of the SAP are to facilitate planning, to document the work to be undertaken, and to provide a roadmap for the statistical programming team. While all noble goals, there are certain practicalities that need to be considered when working with multi-year, longitudinal, observational data and undertaking iterative analyses. This type of work is increasingly common across industries. Taking healthcare as an example, it may not be viable to draft a single SAP containing enough detail for the full set of multi-year and mid-study analyses to be independently replicated. This is because such studies cover numerous and fluid objectives, contain extremely complicated and detailed variable definitions that evolve over time, contend with messy observed or missing data, and can enroll heterogeneous populations that may require very specific handling. The involvement of multiple stakeholders adds to the challenges: who are the target audiences for the SAP? What is their level of technical proficiency? What are their expectations and needs? To whom are they being held accountable and what are their superior’s needs? What business resources are available for this task? This talk will give practical advice on how to achieve the myriad and seemingly disparate objectives in a way that makes sound business sense and satisfies both the reviewers and end-users on the study team.

INTRODUCTION
A “Statistical Analysis Plan” or “SAP” is intended to document the plan(s) for a single analysis or a series of analyses. The format and content of an SAP may vary wildly across or even within industries. That is precisely why we were inspired to write this paper and give a talk. We work in the health care industry where the term SAP is commonly associated with a very specific type of document: one that pre-specifies data analyses conducted in a clinical trial. Yet it turns out that our day to day work calls for a different type of document. This is because we conduct exploratory analysis of observational, multi-year, non-trial data. This paper describes key differences between a clinical trial SAP and one that facilitates iterative, exploratory analyses. These differences extend beyond the content of the file and are driven by the target audience for the document and the primary end user and study team needs. The type of document that we are recommending can serve as a template for data scientists, statisticians, statistical analysts, statistical programmers and quantitative study teams across various industries.

TO “SAP” OR TO “DPA”
At a high level, an SAP is intended to specify the who, what, when, where, why, and how of an analysis:

• Who is included in the analysis? Or what is your unit of analysis, if non-human.
• What data are you collecting and analyzing? What are your methods?
• When is the analysis being conducted? When are/were the data collected?
• Where will the results be disseminated? Where are/were the data collected?
• Why are you doing the analysis?
• How are you analyzing the data? How frequently are you analyzing it? How are/were the data collected and at what frequency?

Outlining these in a single document or a set of related documents will keep the study team organized and provide a roadmap for the statistical programming team. The goal should be to construct a plan that reflects the work to be conducted and is user friendly for the project team as well as for the intended target audience. Ask yourself and/or your team these questions prior to drafting your roadmap:

1. Who will be the lead author of the plan?
2. Who will approve the plan?
3. Who will use the plan?
4. What are the author, user, reviewer, and approver expectations?
5. What is their level of technical proficiency?
6. What business resources (budgetary, time, etc.) are available for the task?
7. Is the document fluid or does it need to be finalized with no further changes?
8. Do the number of and styles of analyses lend themselves to a series of documents or a single document?

There is no right or wrong answer to any of the above questions. However, there may be a more or less ideal way to construct your analysis plan depending on your answers. This is where we caution any “one size fits all” approach to the task at hand. While it is nice to have a template to work from, let’s try to engineer a document to fit our needs, rather than to fit a “standard” that may have been developed for other purposes.

If you are veering from a “standard” in your industry, it may help to use an alternative name for your plan. We frequently use the less common but still very accurate term, “Documentation of Planned Analysis” or DPA. The DPA is the term we use for any form of documenting statistical methodology. Here are some differences between an SAP for a health care clinical trial and a DPA for observational health care research.

<table>
<thead>
<tr>
<th>Questions</th>
<th>SAP</th>
<th>DPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who will be the lead author of the plan?</td>
<td>Statistician</td>
<td>Statistician, statistical analyst, statistical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>programmer, data scientist, and/or subject matter</td>
</tr>
<tr>
<td></td>
<td></td>
<td>expert</td>
</tr>
<tr>
<td>Who will approve the plan?</td>
<td>Statistician and medical director</td>
<td>Often the plan is not formally “signed off” or</td>
</tr>
<tr>
<td></td>
<td></td>
<td>approve. However, reviewers may include any of the</td>
</tr>
<tr>
<td></td>
<td></td>
<td>potential authors and/or project team members</td>
</tr>
<tr>
<td>Who will use the plan?</td>
<td>Statistical and programming team</td>
<td>Statistical and programming team, subject</td>
</tr>
<tr>
<td></td>
<td></td>
<td>matter experts</td>
</tr>
<tr>
<td>What are the author, user, reviewer, and approver expectations?</td>
<td>Formal, fully pre-specified analyses; the</td>
<td>Reader-friendly document for a combination of a</td>
</tr>
<tr>
<td></td>
<td>documents are often over 100 pages in length</td>
<td>technical and non-technical audience outlining the</td>
</tr>
<tr>
<td></td>
<td>with additional mock table and figure shells</td>
<td>high level approach to the analyses at hand; often</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5-15 pages in length with or without additional</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mock table and figure shells and sometimes the mock</td>
</tr>
<tr>
<td></td>
<td></td>
<td>table and figure shells serve as the DPA in and of</td>
</tr>
<tr>
<td></td>
<td></td>
<td>themselves without corresponding text</td>
</tr>
<tr>
<td>What is their level of technical proficiency?</td>
<td>High</td>
<td>Variable</td>
</tr>
<tr>
<td>What business resources (budgetary, time, etc.) are available for the task?</td>
<td>High</td>
<td>Variable</td>
</tr>
<tr>
<td>Is the document fluid or does it need to be finalized with no further changes?</td>
<td>Finalized and approved</td>
<td>Fluid and iterative</td>
</tr>
<tr>
<td>Do the number and styles of analyses lend themselves to a series of documents or a single document?</td>
<td>There may be a large number of analyses; however, the expectation is that the majority of them will be fully pre-specified in a single SAP</td>
<td>Series; heavily weighted towards exploratory analysis</td>
</tr>
</tbody>
</table>

Table 1. SAP vs. DPA

If you are in an industry such as marketing or banking, consider whether your responses fit better with those presented in the SAP vs. DPA column, accounting for the changes in staffing required for your projects.

There is certainly no requirement to rename your document from SAP to DPA or any other alternative. It is simply an option to consider if your document is competing with another “SAP” standard in your industry that may not meet your needs.
THE FRIENDLY DPA

The DPA can take a variety of forms. It truly could be as simple as an email containing a high level summary of the planned analysis. Alternatively, it could involve anything up to a textual description with our without corresponding table shells. If the latter, the DPA may contain these sections, which are also common to an SAP. The difference is that the DPA may or may not provide the level of detail expected in a more formal SAP; this would be left to the study team to decide.

INTRODUCTION

The introduction includes a brief summary of the project and study design. The introduction also clearly states the purpose of the current analysis.

ANALYSIS OBJECTIVES

The analysis objectives section indicates all primary, secondary and other objectives as described in any study background documentation, e.g., within a study protocol for health care research.

ANALYSIS POPULATION

The analysis population section describes the subject population being investigated, the study’s target or current sample size and the analytic cohort(s).

DATA COLLECTION TOOLS

The data collection forms section summarizes all measurements collected in the project and describes all forms used to collect data.

ASSESSMENT SCHEDULE

The assessment schedule section includes information on the timing of the assessments and who completed which assessments (e.g., physician, clinician, patient, caregiver etc., using healthcare as an example).

DATA DICTIONARY

The data dictionary section describes the origin and the nature of the dataset(s) being used and possibly specifications for variables.

ANALYSIS VARIABLE DEFINITIONS

The analysis variable definitions section describes the source and meaning of analysis variables and identifies which variables are investigated as background characteristics and as outcomes.

STATISTICAL METHODS

The statistical methods section identifies which statistical techniques are employed in the examination of the project outcomes. The name of the statistical tests, the alpha level used to determine statistical significance and the covariates to be used in modeling should be clearly identified. In addition, any procedures used to adjust for multiple comparisons are noted. Exploratory analyses may also be detailed in this section.

METHODS FOR HANDLING MISSING DATA

This section explicitly states how missing data is addressed in the analysis.

REFERENCES

This section includes bibliographic references for any non-standard statistical methodology or published references used within the plan.

I “NEED” AN SAP

The manageable size of the DPA makes it possible to generate one for each topic under study on your project. Taken together, the multiple DPAs can serve as modules of a larger SAP, if you encounter a situation where a larger scale or more formal SAP is “required.” As first step, we recommend that you try to educate your project team about the style of document that will truly facilitate the work at hand. However, if faced with resistance to a DPA or demands for an SAP, combining a set of DPAs has satisfied some project teams. Keep in mind that DPAs can be written as
needed during the course of a project prior to conducting an analysis and some may not have been drafted or even thought up in time to meet an arbitrary SAP deadline.

CONCLUSION

SAPs are helpful tools but can be over-engineered for certain types of projects. Work with your project team to design the type of planning tool that will best meet your overall needs. Whether that is termed SAP, DPA, or an alternative name is up to you and your colleagues. However, the alternative naming convention may be a great first step in educating team members about the difference in your tool vs. a “standard” SAP in your industry.

ACKNOWLEDGMENTS

The authors would like to acknowledge the ICON Clinical Research Medical Affairs Statistical Analysis group and Dave Miller for their tireless work to advance the field of exploratory analysis of observational data in health care research.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Michelle Turner
Enterprise: ICON Clinical Research
Address: 456 Montgomery Street Suite 2200
City, State ZIP: San Francisco, CA
Work Phone: +1 215.616.5026
Fax: +1.415.856.0840
E-mail: michelle.turner@icconplc.com
Web: icconplc.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.