Article / Concept Paper

A method to encourage minimum reporting guideline uptake for data analysis in metabolomics

Elizabeth C. Considine [1*], Reza M. Salek [2]

[1]The Irish Centre for Fetal and Neonatal Translational Research (INFANT), Department of Obstetrics and Gynaecology, University College Cork, Cork, Ireland

[2] The International Agency for Research on Cancer (IARC), 150 Cours Albert Thomas, 69372 Lyon CEDEX 08, France.

* lizconsidine@gmail.com

Abstract: Introduction

Despite the proposal of minimum reporting guidelines for metabolomics over a decade ago, reporting on the data analysis step in metabolomics has been shown to be unclear and incomplete with major omissions and lack of logical flow rendering the data analysis' workflows in these studies impossible to follow and therefore replicate or even imitate. Here we propose possible reasons why the original guidelines have had poor adherence and present an approach to improve their uptake. We present in this paper an R markdown reporting template file that guides the production of text and generates workflow diagrams based on user input. This R Markdown template contains, as an example in this instance, a set of minimum information requirements *specifically for the data pre-treatment and data analysis section* of biomarker discovery metabolomics studies, (gleaned directly from the original proposed guidelines by Goodacre at al). These minimum requirements are presented in the format of a questionnaire checklist in an R markdown template file. The R Markdown reporting template proposed here can be presented as a starting point to encourage the data analysis section of a metabolomics manuscript to have a more logical and stepwise presentation and to contain enough information to be understandable and reusable. The idea is that these guidelines would open to user feedback, modification and updating by the metabolomics community via GitHub.

Keywords: reproducibility; minimum guidelines; reporting; data analysis; reporting

.

1. Introduction

Metabolomics data mining describes the application of strategic data analysis methods incorporating artificial intelligence, machine learning, statistics and database operations to extract meaningful and useful information from high-dimensional and high-volume

 doi:10.20944/preprints201901.0246.v1

metabolomics datasets. This is a complex, time-consuming process including many steps with many possible options at each step. Metabolomics analysis is complicated by the metabolome's vast complexity and dynamics. Metabolites are present in a wide range of concentrations, from low abundance signaling molecules to high abundance compounds of central metabolism. Metabolites are also subject to temporal and spatial variability and are affected by environmental influences such as circadian and diet fluctuations to name a few. Further complexity is encountered in metabolomics investigations due to the fact that data analysis is often based on open source tools, each having their own parameter dependencies, also metabolomics datasets typically contain missing values and the handling of these can greatly influence the result of downstream analysis [1, 2]. For detailed discussion and reviews of the data analysis step in metabolomics and its complexities the reader is referred to the following publications [3-7].

Despite the obvious complexity and importance of the data mining step in the overall pipeline of any metabolomics study, this section of the workflow is often given scant attention in the write up of scientific research articles. Metabolomics data analysis sections have been found to be plagued by inconsistent reporting specifically with regards to structure, details reported, and performance metrics used [3].

No standard method for how to analyse metabolomics data exists and therefore data analysis is in constant evolution with new methods frequently being proposed in the literature. To discover the best methods, to build upon existing approaches and to conduct meta-analysis, the data analysis write up in metabolomics studies needs to be understandable and imitable, at a minimum. Furthermore, for those new to metabolomics data analysis, the starting point to construct a data analysis plan would most likely involve examining previously published research in the same field with a view to reusing or adapting the various approaches. For these purposes the current standard of reporting of the data analysis sections in metabolomics studies manuscripts is woefully insufficient. The immediate improvement of reporting of the data analysis step is therefore vital to advance understanding and to promote reuses of data analysis protocols and eventually move closer to the ideal of reproducibility.

Minimum reporting guidelines for data analysis in metabolomics [8] were first published in 2007. These comprehensive guidelines cover: 1: Design of Experiment (sample collection/matching, and data acquisition scheduling of samples); 2: Data Collection 3: Data pre-processing (data cleaning, outlier detection, row/column scaling, or other transformations; Definition and parameterization of subsequent visualizations) 4: Data pre-treatment (row wise and column wise operations such as normalisation, scaling, centering and transformation to make data more amenable to statistical analysis); and finally, 4: Actual data analysis which includes algorithm selection, univariate analysis and multivariate analysis. These reporting guidelines were published as an article but were not subsequently published in the format of a guidelines checklist with an explanation and elaboration document nor were they formally disseminated. Results of our recent review on reporting of the actual data analysis step in metabolomics indicate that these original reporting guidelines have had very poor take up, at least for the data pre-treatment and actual analysis section of metabolomics studies [2]. For example 89% (23 out of a total of 25) of

 doi:10.20944/preprints201901.0246.v1

studies reviewed from the years 2008 to 2014 did not mention the proportion of missing values nor how the missing values were dealt with. Less than half the studies reviewed reported on any kind of quality control procedure and less than half had any mention of outlier detection and/or removal.

Reasons why those original reporting guidelines for metabolomics have had such poor take up may include the fact that they never progressed from the "proposed" stage to being formally published as practical guidelines along with a detailed "explanation and elaboration document" and they are not required by most journals for publishing a metabolomics manuscript. Also their comprehensiveness may have inhibited their uptake as there may have been an overwhelming amount of information to work through.

Strategies to increase the uptake and impact of guidelines can be adopted by their authors such as publishing the guidelines in multiple journals to ensure quicker and wider dissemination; also authors can approach journals and ask them to include the guidelines in their "Instructions to authors" section and publish commentaries to endorse them [9]. Official society and community wide recommendations can also positively influence the uptake of guidelines. There are a number of recommendations on a dedicated page of the EQUATOR (Enhancing the QUAlity and Transparency Of health Research) Network website [10] on how to effectively disseminate your reporting guideline: http://www.equator-network.org/toolkits/developing-a-reporting-guideline/disseminating-your-reporting-guideline/.) This would of course be in addition to previous recommendations of supplementary audit, open code and script sharing [11].

In recognition of the complexity of data analysis in metabolomics we propose that distinct reporting guidelines be drawn up for separate sections of the data analysis pipeline (design of experiment, data reduction and deconvolution, data pre-processing, data annotation and identification, data pre-treatment and data analysis) as their various steps are often carried out at different times, (sometimes years pass between different steps), by different individuals or groups of various skill sets, or even often outsourced to different locations. Adherence to guidelines is more likely to be achieved by discretising each section of the pipeline into succinct guideline sets (modules) which can be adopted by the relevant analyst(s) *at the time of manuscript writing* and incorporated into the report.

In other established omics reporting guidelines', namely MIAME [12] for microarrays and MIAPE [13] for proteomics, instructions do not exist for the data analysis part of the pipeline. The Equator (Enhancing the QUAlity and Transparency Of health Research) Network [10] is an international initiative aimed at promoting transparent and accurate reporting of health research studies to enhance the value and reliability of medical research literature. The Equator Network does not contain any guidelines for reporting of multivariate data analysis/high dimensional data analysis/omics data analysis/supervised data analysis. The biosharing website MIBBI (Minimum Information for Biological and Biomedical Investigations) [14] does not contain any standards for data analysis reporting but it does reference a standard called CIMR- Core Information for Metabolomics Reporting (CIMR) [15] which refers back to the original proposed guidelines [8]. However since these initial

 doi:10.20944/preprints201901.0246.v1

proposed guidelines in 2007 no further work has been published on the development or dissemination of metabolomics data analysis reporting guidelines.

Since the Metabolomics Standards Initiative (MSI) [16] significant developments in data reporting standards in metabolomics have been made through many initiatives including COSMOS [17], MetaboLights [18] and FAIR [19] which endeavour to ensure consistency of metadata between datasets, and facilitate data reuse and data merger across studies [20]. However with regards to reporting of the data *analysis* of metabolomics studies, since the original guidelines [8] there have been no further advancements.

There has recently been a proliferation of reporting guidelines in biomedical research [21], there are currently 407 reporting guidelines on the Equator Network [10] many containing extensions and different versions. However despite this, compliance levels with these guidelines have been disappointing [22]. It has been noted that the "main problem" preventing the uptake of guidelines is that they are used too late in the research process, when it is too late to discover important things that have been missed or could have been done better [23].

1.2 Data analysis reporting using R Markdown

With this information in mind we suggest that adherence to guidelines could be facilitated if reporting guideline modules were contained in authoring tools such as the one we present in this study using R Markdown. R Markdown is a free and open source authoring framework [24]. R Markdown documents are fully reproducible and support dozens of static and dynamic output formats. A single R Markdown file can be used to both save and execute code and generate high quality reports that can be shared with an audience. R Markdown starts with a plain text file that is edited by the user that has the extension .Rmd. This plain text file then generates a new file that contains user selected text, code, and results from the .Rmd file. The new file can be a finished web page, PDF, MS Word document, slideshow, notebook, handout, book, dashboard, package vignette or other format.

As efforts towards computational reproducibility continue, an authoring tool in R markdown such as the one we present has the advantage that it can simultaneously achieve the aims of both ensuring reporting standards are adhered to while also having the potential to embed the code to perform the analysis. Such an authoring tool could therefore ultimately provide an uninterrupted and transparent workflow from the initial stage of data in to the final output of an analysis report. Our example of an authoring tool in this instance is solely focused on reporting of the statistical data analysis step of the pipeline, incorporating the data pre-treatment step and the actual data analysis step.

 doi:10.20944/preprints201901.0246.v1

1.3 Objectives

1. To present a set of previously proposed minimum reporting guidelines in the form of a checklist specifically for the data analysis step of metabolomics biomarker discovery studies. There are typically 4 phases to this data analysis pipeline, although aside from pre-treatment the other steps are not essential but are commonly used
    - data pre-treatment
    - univariate data analysis to identify significant features that differ between groups
    - unsupervised data analysis to discover correlated features or identify hidden subgroups or to visualise separation and identify outliers
    - Supervised data analysis, specifically for developing prediction models and/or biomarker identification.
    - Receiver Operating Curve (ROC) analysis

2. To provide an authoring tool to promote standardised comprehensive reporting on data analysis which will also generate workflow diagrams.

2. Methods

2.1: Checklist of minimum information for reporting data analysis in metabolomics

The development of a reporting guideline checklist, specifically for the data pre-treatment and data analysis sections of the metabolomics pipeline, with a view to general applicability to other omic domains.

This checklist was compiled based on the complete information required to construct a workflow diagram and to repeat the analysis using the reader's own version of code. The main areas of omissions which lead to confusion and ambiguity when conducting our review [3] helped to inform this checklist.

Existing guidelines which helped to shape this guideline list included the TRIPOD [25] statement, GRIPS [26] statement and REMARK guidelines [27]. Of course, the main document informing these guidelines is the original proposed guidelines for data analysis reporting in metabolomics by Goodacre at al [8] which covers the reporting of every part of the data analysis of a metabolomics experiment from design of experiment through pre-processing to data pre-treatment and final analysis.

2.2: An authoring tool for reporting statistical analysis of predictive omics

The development of an authoring tool using R Markdown. This reporting guideline checklist is presented as a questionnaire in an R markdown file that guides the production of text and workflow diagrams based on user input. These reporting guidelines are intended to form a neutral and malleable framework and have general applicability and interoperability across various omics domains. These can be extended as needed by different domains or studies but

 doi:10.20944/preprints201901.0246.v1

200 would represent a minimum set of information to be supplied whenever predictive data
201 mining in metabolomics is carried out. We purposely do not develop a “user friendly” web
202 interface as the goal is for users to operate within the R Markdown environment.

203

 doi:10.20944/preprints201901.0246.v1

3. Results

3.1: Minimum Information about a Data AnalysiS (MIDAS) checklist (Guidelines checklist specifically for the data analysis step)

Guidelines of two types generally exist: guidelines for reporting and guidelines for protocols. Since the area of data mining for metabolomics is still nascent we suggest that guidelines or limitations on methodology at this point would be premature as the optimal methods for extracting clinically useful biomarkers has clearly not been established. Therefore our guidelines pertain only to reporting.

MIDAS Guidelines Checklist

Pre-treatment

- What are the dimensions of the dataset entering this phase of analysis?
- What percentage of the data is missing values?
- Is imputation (I) performed?
- If yes describe method
- Is normalisation (N) performed?
- If yes describe method
- Is transformation (T) performed?
- If yes describe method
- Is scaling (S) performed?
- If yes describe method
- Is filtering (F) applied to the dataset at this point?
- If yes describe method
- Is a QC (QC) method employed on the dataset?
- Please describe
- Outline the order of the pre-treatment steps performed on the dataset
    - E.g. I-> T-> S->N->F->QC
- Have the dimensions of the dataset changed from the outset of pre-treatment to the end of pre-treatment?
- Provide details on the package or program used for this phase of the analysis
- If in house code is used provide it or a link to it and also the language the code is written in.

Univariate analysis

- What are the dimensions of the dataset entering this phase of analysis?
- Is univariate testing performed?
- If yes describe method
- Is a multiple testing correction employed with this method?
- If yes describe method

 doi:10.20944/preprints201901.0246.v1

- Are other methods of univariate testing performed?
- If yes describe methods
- Are multiple testing correction employed with these methods?
- If yes describe method
- Please report p-values and adjusted p-values.
- Please report test statistics and confidence intervals.
- Have the dimensions of the dataset changed from the outset of univariate analysis to the end of univariate analysis? If yes provide the dimensions of the dataset at the end of univariate analysis and make clear how the dimensions have changed
- Provide details on the package or program used for this phase of the analysis
- If in house code is used provide it or a link to it and also the language the code is written in.
- If a list of potential biomarkers is produced at this point please state this explicitly

Unsupervised analysis

- What are the dimensions of the dataset entering this phase of the analysis?
- Are unsupervised methods employed for visualisation and/ or data reduction and/or correlation analysis?
- If yes describe the algorithm used.
- Is outlier detection and removal addressed at this point? If yes please describe and specify the outliers removed.
- Are unsupervised analysis methods used for clustering?
- If yes describe and provide distance metric.
- Have the dimensions of the dataset changed? If yes how and why?
- Provide the dimensions of the dataset at the end of unsupervised analysis.
- Provide details on the package or program used for this phase of the analysis
- If in house code is used provide it or a link to it and also the language the code is written in.
- If a list of potential biomarkers is produced at this point please state this explicitly

Supervised analysis

- What are the dimensions of the dataset at this point?
- Are supervised methods employed?
- If yes describe the supervised analysis described fully enough to allow imitation of the exact procedure. This would require reporting all the following information: all parameters; details of how data is split; details of how internal validation is conducted (i.e. Cross Validation); details of how meta-parameter optimization is performed; details about the chosen metric for assessing the predictive ability of the model and finally the overall description of the workflow.
- Is more than one supervised method employed?
- If yes describe the implementation of the other algorithm(s) fully enough to allow imitation of the exact procedure. This would require reporting all the following

 doi:10.20944/preprints201901.0246.v1

information: all parameters; details of how data is split; details of how internal validation is conducted (i.e. Cross Validation); details of how meta-parameter optimization is performed; details about the chosen metric for assessing the predictive ability of the model and finally the overall description of the workflow.

- Is external validation employed?
- If yes describe the source of external data. Is the data from the same location/ lab/timeline or a hold-out set from the original data?
- Provide a confusion matrix of results
- Provide results as average of N leave-multiple-out and external predictions
- Are potential biomarkers identified? If yes, list them.
- Have the dimensions of the dataset changed? If yes how and why?
- Provide the dimensions of the dataset at the end of supervised analysis.
- Provide details on the package or program used for this phase of the analysis
- If in house code is used provide it or a link to it and also the language the code is written in.
- If a list of potential biomarkers is produced at this point please state this explicitly.

Receiver Operating Curve (ROC) Analysis

- Is ROC analysis performed on the identified putative biomarkers?
- If yes please report on AUC, sensitivity and specificity
- Provide details on the package or program used for this phase of the analysis
- If in house code is used provide it or a link to it and also the language the code is written in.

For data analysis methods currently not covered here (for example, cluster analysis and other classification and feature selection methods) similar templates can be generated with their required parameter reporting and add it to the existing templates via adding branches to the GitHub repository. We actively invite participation from metabolomics community users to become involved in this collaborative venture.

3.2 Link to GitHub repository containing markdown template.

https://github.com/MSI-Metabolomics-Standards-Initiative/MIDAS

This R Markdown file containing our authoring template is contained in a GitHub repository. Having started as a code developer's collaborative platform, GitHub [28] is now the largest online storage space of collaborative works that exists in the world which makes it the ideal platform to share this R Markdown template file.

The beauty of R Markdown as stated above is that it can embed and execute code and this code can then be hidden or displayed in the final document. Even in the most basic of report writing templates such as the one presented here this is very useful as we can use

 doi:10.20944/preprints201901.0246.v1

DiagrammeR [29], a flexible and powerful R package for generating graph and flowchart diagrams. It is necessary to state that because these diagrams depend on HTML and JavaScript for rendering they can only be used in HTML based output formats (they don't work in PDFs or MS Word documents). We can get around this by can saving our workflow diagrams within RStudio as an image (JPEG, PNG, BMP, etc.) and inserting them into the final PDF or Word document version of the report.

The authoring tool presented here does not dictate or control the report produced by the data analyst and the report produced can continue to be edited after the final report is generated. It merely serves as a guide for the writer to construct their analysis report by reminding them of points to include. To our knowledge this is the first instance that such a markdown tool has been proposed to aid and formalize reporting guideline uptake.

This authoring template is currently available on the Metabolomics Standards Initiative GitHub Repository and is open and welcoming to extension, modification and improvement from the metabolomics community and as such is considered a work in progress and a dynamic tool.

4. Discussion

Currently a copy-and paste paradigm in which results are generated in a statistical package and copied and pasted to a report document dominates data analysis reporting. Eventually a complete move away from this antiquated copy-and-paste system which is error prone and enables selective reporting is needed in order to fulfil the requirements of reproducible research. However, in the meantime, in this instance, our version of the MIDAS reporting template allows users to manually input results that they have obtained using other software, whilst also having the potential to contain fully executable code. This is so as to not exclude users other than R users from benefiting from using this reporting template as an authoring tool and to encourage the first steps towards reproducible research.

Employing R Markdown to help uptake of minimum reporting standards goes further than providing a checklist. By encouraging scientists to consider reporting standards at the time of manuscript writing it actively helps authors adhere to guidelines. These guidelines should be viewed by authors as helpful to the writing process as opposed to a "yet another hurdle along the journey to publication" [21]. Also, requiring that parameter choices are revealed in reporting, even default ones from online data analysis tools, will encourage non experts to deliberate on the choices they are making regarding the appropriate algorithms and parameters for the type of analysis that they are doing, which will further the advancement of the field.

We purposely do not provide a web end user interface for accessing this R Markdown template as we believe that data analysts need to become comfortable in environments such as R Markdown if the production of reproducible research papers is to become a reality.

 doi:10.20944/preprints201901.0246.v1

Furthermore we feel that anyone who is capable of data analysis is more than capable of using R Markdown without the need of a "user friendly" web interface.

A modularised system of authoring tools would encourage the uptake of guidelines on two fronts: The modularised facet would ensure that each researcher at each stage of the pipeline would be responsible for following the appropriate guidelines pertaining to *their own* area of expertise. The authoring tool part would ensure that guidelines are addressed *at the time of writing* up that particular section, as opposed to a set of rules to consider for application to the manuscript just before journal submission when the entire article has already been written.

Ideally, it is envisaged that such a modularised system of reporting guideline authoring tools would evolve in the omics community whereby these modules could be concatenated as needed depending on the experiment, each module corresponding to a stage in the workflow pipeline, with all modules being extensible and modifiable according to the domain and experiment in question.

In computational biology extensibility and modifiability of tools are essential so that new methods can develop and build on the old ones without repetition or reinventing the wheel. For this reason this R Markdown file is not presented here as an end result but is proposed as a starting point to encourage the data analysis section of metabolomics papers to have a more logical and stepwise presentation and to contain enough information to be understandable. So, even though this R Markdown file only attends to the authoring and not the analysis of metabolomics data we hope that it will coax data analysts into the environment of R Markdown (and GitHub) and therefore be a nudge along the road towards readable, and ultimately, reproducible, metabolomics research.

Here are the instructions to use this R Markdown authoring template.

1. Go to the **GitHub** repository: https://github.com/MSI-Metabolomics-Standards-Initiative/MIDAS
2. Click the "clone or download" button on the right hand side of the page and download the folder as a zip file.
3. Download latest version of **R Studio** if you do not have it.
4. Open the folder and open the *MIDAS.rmd* file in **R studio**.
5. Start editing and writing the report of your data analysis guided by the questions in green directly inside the *MIDAS.rmd* file.
6. After the pre-treatment, univariate analysis, unsupervised analysis and supervised analysis sections have been completed the next section is to produce workflow diagrams.
7. Follow the instructions in green to produce a workflow diagram of pre-treatment steps.
8. Click on the knit button and knit to HTML to see how the generated report looks.
9. Knit to PDF or Word to render the report to a pdf or word document as you wish.
10. PDF and Word reports will not contain the diagrams so these need to be saved in the viewer pane as an image (JPG /BMP etc) to your local folder

11. Insert the workflow diagrams into your Word or PDF report that you have saved to your local folder.
12. Render the document to HTML and workflow diagrams will be included anyway.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "conceptualization, E.C; methodology, E.C.; software, E.C.; validation, RS; writing—original draft preparation, EC an.; writing—review and editing, R.S.",

Funding: This research was funded by Science Foundation Ireland

References

1. Gromski, P.S., et al., *Influence of missing values substitutes on multivariate analysis of metabolomics data.* Metabolites, 2014. **4**(2): p. 433-452.
2. van den Berg, R.A., et al., *Centering, scaling, and transformations: improving the biological information content of metabolomics data.* BMC Genomics, 2006. **7**: p. 142-142.
3. Considine, E.C., et al., *Critical review of reporting of the data analysis step in metabolomics.* Metabolomics, 2018. **14**(1): p. 7.
4. Cambiaghi, A., M. Ferrario, and M. Masseroli, *Analysis of metabolomic data: tools, current strategies and future challenges for omics data integration.* Briefings in bioinformatics, 2017. **18**(3): p. 498-510.
5. Bartel, J., J. Krumsiek, and F.J. Theis, *Statistical methods for the analysis of high-throughput metabolomics data.* Computational and structural biotechnology journal, 2013. **4**(5): p. e201301009.
6. Ren, S., et al., *Computational and statistical analysis of metabolomics data.* Metabolomics, 2015. **11**(6): p. 1492-1513.
7. Tugizimana, F., et al., *A conversation on data mining strategies in LC-MS untargeted metabolomics: Pre-processing and pre-treatment steps.* Metabolites, 2016. **6**(4): p. 40.
8. Goodacre, R., et al., *Proposed minimum reporting standards for data analysis in metabolomics.* Metabolomics, 2007. **3**(3): p. 231-241.
9. Simera, I., et al., *Guidelines for Reporting Health Research: The EQUATOR Network's Survey of Guideline Authors.* PLOS Medicine, 2008. **5**(6): p. e139.
10. *The Equator Network*.
11. Meier, R., et al., *Bioinformatics can boost metabolomics research.* Journal of biotechnology, 2017. **261**: p. 137-141.
12. Brazma, A., et al., *Minimum information about a microarray experiment (MIAME)-toward standards for microarray data.* Nat Genet, 2001. **29**(4): p. 365-71.
13. Taylor, C.F., et al., *The minimum information about a proteomics experiment (MIAPE).* Nat Biotech, 2007. **25**(8): p. 887-893.
14. *The Biosharing Website*. Available from: https://biosharing.org/standards/?selected_facets=isMIBBI:true.
15. *CIMR-Core Information for Metabolomics Reporting*.
16. Fiehn, O., et al., *The metabolomics standards initiative (MSI).* Metabolomics, 2007. **3**(3): p. 175-178.

17. Salek, R.M., et al., *COordination of Standards in MetabOlomicS (COSMOS): facilitating integrated metabolomics data access.* Metabolomics, 2015. **11**(6): p. 1587-1597.
18. Steinbeck, C., et al., *MetaboLights: towards a new COSMOS of metabolomics data management.* Metabolomics, 2012. **8**(5): p. 757-760.
19. Wilkinson, M.D., et al., *The FAIR Guiding Principles for scientific data management and stewardship.* 2016. **3**: p. 160018.
20. Spicer, R.A., R. Salek, and C. Steinbeck, *Compliance with minimum information guidelines in public metabolomics repositories.* 2017. **4**: p. 170137.
21. The, P.M.E., *From Checklists to Tools: Lowering the Barrier to Better Research Reporting.* PLoS Medicine, 2015. **12**(11): p. e1001910.
22. Glasziou, P., et al., *Reducing waste from incomplete or unusable reports of biomedical research.* Lancet, 2014. **383**(9913): p. 267-76.
23. Marusic, A., *A tool to make reporting checklists work.* BMC Medicine, 2015. **13**(1): p. 243.
24. Baumer, B. and D. Udwin, *R Markdown.* Wiley Interdisciplinary Reviews: Computational Statistics, 2015. **7**(3): p. 167-177.
25. Moons, K.G., et al., *Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration.* Annals of Internal Medicine, 2015. **162**(1): p. W1-73.
26. Janssens, A.C.J.W., et al., *Strengthening the Reporting of Genetic Risk Prediction Studies: The GRIPS Statement.* PLOS Medicine, 2011. **8**(3): p. e1000420.
27. McShane, L.M., et al., *REporting recommendations for tumour MARKer prognostic studies (REMARK).* British Journal of Cancer, 2005. **93**(4): p. 387-391.
28. *GitHub*. Available from: https://github.com/.
29. Sveidqvist, K., et al., *DiagrammeR: Create Graph Diagrams and Flowcharts Using R.* R package version 0.9. 0. URL: https://CRAN. R-project. org/package= DiagrammeR, 2017.