# Data Synthesis

To study the prevalence of FASD, behaviours that predict FASD, prevention efforts which have been made   
and normative questions which have been raised for FASD and analogous conditions.

Table of Contents

[Data Synthesis 1](#_Toc413070912)

[Screener Instructions 2](#_Toc413070913)

[Preparation and collaboration 2](#_Toc413070914)

[Preparation 2](#_Toc413070915)

[Collaboration 2](#_Toc413070916)

[First screening sweep: titles 3](#_Toc413070917)

[Screening an entry 3](#_Toc413070918)

[Second screening sweep: titles and abstracts 4](#_Toc413070919)

[Preparation 4](#_Toc413070920)

[Acquiring the abstracts 4](#_Toc413070921)

[Screening an entry 5](#_Toc413070922)

[Third screening sweep: full texts 6](#_Toc413070923)

[Preparation 6](#_Toc413070924)

[Extractor Instructions 8](#_Toc413070925)

[Preparation and collaboration 8](#_Toc413070926)

[Preparation 8](#_Toc413070927)

[Collaboration 8](#_Toc413070928)

[Example: Method & results extraction: general article information 9](#_Toc413070929)

[Entering the extracted information 12](#_Toc413070930)

[Bibtex Key 12](#_Toc413070931)

[Sample information and subsamples 12](#_Toc413070932)

[Methodological information 13](#_Toc413070933)

[Method & results extraction: results and measurements 13](#_Toc413070934)

[Associations 13](#_Toc413070935)

[Univariate results 15](#_Toc413070936)

[Example 1 (univariate results extracted from Druschel2007a): 16](#_Toc413070937)

[Measurements and manipulations 17](#_Toc413070938)

[Example 3 (measurements extracted from Druschel2007a): 20](#_Toc413070939)

[Summary 21](#_Toc413070940)

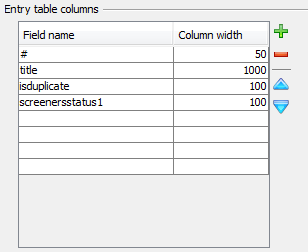
# Screener Instructions

To study the prevalence of FASD, behaviours that predict FASD, prevention efforts which have been made   
and normative questions which have been raised for FASD and analogous conditions.

As a screener, you will use the portable version of JabRef, an open source reference management program, to screen a number of articles that are eligible to be included in a data synthesis (a qualitative review and meta-analysis) of studies into the prevalence of FASD, behaviours that predict FASD, prevention efforts which have been made and normative questions which have been raised for FASD and analogous conditions. The screening comprises three sweeps. This document explains what your tasks as a screener are in each sweep.

## Preparation and collaboration

### Preparation

Before you can start screening, you need to make sure that you have the correct version of JabRef and that it is configured correctly. You can download JabRef <http://jabref.sourceforge.net/> . Also, download the 7-zip application (this is an open source archiving application that you can download at <http://7-zip.org/> - if you don’t want to install any software, you can get a portable version at <http://portableapps.com/apps/utilities/7-zip_portable>). You will receive a Screener Code from the project coordinator. Make sure you download the version corresponding to your Screener Code (E for Eline or S for Sylvia). When opening the windows batch file in your folder, you will see that JabRef opens with an empty screen. Drag the file .bib to your opened JabRef and now you see a full screen in front of you. Once you started this program, open the Options menu and select Preferences. In the Entry editor section, verify that “Show BibTex source panel” is unchecked. Then, go to the Entry Table Columns section and verify that the selected columns (and their widths, roughly) correspond to the picture on the right. If not, copy this example.

Then, in the Options menu select Set up general fields, and verify that the only text in the textbox is (substitute your screener code, E or S, for the X):

Screening:title;isduplicate;screenerxstatus1  
Screening2:title;screenerxincl;screenerxpreval;screenerxbehav;screenerxpreven;screenerxethic;screenerxmeth;screenerxcat

If the textbox contains anything else, replace the contents with this text. Then, contact the project leader. He will then send you the library with references to screen. You can open this library in JabRef by pressing CTRL-O or opening the File menu and choosing Open Database, and then browsing to the location where you stored the library, and selecting it. Every time you restart JabRef, before opening the library, verify that these three JabRef settings (BibTex source panel unchecked; correct columns in Entry Table; correct General Fields in Entry Editor) are as described here, and if not, reset them. Also, start by organizing the references alphabetically by article title by clicking the heading of the Title column of the entry table until it shows a triangle next to the word Title, like this:

By sorting the entries like this, it is easier to spot duplicate entries.

During the screening, make sure that every session, you keep track of the number of entries you screened and how long you took! This is necessary to determine the costs of screening, and how quickly people become faster as screeners, for future projects.

### When you experience problems, follow this:

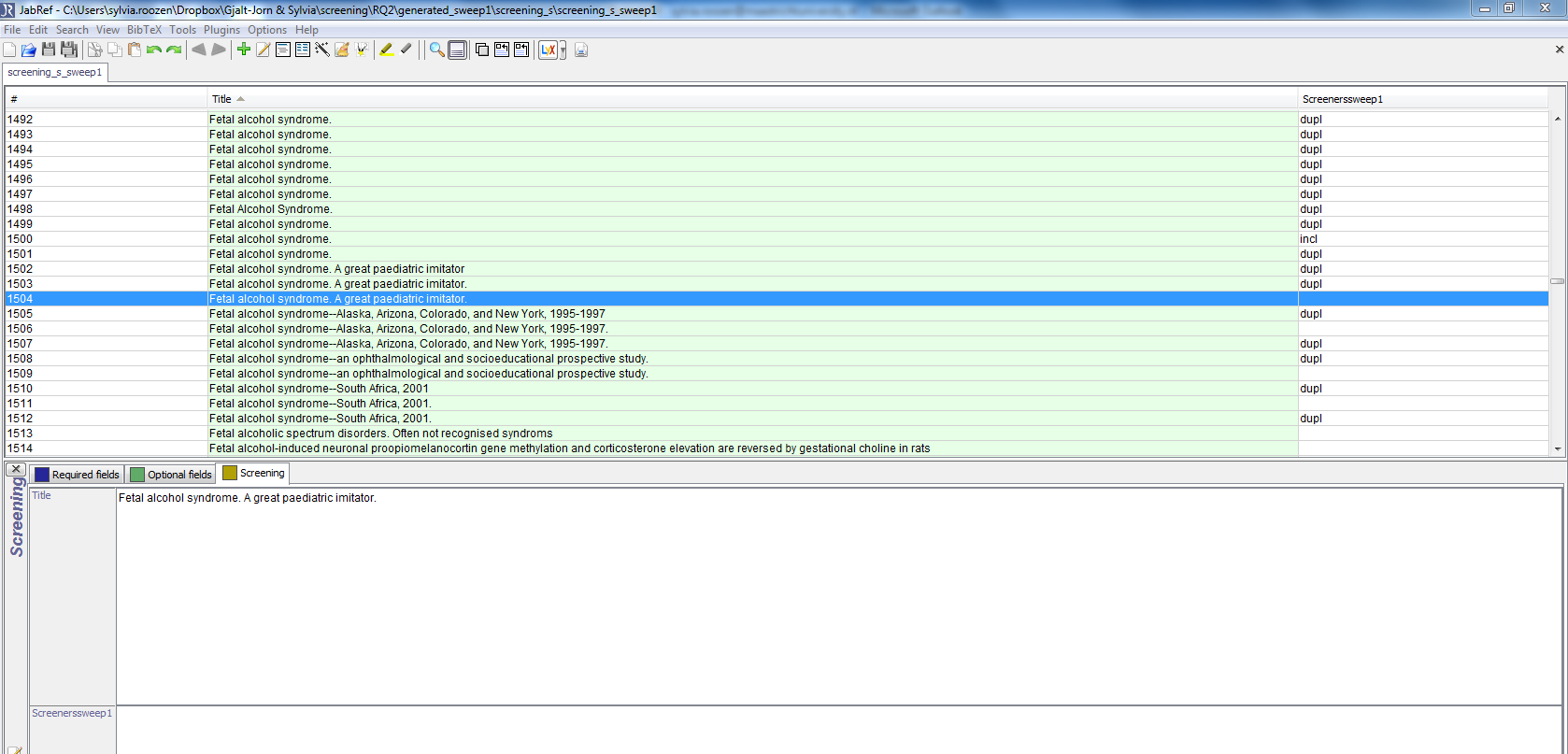
1) Go to your dropbox, navigate to \screening\RQ2\screened\_sweep1   
2) In that directory, open screening\_e (Eline) or screening\_s (Sylvia)  
3) run "remove jabref settings (windows only).bat" - provide administrator permission if Windows requests it (this step deletes old JabRef settings from the Windows Registry)  
4) drag-drop ‘screening\_s.bibtex' (Sylvia) or 'screening\_s.bibtex' (Eline) into JabRef  
5) JabRef should now show columns #, title, abstract, and screenerXsweep1  
8) use the exclusion criteria for Screening phase 1, 2 or 3  
9) enter screening (exclusion/inclusion) decision/criterion in field screenerXsweep1  
(replace the X with your screener letter, i.e. 's' or 'e')

**Collaboration**

There are two independent screeners who can work in the JabRef database simultaneously.

When you save the database for the first time, use the Save Database As … functionality (CTRL-SHIFT-S) to save the database under a new name, "screening\_s\_sweep1.bibtex" or "screening\_e\_sweep1.bibtex", depending on your screening code. The next time you want to start screening again, repeat the steps exactly as mentioned in the section preparation. In this way, you are sure all functions of the file will work in the program exactly how they need to work.

## First screening sweep: titles



In the first screening sweep, you will use JabRef to examine each entry in a .bibtex file. On the basis of your assessment, you will type a code corresponding to your assessment into JabRef. For every screening session, you will keep track of the time you spend on it, to enable assessment of how long screening takes, and whether people speed up as they gain experience.

### Screening an entry

Doubleclick the first entry in the list (or, if you resume screening, locate and open the first entry you haven’t screened yet by looking at the last column in the Entry Table). In the panel that opens now (the Entry Editor), you should see three tabs: Required fields, optional fields, and Screening. For this first screening sweep, you will use Screening1. Click it, and you will see the title field of the first article. In addition, the fourth field should be screenerXsweep, where the X corresponds to your screener code (E or S). Closely study the title and determine the following things and enter the corresponding text in the screenerXsweep field.

**Description Code:**

* The study is a duplicate of an entry which was already coded **Dupl**
* The study is a review or meta-analysis **Review**
* The study does not involve human subjects **Animal**
* The study is in another language **English**
* The study addresses issues not relevant to parental alcohol-related behaviors that predict FASD\* **Var\*\***
* All others **Incl**

**\*Note**: Fetal Alcohol Spectrum Disorder (FASD) includes Fetal Alcohol Syndrome, partial Fetal Alcohol Syndrome, Alcohol Related Neurodevelopmental Disorder, Alcohol Related Birth Defects, Prenatal Alcohol Exposure or Fetal Alcohol Exposure.

**\*\***The code **Var** during this first sweep may include: health status of pregnant women, perception and knowledge concerning FASD, interventions, biomarkers, workshops or erratums, studies that examine different brain dysfunctions of children with FASD but not in relation to paternal drinking behavior, studies that do not involve FASD.   
  
As you can see, this list has been developed so as to be progressive. For example, once you determined that an entry does not report a study with human participants (e.g. with animals, or just a discussion or opinion piece), you can code it as ‘animal’ and you can move on; after all, the categories below ‘animal’ can no longer apply. Note also that the goal of this screening sweep is to exclude as many articles as possible – but that you should only exclude an article if you are confident that the article does not report FASD and/or behaviours that predict FASD. In the next stage, you will examine the titles, abstracts and keywords of all articles that are marked as ‘incl’ by at least one screener. Therefore, if you are not certain that an article should be excluded, mark it as ‘incl’.

## Second screening sweep: titles and abstracts

### Preparation

Before you can start screening for the second screening sweep, you need to make sure that you follow these steps:

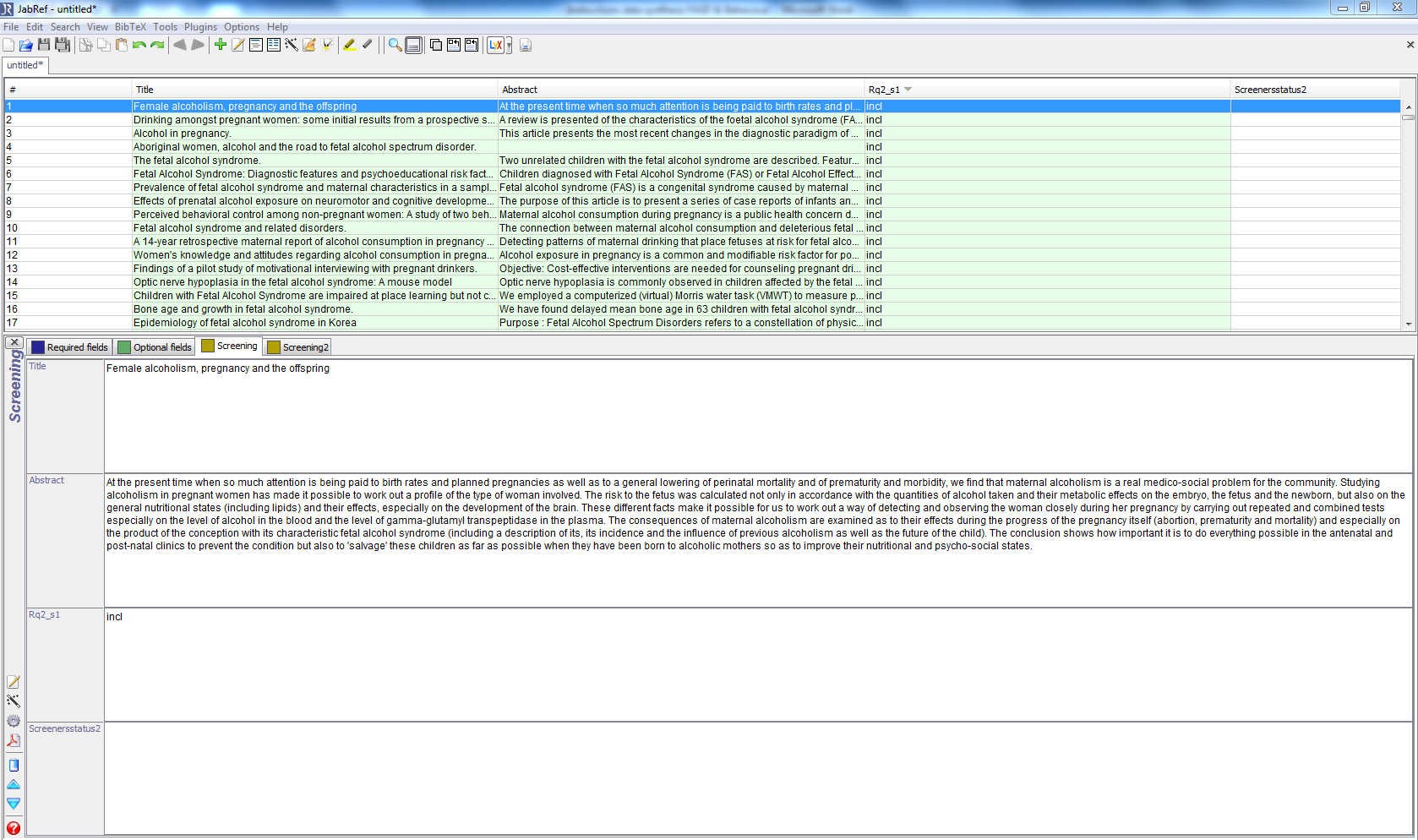
1) Go to your dropbox, navigate to \screening\RQ2\screened\_sweep2   
3) run "remove jabref settings (windows only).bat" - provide administrator permission if Windows requests it (this step deletes old JabRef settings from the Windows Registry)  
4) drag-drop ‘rq2\_s2\_s.bibtex' (Sylvia) or rq2\_s2\_e.bibtex' (Eline) into JabRef  
5) JabRef should now show columns #, title, abstract, and screenerXsweep1  
8) use the exclusion criteria for Screening phase 1, 2 or 3  
9) enter screening (exclusion/inclusion) decision/criterion in field screenerXsweep1  
(replace the X with your screener letter, i.e. 's' or 'e')

When you still encounter problems with the correct settings (step 4), try the followings:

Import setting into JabRef:  
  
1) Dubbelclick on the .jar -> that will start Java and in Java will start .jar (JabRef) in return  
2) Options -> preferences -> bottom left, 'import preferences'  
3) Select the .xml in the good screening folder   
  
If this works, you will see the same screen as the example below.

### Acquiring the abstracts

After the first coding sweep, the abstracts will be acquired of all articles that were marked as ‘incl’ by at least one screener. Then the second screening sweep will start, where you will examine the abstracts to verify whether each paper should indeed be included. For the included papers, you will specify what type of data is reported.



### Screening an entry

Doubleclick the first entry in the list (or, if you resume screening, locate and open the first entry you haven’t screened yet by looking at the last column in the Entry Table). In the panel that opens now (the Entry Editor), you should see three tabs: Required fields, Optional fields, and Screening. For this second screening sweep, you will use Screening. Click it, and you will see the title and abstract fields of the first article. In addition, the fourth field should be screenerXstatus2, where the X corresponds to your screener code (E or S). If your window doesn’t look like the example, please contact the project leader. Closely study the title, abstract and determine the following things and enter the corresponding text in the screenerXstatus2 field.

**Description Code:**

* The study is a duplicate of an entry which was already coded **Dupl**
* The study is a review or meta-analysis **Review**
* The study does not involve human subjects **Animal**
* The study is in another language **English**
* VAR – the study does not include the following **Var\***
  + The study does not address the association between i) maternal alcohol   
    drinking behavior (e.g. quantity drinking, style, before or during pregnancy) and ii)   
    fetal alcohol exposure or any diagnose of FASD\*\* related to alcohol consumption   
    during pregnancy.
* Important: also include every article where the abstract is missing. During third screening **Incl**  
  sweep we will examine the full texts of these missing abstracts. Also, include studies that established  
  FASD diagnosis. Maternal drinking history is one of the criteria for diagnosis, so most probably the studies  
  will explain behavior for the FASD diagnosis.

**\***The code **Var** during this first sweep may include: health status of pregnant women, perception and knowledge concerning FASD, interventions, biomarkers, workshops or erratums, chapters from books, overview studies, studies that examine different brain dysfunctions of children with FASD but not in relation to paternal drinking behavior, or studies that do not involve FASD.

**\*\*Note**: Fetal Alcohol Spectrum Disorder (FASD) includes Fetal Alcohol Syndrome, partial Fetal Alcohol Syndrome, Alcohol Related Neurodevelopmental Disorder, Alcohol Related Birth Defects, Prenatal Alcohol Exposure or Fetal Alcohol Exposure. Please note that we are interested in maternal behavior as part of a FASD diagnosis.

When you save the database for this second, use the Save Database As … functionality (CTRL-SHIFT-S) again to save the database under a new name, "screening\_s\_sweep2.bibtex" or "screening\_e\_sweep2.bibtex", depending on your screening code.

## Third screening sweep: full texts

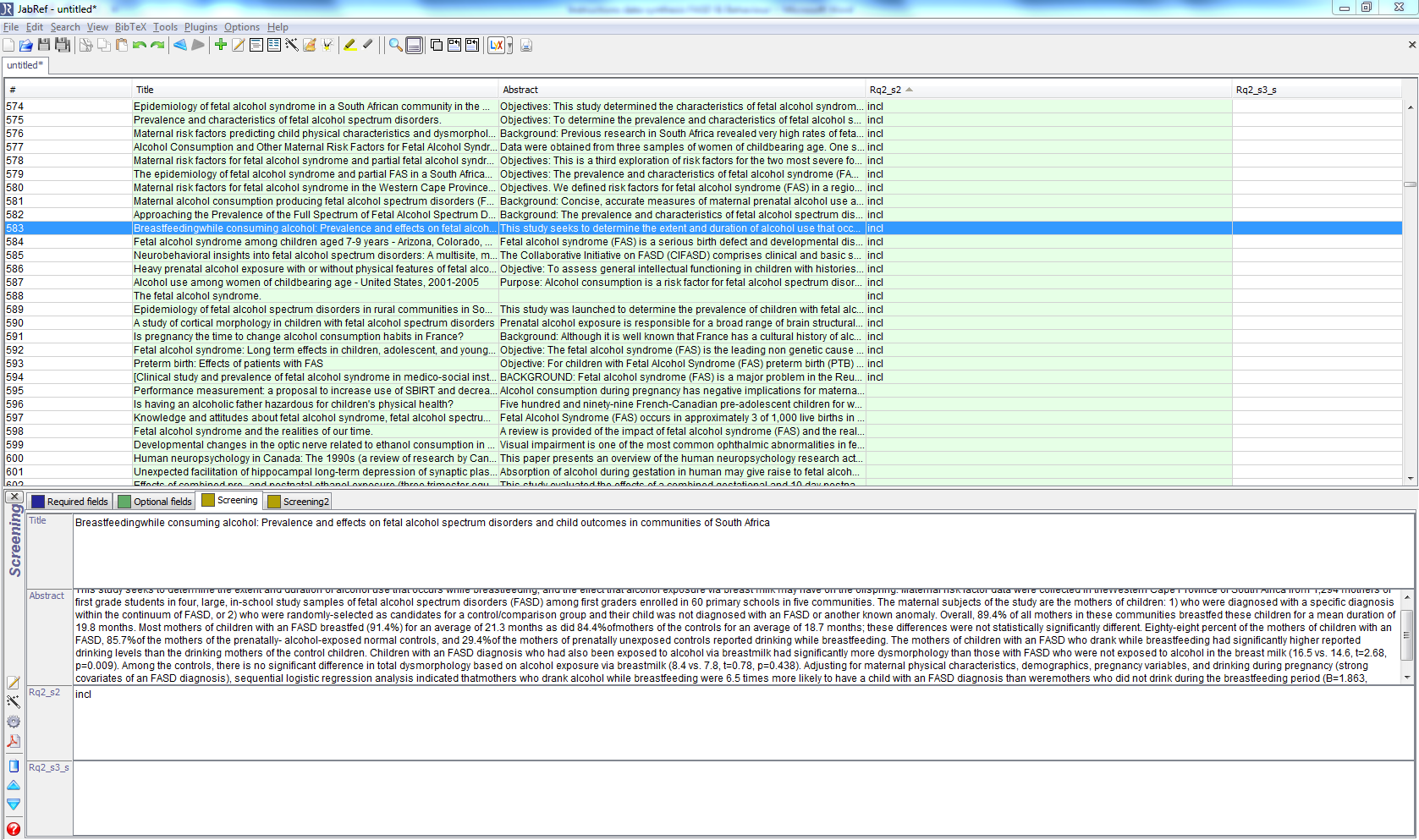
### Preparation

Before you can start screening for the third screening sweep, you need to make sure that you follow these steps:

1) Go to your dropbox, navigate to \screening\RQ2\screening\_sweep3   
3) run "remove jabref settings (windows only).bat" - provide administrator permission if Windows requests it (this step deletes old JabRef settings from the Windows Registry)  
4) drag-drop ‘rq2\_s3\_s.bibtex' (Sylvia) or ‘rq2\_s3\_e.bibtex' (Eline) into JabRef  
5) JabRef should now show columns #, title, abstract, and screenerXsweep1  
8) use the exclusion criteria for Screening phase 1, 2 or 3  
9) enter screening (exclusion/inclusion) decision/criterion in field screenerXsweep1  
(replace the X with your screener letter, i.e. 's' or 'e')

When you still encounter problems with the correct settings (step 4), try the followings:

Import setting into JabRef:  
  
1) Dubbelclick on the .jar -> that will start Java and in Java will start .jar (JabRef) in return  
2) Options -> preferences -> bottom left, 'import preferences'  
3) Select the .xml in the good screening folder   
  
If this works, you will see the same screen as the example below.



**Acquiring the full texts**

After the first and second coding sweep, the full texts will be acquired of all articles that were marked as ‘incl’ during the second sweep by at least one screener. Note: in the field Rq2\_x2 you will see ‘empty’, ‘incl’ and ‘excl’ fields. These empty fields where excluded after the first sweep and de ‘excl’ are excluded during the second sweep. Only acquire the full texts from the ‘incl’ articles. Then the third screening sweep will start, where you will examine the full texts from the PDFs to verify whether each paper should indeed be included.

These full texts can be acquired from their PDFs in your folder (Dropbox\Gjalt-Jorn & Sylvia\screening\RQ2\screened\_sweep3\Articles). You simply open each PDF in the folder and look for the corresponding entry in the JabRef database. Remember where in the file list you are when you stop screening and continue at a later time; or, alternatively, rename the PDFs you already screened (e.g. by adding a dash or the string ‘[done]’ to the start of the filename).

**Screening an article**  
Double click the first entry in the list (or, if you resume screening, locate and open the first entry you haven’t screened yet by looking at the last column in the Entry Table). In the panel that opens now (the Entry Editor), you should see three tabs: Required fields, Optional fields, and Screening. For this third screening sweep, you will use Screening. Click it, and you will see the title and abstract fields of the first article. In addition, the fourth field should be screenerXstatus3, where the X corresponds to your screener code (E or S). If your window doesn’t look like the example, please contact the project leader. Closely study the full text of the article in de pdf file and determine the following things and enter the corresponding text in the screenerXstatus2 field.

**Description Code:**

* The study has no full text available (please keep article separate in folder ‘Picarta full text request’ **Empty**
* The study is a duplicate of an entry which was already coded. **Dupl**
* The study is a review or meta-analysis. **Review**
* The study is an opinion or not a full text article (f.e. conference abstracts) **Opinion**
* The study does not involve human subjects. **Animal**
* The study is not published in English language**. English**
* The study does not report empirically examined data (e.g. book) **Empiri**
* VAR – the study does not include one of the followings **Var\***
  + The study does not address the association between i) maternal alcohol   
    drinking behavior (e.g. quantity drinking, style, before or during pregnancy) and ii)   
    fetal alcohol exposure or any diagnose of FASD\*\* related to alcohol consumption   
    during pregnancy.
* Include all other studies that mention the association as described above **Incl**   
  Please note that it is also necessary to have a control group in the sample.

**\***The code **Var** during this first sweep may include: health status of pregnant women, perception and knowledge concerning FASD, interventions, biomarkers, workshops or erratums, chapters from books, overview studies, studies that examine different brain dysfunctions of children with FASD but not in relation to paternal drinking behavior, or studies that do not involve FASD.

**\*\*Note**: Fetal Alcohol Spectrum Disorder (FASD) includes Fetal Alcohol Syndrome, partial Fetal Alcohol Syndrome, Alcohol Related Neurodevelopmental Disorder, Alcohol Related Birth Defects, Prenatal Alcohol Exposure or Fetal Alcohol Exposure.

When you save the database, use the Save Database As … functionality (CTRL-SHIFT-S) again to save the database under a new name, "screening\_sweep3\_s.bibtex" or " screening\_sweep3\_e.bibtex", depending on your screening code.

Iterative query development, ascendancy and descendancy

Throughout your screening, you may discover keywords that have not yet been included in the query. If you do, email the project leader. The query will be updated and repeated, so that additional relevant articles may be located. In addition, once the final list of included articles is available, the references will be studied to locate additional articles to include. Finally, a list of the articles citing each included article will be retrieved and scanned to look for additional articles to include. Then the screeners will continue with screening of these new references following the screening steps 1 – 3 in this manual. The files can be found in this directory: Dropbox\Gjalt-Jorn & Sylvia\screening\RQ1\_RQ2\_update (2015-06-23)\generated\_q2\_s1.

# Extractor Instructions

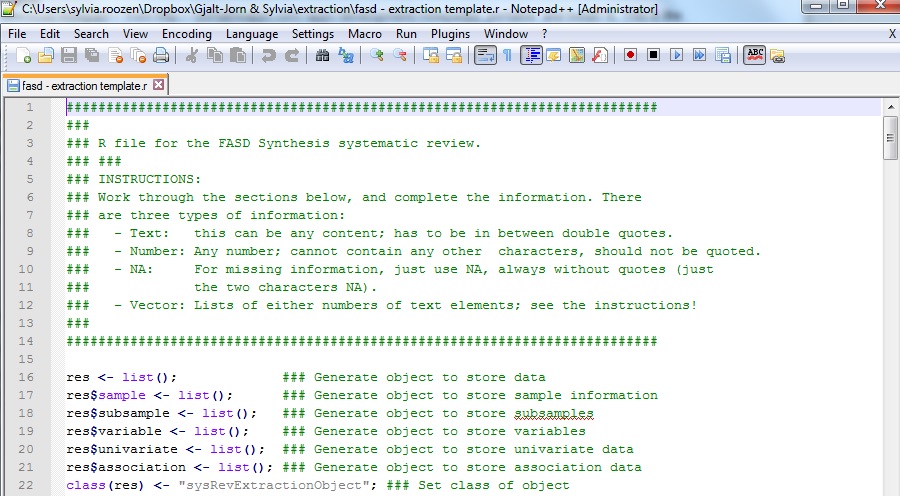
To study the prevalence of FASD, behaviours that predict FASD, prevention efforts which have been made   
and normative questions which have been raised for FASD and analogous conditions.

As an extractor, you will use a portable version of Notepad++, an open source text editor, to generate plain text files that specify which results were found in a study. These files are in the format of R, an open source statistical analysis program and programming language; however, you will not need to understand this language. This document details what you do to extract the methods and results from each article.

## Preparation and collaboration

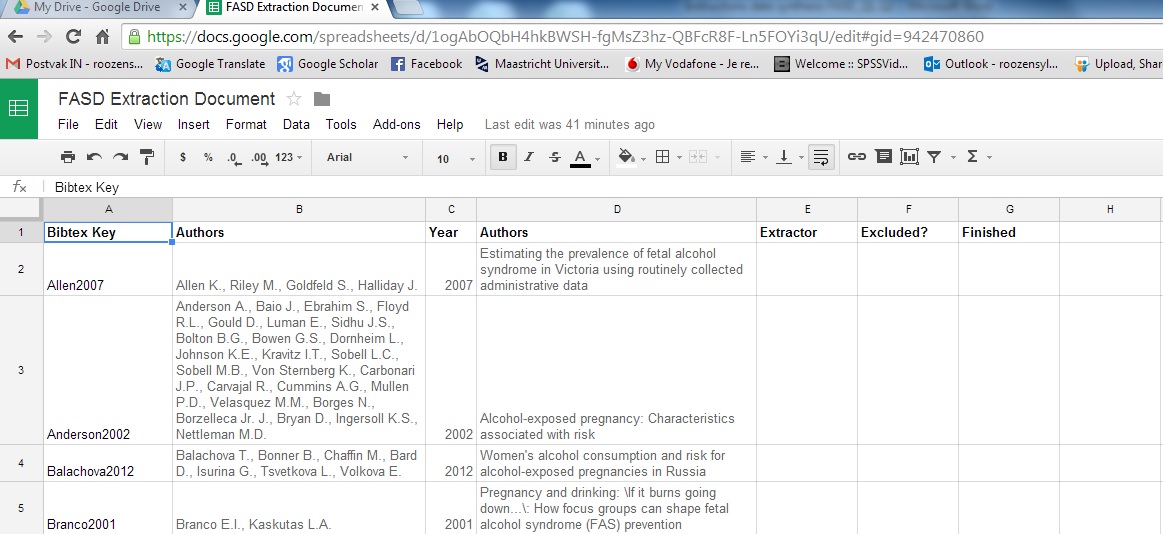
### Preparation

First, download Notepad++ from <http://portableapps.com/apps/development/notepadpp_portable> and install it. This is the text editor you will use. You can also use another text editor if you prefer, as long as it has syntax coloring for .r files, which will help prevent errors (e.g. forgetting a closing quote). When you open Notepad++, it will look roughly like this:



### Collaboration

Because you will not be the only extractor, we use an online spreadsheet that can be edited and viewed by several people simultaneously. It is located at [**this link**](https://docs.google.com/spreadsheets/d/1ogAbOQbH4hkBWSH-fgMsZ3hz-QBFcR8F-Ln5FOYi3qU/edit?usp=sharing), and the sheet ‘Extraction’ contains a list of the articles that were included after the two screening phases. As an example, the document looks like this:



When you start extracting the methods and results from an article, please start with placing your name in the ‘E’ column, so that nobody else starts with the same article while you’re working on it. If closer inspection reveals that the article should in fact be excluded, indicate this, and the reason, in the ‘F’ column. If you finished an article, please indicate this in the ‘G’ column.

## Example: Method & results extraction: general article information

Above, a screenshot is shown of Notepad++. As you can see, the characters in the text file have been colored in different colors. These colors represent the meaning of the symbols and words. Green is used for comments; grey is used for text strings; orange is used for numbers, and light blue is used for NA, the special symbol that means ‘missing’ or ‘empty’ (black is used for variable names and dark blue for operators). Purple is used for recognized special keywords (functions that can be used in R – but you don’t need to worry about these). On the following pages, the text of the default ‘template’ document will be shown, with explanations and examples of how to enter the extracted methods and results. First, an example of an earlier study template document will be shown in its entirety; then, it will be explained section by section.

##########################################################################

###

### R file for the FASD Synthesis systematic review.

### ###

### INSTRUCTIONS:

### Work through the sections below, and complete the information. There

### are three types of information:

### - Text: this can be any content; has to be in between double quotes.

### - Number: Any number; cannot contain any other characters, should not be quoted.

### - NA: For missing information, just use NA, always without quotes (just

### the two characters NA).

### – Vector: Lists of either numbers of text elements; see the instructions!

###

##########################################################################

res <- list(); ### Generate object to store data

res$sample <- list(); ### Generate object to store sample information

res$subsample <- list(); ### Generate object to store subsamples

res$variable <- list(); ### Generate object to store variables

res$univariate <- list(); ### Generate object to store univariate data

res$association <- list(); ### Generate object to store association data

class(res) <- "sysRevExtractionObject"; ### Set class of object

##########################################################################

### BIBTEX KEY

##########################################################################

### Store the Bibtex key as text

res$bibTexKey <- "author9999"

##########################################################################

##########################################################################

###

### GENERAL INFORMATION / COMMENTS ON THIS STUDY:

###

res$generalComments <- "

You can enter comments here.

";### Make sure this line starts with a double quote and a semicolon (";)

##########################################################################

##########################################################################

##########################################################################

### SAMPLING AND SAMPLE DESCRIPTION

##########################################################################

### Description of sampling method

res$sample$samplingMethod <- NA; ### 'prospective' vs 'retrospective'

res$sample$samplingAselect <- FALSE; ### can also be TRUE

res$sample$samplingControls <- NA; ### NA for prospective designs; list of

### variables on which controls were matched,

### or "none" when controls were not matched.

### For example, c("SES mother", "age child");

res$sample$recruitmentSetting <- NA; ## Is "records", "school", "population", "hospital", or NA

### Sample Method

res$sample$native <- FALSE; ### TRUE for native/aboriginal populations (e.g. inuit, aboriginals, etc)

### Description of sample, in text

res$sample$description <- "Description of sample"

### Description of geography of sample, in text

res$sample$geography <- "For example 'Netherlands'"

### Study year, in text

res$sample$year <- NA;

### Sample size, as a number

res$sample$size <- NA;

### SUBSAMPLES - REMOVE THIS SECTION IF THERE ARE NO SUBSAMPLES

### Indicate which information you want to override. Specify the subsample

### you're providing new data for in the single quotes between the double

### brackets; make sure the spelling is exactly identical to the list above.

### Use NA to specify data that should not be overridden.

res$subsample[[length(res$subsample) + 1]] <- list( ### Create object for this subsample

name = NA,

description = NA,

size = NA ### Additional variables specified as univariate data for this subsample

)

##########################################################################

### METHODOLOGICAL INFORMATION

##########################################################################

### Specify how the data was collected ("qualitative" or "quantitative")

res$datacollectionmethod <- NA;

### Specify how many measurement moments the study has, as a number (of

### course, this will just be 1 for most studies)

res$measurementMoments <- 1

### For each measurement moment, specify the number of days since the

### first measurement moment. This is a vector of days; e.g., for three

### measurement moments, each a month apart, this would be:

### res$measurementTimes <- c(0, 30, 60);

res$measurementTimes <- c(0)

##########################################################################

### ASSOCIATIONS: CORRELATIONS, T-TESTS, CHI-SQUARES, ETC

##########################################################################

res$association[[length(res$association) + 1]] <- list(

variable1 = "Variablename 1", ### Name of first variable

var1values = NA, ### Values of var 1 in this analysis (e.g. post-hoc)

var1moment = 1, ### Measurement moment of var 1

variable2 = "Variablename 2", ### Name of second variable

var2values = NA, ### Values of var 2 in this analysis (e.g. post-hoc)

var2moment = 1, ### Measurement moment of var 2

subsample = NA, ### Name of relevant subsample

df1 = NA, ### degrees of freedom (numerator)

df2 = NA, ### for anova, second degrees of freedom (denominator)

t = NA, ### t-value

F = NA, ### F-value

chisq = NA, ### Chi square value

r = NA, ### Pearson correlation

d = NA, ### Cohen's d value

etasq = NA, ### Eta squared value

OR = NA, ### Odds ratio

p = NA, ### P-value (if nothing else is available)

covariates = 0, ### Number of covariates (for multivariate analyses)

comment = NA ### Comment (e.g. calculation, page, etc)

);

##########################################################################

##########################################################################

### UNIVARIATE RESULTS: PERCENTAGES, MEANS, ETC

##########################################################################

res$univariate[[length(res$univariate) + 1]] <- list(

variable = "variable name", ### Name of variable this result pertains to

subsample = NA, ### Name of relevant subsample, or NA

value = NA, ### Value this result pertains to

moment = 1, ### Measurement moment this datapoint pertains to

minimum = NA, ### Minimum or NA

maximum = NA, ### Maximum or NA

median = NA, ### Median or NA

mean = NA, ### Mean or NA

percentage = NA, ### Percentage that endorsed 1, or NA

sd = NA, ### Standard deviation, or NA

qualitative = NA, ### Description (qualitative research)

comment = NA ### Comment or NA

);

##########################################################################

##########################################################################

### MEASUREMENTS AND MANIPULATIONS: THE VARIABLES IN THE STUDY

##########################################################################

res$variable[[length(res$variable) + 1]] <- list(

variable = "example", ### Replace 'example' with variable name

moment = 1, ### moment this variable was measured/manipulated

type = "question", ### "question", "aggregate", or "manipulation"

datatype = "numeric", ### "numeric", "logical", "nominal", "ordinal", or "text"

values = NA, ### Possible values; NA or a vector, e.g. c(1, 2, 3)

labels = NA, ### Labels for the values; NA or e.g. c("no", "maybe", "sometimes")

####################################### ONLY FOR FAS DIAGNOSIS VARIABLES

diagnosisMethod = NA, ### "IOM96", "IOM05", or "4digit", etc

syndromeCategory = NA, ### 1 (FAS), 2 (pFAS), 3 (ARBD), 4 (ARND), or 5 (FASD)

maternalDrinkingConfirmed = NA, ### "never", "required", "if possible"

caseascertainment = NA, ### "active", "passive"

####################################### ONLY FOR MATERNAL DRINKING BEHAVIOR VARIABLES

dataCollectionMethod = NA, ### "self-report" or "interview"

timeframe = NA, ### "retrospective" vs "concurrent"

period = NA, ### "before", "1st", "2nd", "3rd", "during", "after", "other"

varType = NA, ### "frequency" or "units"

intensitySpecification = NA, ### NA, "any day", "weekday", "weekendday", "friday", "saturday", etc

specificationUnits = NA, ### "ml", "mg", "oz"

specificationTimeframe = NA, ### "per week", "per month", "per year"

aboutBinging = FALSE, ### TRUE if the variable is about binge drinking

aboutAlcoholism = FALSE, ### TRUE if the variable is about alcoholism

description = NA, ### Description, as text; or NA

comment = NA ### Comment, as text; or NA;

##########################################################################

##########################################################################

### END

##########################################################################

### Some verification functions; note that these need the 'userfriendlyscience'

### package to be installed and loaded (with 'require').

### This function checks whether for all variables that are used, the

### operationalisation is also extracted, and vice versa:

#extractionVerification(res);

### For showing the entire dataframes

#print.data.frame(extractionVerification(res)$dat$extractedVariables)

#print.data.frame(extractionVerification(res)$dat$extractedUnivariate)

#print.data.frame(extractionVerification(res)$dat$extractedAssociations)

### Entering the extracted information

Information is entered by specifying variable values. In fact, the only thing you have to do as an extractor is enter information in the template file.There are five basic types of input that can be used. They are listed here, with examples:

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Description** | **Example 1** | **Example 2** |
| NA | This means ‘missing value’; use this when the relevant variable does not apply, or when the value is not specified in the article. NA becomes bright blue in Notepad++. NA can in fact be entered anywhere where the values below can be entered to indicate no relevant value is available. It’s the only one with such flexibility. | NA | NA |
| Text | This can be used for all text values. Text values are contained between double quotes ("like this"), and become grey if entered correctly. | "First example" | "Second example" |
| Number | Numbers are, well, numbers. No quotes, and use de point (.) when entering decimals. Numbers become orange when entered correctly. | 437 | 3.87 |
| Textvec | Vectors are basically just lists of values. They are specified with the function c**()**. The c becomes purple; the parentheses dark blue; and the list elements are separated by comma’s and show up in grey. Vectors can also contain only one value, in which case no comma’s are used. | c**(**"First element"**,** "Another one"**,** "Final"**)** | c**(**"This is not a number:"**,** "65"**,** "Because of the quotes"**)** |
| Numvec | These vector are the same as vectors of text, except that they contain numbers. | c**(**543**,** 12**,** 3424**)** | c**(**64**,** 12**)** |

For all information you enter, you will use one of these five types. (In case you were wondering: confusingly enough, in R, ‘vector’ is the name for a list of elements; another data type, which is actually a more abstract ‘object’, is called a ‘list’.)

### Bibtex Key

The file starts with comments explaining the function of the file, which are followed by a number of commands that can be ignored (they create a number of variables that will be used later on). The first relevant line of this file is line 29, where the Bibtex Key of the article you’re extracting data from has to be entered. The Bibtex key is a text string, and should therefore be contained between double quotes. Therefore, for the Bibtex key ‘ahmed2004’, the line would become:

### Store the Bibtex key as text

res**$**bibTexKey **<-** "ahmed2004"

Lines 34 to 40 are useful for providing additional comments about a study. For example, if you think it’s useful to indicate, for later, where some information is provided in the article, this is the place to explain it. Also, if you make a decision to NOT include some information from the article, you can justify that decision there.

### Sample information and subsamples

Line 66 and to 80 of the template file contain information about the study sample, and, if applicable any subsamples. This information is usually found in the Methods section of articles. This consists of three information elements: a description of the sample, a description of the geography, and the sample size. Both descriptions are just plain text; like the Bibtex key, therefore, they need to be enclosed within double quotes. The sample size is a number, and therefore it should NOT be quoted. So, if a study describes the data from 238 fifth and sixth-grade students in Nashville, Tennessee, this section would become:

### Description of sample, in text

res**$**sample**$**description **<-** "Fifth and sixth-grade students"

### Description of geography of sample, in text

res**$**sample**$**geography **<-** "Nashville, TN"

### Sample size, as a number

res**$**sample**$**size **<-** 238

Of course, samples are defined by more characteristics than just these three. However, the rest of these are normally variables in a datafile, and therefore, they are inserted below as univariate data.

Some studies do analyses on subsamples, and in that case, it is important to know to which subsample these analyses pertain. Therefore, it is possible to specify subsamples. When specifying such a subsample, it is possible to override one or several (or all) of the characteristics that were provided for the main sample. Basically, the characteristics are just repeated: if they should be overridden, the new information is provided; if they should not be overridden, this is indicated by using NA, and in that case, the information from the main sample is used. You have to give each subsample a name that is unique to the current study.

res$subsample[["subSample1"]] <- list( ### Create object for this subsample

description = NA,

size = NA, ### Additional variables specified as univariate data for this   
 subsample

adoptionInfo = "Qualitative description of adoption information and a description of the location in the   
 article"

### Methodological information

After specifying the sample, some methodological study characteristics are provided. Specifically, the method that was used to collect data (with possible values “qualitative” and “quantitative”), the number of measurement moments (a simple integer, i.e. a number), and the day for each measurement moment, counting the first measurement moment as day 0 (a vector, or list, of integers). For example, if we have a quantitative study with one measurement moment, we would indicate this like this:

### Specify how the data was collected ("qualitative" or "quantitative")

res**$**datacollectionmethod **<-** "quantitative"

### Specify how many measurement moments the study has, as a number (of

### course, this will just be 1 for most studies)

res**$**measurementMoments **<-** 1

### For each measurement moment, specify the number of days since the

### first measurement moment. This is a vector of days; e.g., for three

### measurement moments, each a month apart, this would be:

### res$measurementTimes <- c(0, 30, 30);

res**$**measurementTimes **<-** c(0)

And if we would have a quantitative study with three measurement moments; the baseline, one after a month, and one after three months, we would indicate this like this:

### Specify how the data was collected ("qualitative" or "quantitative")

res**$**datacollectionmethod **<-** "quantitative"

### Specify how many measurement moments the study has, as a number (of

### course, this will just be 1 for most studies)

res**$**measurementMoments **<-** 3

### For each measurement moment, specify the number of days since the

### first measurement moment. This is a vector of days; e.g., for three

### measurement moments, each a month apart, this would be:

### res$measurementTimes <- c(0, 30, 60);

res**$**measurementTimes **<-** c(0, 30, 90)

## Method & results extraction: results and measurements

NOTE -> Associations: Not applicable for research question 1 (prevalence rates of FASD). Continue with univariate results

### Associations

After having specified this information, we go examine the Results section of the article. It may seem counter-intuitive to immediately jump forward to the associations: after all, the Methods section also contains the Measurements section, which we also need to extract. However, it is useful to first get a clear idea of the variables for which we actually need to extract the measurement (or manipulation) information. After all, research articles often measure a lot of variables for which they don’t actually report any statistical tests; and similarly, a lot of variables that are not relevant in the context of this data synthesis. We start with the most interesting results, the bivariate associations. The template file contains one example which you will copy and paste several times until you entered all the relevant associations:

##########################################################################

### ASSOCIATIONS: CORRELATIONS, T-TESTS, CHI-SQUARES, ETC

##########################################################################

res$association[[length(res$association) + 1]] <- list(

variable1 = "Variablename 1", ### Name of first variable

var1values = NA, ### Values of var 1 in this analysis (e.g. post-hoc)

var1moment = 1, ### Measurement moment of var 1

variable2 = "Variablename 2", ### Name of second variable

var2values = NA, ### Values of var 2 in this analysis (e.g. post-hoc)

var2moment = 1, ### Measurement moment of var 2

subsample = NA, ### Name of relevant subsample

df1 = NA, ### degrees of freedom (numerator)

df2 = NA, ### for anova, second degrees of freedom (denominator)

t = NA, ### t-value

F = NA, ### F-value

chisq = NA, ### Chi square value

r = NA, ### Pearson correlation

d = NA, ### Cohen's d value

etasq = NA, ### Eta squared value

OR = NA, ### Odds ratio

p = NA, ### P-value (if nothing else is available)

covariates = 0, ### Number of covariates (for multivariate analyses)

comment = NA ### Comment (e.g. calculation, page, etc)

);

##########################################################################

For every bivariate association, it is necessary to report at least between which two variables this is an association, as well as of course the association itself. Specifically, this consists of the following elements:

|  |  |  |
| --- | --- | --- |
| **Name** | **Type** | **What you should enter here (if it’s not NA)** |
| variable1 | Text | The variable name of the variable that is the independent variable: thus, a belief or determinant, such as attitude or self-efficacy. Note that variable names have to be unique for every variable in this study (but different studies can use the same variable names); you will use the exact same variable names when you specify how it’s measured in the ‘MEASUREMENTS AND MANIPULATIONS’ section below. |
| var1moment | Number | The measurement moment to which this measurement of variable 1 pertains. This is always 1 for cross-sectional designs. Entering NA or omitting it is equivalent to entering 1. |
| var1values | Textvec | If this variable is a categorical variable (i.e. a nominal or ordinal variable, thus with three or more possible values that are in fact categories, rather than values on an interval scale), it can be involved in a test where all levels are compared (e.g. anova or chi-square) or in a test where only two levels are compared (e.g. a t-test as post-hoc test after an anova). In the latter case, specify these values here in a vector. For example, if a variable has 4 possible values (1, 2, 3 and 4), and you’re entering the t-value for the comparison of levels 2 and 3, add c**(**2, 3**)** here. A more concrete example. Imagine that we have an ordinal variable smokingRiskPerception, with four levels, specified in MEASUREMENTS AND MANIPULATIONS with c**(1,** 2, 3, 4**)**, and with four labels that are specified in MEASUREMENTS AND MANIPULATIONS as c**(**"low"**,** "medium", "high", "very high"**)**. Imagine further that for some reason, the authors only reported the t-test that compares the categories “low” and “very high”. In that case, you would specify c**(1**, 4**)** here. Note that you will need var1values and var2values very, very rarely. |
| variable2 | Text | The variable name for the dependent variable in the comparison: this will be something like intention or behavior, but you can also be more explicit (which is helpful), like in the explanation of var1values above. |
| var2moment | Number | The measurement moment to which this measurement of variable 2 pertains. This is always 1 for cross-sectional designs. Entering NA or omitting it is equivalent to entering 1. |
| var2values | Textvec | See the explanation of var1values above: note that the dependent variable can be dichotomous or categorical (nominal or ordinal) as well. For example, a t-test is symmetric; if the independent variable is on interval level, and the dependent variable is dichotomous, you would conduct a t-test to test for the association between the variables. |
| subsample | Text | If this association was not tested in the entire sample, but in a subsample, include the name of the subsample here. Note that this name must be exactly equivalent to the one used in the ‘Sample information and subsamples’ section above. |
| df1 | Number | The degree of freedom for whichever test statistic is reported below. If this is not provided, it will later be calculated on the basis of the sample size (or subsample size). If the statistics that you report is an anova, this is the degrees of freedom for the numerator (i.e. for the effect, the between-groups degrees of freedom; this is the smaller of the two). [[1]](#footnote-1) |
| df2 | Number | For anovas, the is the degrees of freedom for the denominator (i.e. for the error, the within-groups degrees of freedom; this is the larger of the two). |
| t | Number | The t-value (the difference in means divided by the standard error) |
| F | Number | The F value (the Mean Squares (MS) for the effect (between groups), divided by the Mean Squares (MS) for the error (within groups). |
| chisq | Number | The chi-square value |
| r | Number | The pearson correlation |
| d | Number | The Cohen’s d value (the difference in means divided by the standard deviation) |
| etasq | Number | The Eta Squared value (the Sum of Squares (SS) explained by the effect (the between groups SS), divided by the total Sum of Squares (SS) of the dependent variable) |
| OR | Number | Odds Ratio |
| p | Number | Here, you can report an extracted p-value, because sometimes, that’s all that’s reported. |
| covariates | Number | If this association is based on a multivariate analysis, enter the number of covariates (i.e. the number of other predictors in the same analysis) in that analysis here. |
| comment | Text | Any additional explanation, such as the page number and description of the location on the page where you found the value(s), any calculations you performed, etc. |

It is also possible that a study does not report any of these numbers, but that it reports different statistics from which these numbers can perhaps be calculated. In that case, please contact the principal investigator to discuss the situation.

Basically, what you will be doing in this phase, is just looking through the text of the results section of the relevant article, scanning for numbers. For every number you find, you establish whether it describes a bivariate association for example between i) maternal alcohol drinking behavior (quantity drinking, style, before or during pregnancy) and ii) fetal alcohol exposure or any diagnose of FASD\* related to alcohol consumption during pregnancy. All bivariate associations between other variables are irrelevant. If an article only reports outcomes from multivariate analyses, make sure you include the number of covariates (i.e. the number of other predictors in the same analysis) in the covariates field. If an article reports both bivariate associations and associations from multivariate analyses, only extract the associations from the bivariate analyses[[2]](#footnote-2).

For every association copy-paste the bivariate assocation fragment of the template file shown above, and add the relevant information. Make sure you include the first and the last line, as shown in the examples below. Other lines can be left out if you want; this is equivalent to entering NA.

### Univariate results

Sometimes articles only report means or percentages for a variable such as “prevalence”; Therefore, such studies can only provide us information about either prevalence of the syndrome. For this, we use the section for univariate results:

##########################################################################

### UNIVARIATE RESULTS: PERCENTAGES, MEANS, ETC

##########################################################################

res$univariate[[length(res$univariate) + 1]] <- list(

variable = "variable name", ### Name of variable this result pertains to

subsample = NA, ### Name of relevant subsample, or NA

value = NA, ### Value this result pertains to

moment = 1, ### Measurement moment this datapoint pertains to

minimum = NA, ### Minimum or NA

maximum = NA, ### Maximum or NA

percentage = NA, ### Percentage that endorsed 1, or NA

comment = NA ### Comment or NA

);

##########################################################################

For every univariate result, it is necessary to report at least which variables it pertains to, as well as an indication as the information that is available about the given variable:

|  |  |  |
| --- | --- | --- |
| **Name** | **Type** | **What you should enter here (if it’s not NA)** |
| variable | Text | The variable name that this result pertains to: thus, a belief or determinant, such as attitude or self-efficacy, or “partner support” . Note that variable names have to be unique for every variable in this study (but different studies can use the same variable names); you will use the exact same variable names when you specify how it’s measured in the ‘MEASUREMENTS AND MANIPULATIONS’ section below. |
| subsample | Text | If this result does not come from the entire sample, but from a subsample, include the name of the subsample here. Note that this name must be exactly equivalent to the one used in the ‘Sample information and subsamples’ section above. |
| moment | Number | The measurement moment to which this result pertains. This is always 1 for cross-sectional designs. Entering NA or omitting it is equivalent to entering 1. |
| minimum | Number | Minimum value (i.e. lowest possible score). This is helpful to interpret the mean/median etc. Note that this is only useful for interval (scale) variables! |
| maximum | Number | Maximum value (i.e. highest possible score). This is helpful to interpret the mean/median etc. Note that this is only useful for interval (scale) variables! |
| percentage | Number | For dichotomous measures, or when this univariate result pertains to one level of a nominal or ordinal variable the percentage of participants who answered ‘1’ or who selected this category. For example, if people could check all reasons that they thought contributed to their smoking cessation, each of those reasons would be one variable, the this percentage would be the percentage of participants that endorsed the relevant option. (That means that if the list of potential reasons has 8 reasons, 8 variables and 8 univariate results will be entered into this extraction file.) |
| comment | Text | Any additional explanation, such as the page number and description of the location on the page where you found the reasons/beliefs/thoughts/emotions or value(s), any calculations you performed, etc. |

If you run into anything of which you’re not sure how to enter it, or whether it should be entered, just contact the principal investigator.

How you approach this is different for a quantitative and a qualitative study.

For a quantitative study, you scan the results section for any useful univariate information that is provided. For quantitative research, if no associations are reported, usually frequencies, percentages, or means/standard deviations are reported for the measured variables. Per variable (or for each variable/measurement moment combination, if the study has a longitudinal design) you add a section with univariate data.

#### Example 1 (univariate results extracted from Druschel2007a):



This study reports 4 different rates of FAS prevalence. As table 2 shows, the states Erie and Monroe report their own prevalence rate and use these within two subgroups. To put this table into script, we will have 4 variables: Prevalence Erie White, Non-Hispanic; Prevalence Erie Black; Prevalence Monroe White, Non-Hispanic; Prevalence Monroe Black.

res$univariate[[length(res$univariate) + 1]] <- list(

variable = "prevalence FAS", ### Name of variable this result pertains to

subsample = "Erie, White, Non-Hispanic", ### Name of relevant subsample, or NA

percentage = 3.4, ### Percentage that endorsed 1, or NA

comment = "Extracted from Table 2, p. e387" ### Comment or NA

);

res$univariate[[length(res$univariate) + 1]] <- list(

variable = "prevalence FAS ", ### Name of variable this result pertains to

subsample = "Erie, Black", ### Name of relevant subsample, or NA

percentage = 33.1, ### Percentage that endorsed 1, or NA

comment = "Extracted from Table 2, p. e387" ### Comment or NA

);

res$univariate[[length(res$univariate) + 1]] <- list(

variable = "prevalence FAS ", ### Name of variable this result pertains to

subsample = "Monroe, White, Non-Hispanic", ### Name of relevant subsample, or NA

percentage = 1.8, ### Percentage that endorsed 1, or NA

comment = "Extracted from Table 2, p. e387" ### Comment or NA

);

res$univariate[[length(res$univariate) + 1]] <- list(

variable = "prevalence FAS ", ### Name of variable this result pertains to

subsample = "Monroe, Black", ### Name of relevant subsample, or NA

percentage = 4.0, ### Percentage that endorsed 1, or NA

comment = "Extracted from Table 2, p. e387" ### Comment or NA

);

*Example 2 (univariate results extracted from Viljoen2005a):*

This study reports three different prevalence rates. As we are not sure why they used three different rates to describe the prevalence of the 13 schools which were examined, report all three of them as separate variables.



res$univariate[[length(res$univariate) + 1]] <- list(

variable = "prevalence", ### Name of variable this result pertains to

subsample = "1st grade children screened", ### Name of relevant subsample, or NA

percentage = 7.4, ### Percentage that endorsed 1, or NA

comment = "Extracted from Table 5, p. 600" ### Comment or NA

);

res$univariate[[length(res$univariate) + 1]] <- list(

variable = "prevalence", ### Name of variable this result pertains to

subsample = "1st grade children per all enrolled in 12 schools", ### Name of relevant   
 subsample, or NA

percentage = 6.9, ### Percentage that endorsed 1, or NA

comment = "Extracted from Table 5, p. 600" ### Comment or NA

);

res$univariate[[length(res$univariate) + 1]] <- list(

variable = "prevalence", ### Name of variable this result pertains to

subsample = "1st grade children per all enrolled in all schools", ### Name of relevant   
 subsample, or NA

percentage = 6.5, ### Percentage that endorsed 1, or NA

comment = "Extracted from Table 5, p. 600" ### Comment or NA

);

### Measurements and manipulations

Now we have extracted all associations, and univariate information, about all relevant variables, we need to provide information about the nature of these variables. In the current data synthesis, we only extracted measured variables, so manipulations will not be explained in this section.

To know which variables need to be included in this section, we simply go back to the Associations and Univariate Results sections and look at the variables we included there. Every variable name that we included there has to be described in more detail in the Measurements and manipulations section. This information will usually be provided in the Methods section of articles, although sometimes parts of it can be in the Results section or even the Introduction section.

We extract the following information:

##########################################################################

### MEASUREMENTS AND MANIPULATIONS: THE VARIABLES IN THE STUDY

##########################################################################

res$variable[[length(res$variable) + 1]] <- list(

variable = "ethnicity", ### Replace 'example' with variable name

moment = 1, ### moment this variable was measured/manipulated

type = "question", ### "question", "aggregate", or "manipulation"

datatype = "nominal", ### "numeric", "logical", "nominal", "ordinal", or "text"

values = c(1, 2, 3), ### Possible values; NA or a vector, e.g. c(1, 2, 3)

labels = c('Black', 'White', 'Coloured'), ### Labels for the values; NA or e.g. c("no", "maybe", "sometimes")

comment = NA ### Comment, as text; or NA

);

res$variable[[length(res$variable) + 1]] <- list(

variable = "prevalence", ### Replace 'example' with variable name

moment = 1, ### moment this variable was measured/manipulated

type = "question", ### "question", "aggregate", or "manipulation"

datatype = "ratio", ### "numeric", "logical", "nominal", "ordinal", or "text"

values = NA, ### Possible values; NA or a vector, e.g. c(1, 2, 3)

labels = NA, ### Labels for the values; NA or e.g. c("no", "maybe", "sometimes")

diagnosisMethod = "IOM, Two-tier", ### NA or a value from the Google Document at https://docs.google.com/spreadsheets/d/1ogAbOQbH4hkBWSH-fgMsZ3hz-QBFcR8F-Ln5FOYi3qU/edit#gid=0

diagnosisCutoff = NA, ### NA or description of cut-off, as text

diagnosisProvider = "pediatric dysmorphologist, physician", ### NA or a value from the Google Document at https://docs.google.com/spreadsheets/d/1ogAbOQbH4hkBWSH-fgMsZ3hz-QBFcR8F-Ln5FOYi3qU/edit#gid=0

prevalence = TRUE, ### FALSE or TRUE - TRUE if this concerns a prevalence

parent = NA, ### Name of parent variable into which this one is aggregated

description = NA, ### Description, as text; or NA

extrapolated = TRUE, ### FALSE or TRUE - TRUE is this concerns extrapolation of prevalence

extrapolationComment = "Table 5, page 600",

comment = NA ### Comment, as text; or NA

);##########################################################################

For each variable, it is necessary to at least provide the variable name (note: this is done by replacing "example" on the first line, instead of completing one of the rows with variables below that statement), indicate the type, the datatype, the psytype, and the dependent variable. Explanations of these fields follow.

|  |  |  |
| --- | --- | --- |
| **Name** | **Type** | **What you should enter here (if it’s not NA)** |
| variable | Text | The variable name, which has to match the name using in the Associations and Univariate Results sections perfectly, is entered on the first line, as text (i.e. between double quotes), between the double opening and closing brackets (i.e. ‘straight parentheses’). Note that variable names have to be unique for every variable in this study (but different studies can use the same variable names). |
| moment | Number | The measurement moment to which this measurement specification pertains. This is always 1 for cross-sectional designs. Entering NA or omitting it is equivalent to entering 1. |
| type | Text | The type of this variable. This can be either a question, if it’s a question (or item) in a survey, or more generally, a single measurement of anything; or it can be an aggregate variable, which means it’s the mean, sum, or otherwise derived aggregate measure of a number of other variables (usually questions/items in a scale); or it can be a manipulation, in the case of experiments. The current data synthesis does not concern manipulations, so you just have to indicate whether a variable is an item/question, or whether it’s an aggregate measure (i.e. a sum, scale, mean, etc). |
| datatype | Text | The data type of a variable roughly conforms to the measurement level as construed in statistics: ‘numeric’ for interval variables, ‘logical’ for dichotomous variables, ‘nominal’ for categorical variables with categories that cannot be ranked, ‘ordinal’ for categorical variables with categories that *can* be ranked, and finally ‘text’ for strings (variables from qualitative research). |
| values | Vecnum | The values the variable can take. For categorical variables (nominal and ordinal data types), you have to specify which values correspond to which categories. You do this by first specifying the possible values here. You use a vector of numbers here, such as c(1, 2, 3, 4) |
| Labels | Vectext | Once you specified the values representing each category in the values variable above, now specify what each value corresponds to, again using a vector, such as c(“not at all”, “a bit”, “quite”, “extremely”) |
| diagnosisMethod | Text | The diagnostic method type of a variable describes a ‘nominal’ value of the method used to diagnose the syndrome within this study. Please specifiy in the Google document per study which method the article used to examine prevalence rate. |
| diagnosisCutoff | Text | The diagnostic cutoff of a variable describes a number. This number needs to represent a cutoff point who can be diagnosed with the disorder and who can’t. Often used are percentiles. |
| diagnosisProvider | Text | The diagnostic provider of a variable describes who performed the diagnose. Update the Google document if the study describes the provider of a FASD diagnose. |
| prevalence | Text | Prevalence can be coded in two ways. State TRUE if the study reports a prevalence rate. State FALSE if this study doesn’t report a prevalence rate. |
| parent | Text | If this variable is part of an aggregate variable (i.e. if this is a question/item in a scale that is also included), provide the name of the parent variable here. For example, is the belief “if I smoke, I will likely get lung cancer” is aggregated into attitude, and both data on measures are reported, provide the name you gave to the attitude variable in this slot for the belief variable. |
| description | Text | Here, provide a description of this variable. For example, you can include the question here for questions, or a description of how an aggregate variable is described in the article. |
| Extrapolated | Text | Extrapolation of a variable can be described as TRUE or FALSE. State TRUE if the study extrapolates a prevalence rate. |
| extrapolationComment | Text | Any additional explanation, such as how the extrapolation took place. |
| comment | Text | Any additional explanation, such as the page number and description of the location on the page where you found the reasons/beliefs/thoughts/emotions or value(s). |

If you run into anything of which you’re not sure how to enter it, or whether it should be entered, just contact the principal investigator.

When you get to this phase of the extraction, you will not have to search for variables to include any more. All variables of interest already have entries above, in the Associations or Univariate Results sections. Therefore, you can just go back to the first entry in the Associations section, and go look in the article for the description of the measurement of the variable(s) reported in that first entry (in practice, of course, you will already add variables to the Measurement section as you enter them in the Associations and Univariate sections, to avoid this step later on). Then, you’ll go look for the relevant information. For qualitative studies, you don’t even have to go back to the article; after all, qualitative data is collected through interviews, and no measurement instruments are used. You will see an example below.

#### Example 3 (measurements extracted from Druschel2007a):



**##########################################################################**

**### MEASUREMENTS AND MANIPULATIONS: THE VARIABLES IN THE STUDY**

**##########################################################################**

**res$variable[[length(res$variable) + 1]] <- list(**

**variable = "region", ### Replace 'example' with variable name**

**moment = 1, ### moment this variable was measured/manipulated**

**type = "question", ### "question", "aggregate", or "manipulation"**

**datatype = "ordinal", ### "numeric", "logical", "nominal", "ordinal", or "text"**

**values = c(1, 2, 3, 4), ### Possible values; NA or a vector, e.g. c(1, 2, 3)**

**labels = c('United States, Erie, White Non-Hispanic', 'United States, Monroe, White Non-Hispanic', 'United States, Erie, Black', 'United States, Monroe, Black'), ### Labels for the values; NA or e.g. c("no", "maybe", "sometimes")**

**comment = NA ### Comment, as text; or NA**

**);**

**res$variable[[length(res$variable) + 1]] <- list(**

**variable = "prevalence FAS", ### Replace 'example' with variable name**

**moment = 1, ### moment this variable was measured/manipulated**

**type = "question", ### "question", "aggregate", or "manipulation"**

**datatype = "ratio", ### "numeric", "logical", "nominal", "ordinal", or "text"**

**values = NA, ### Possible values; NA or a vector, e.g. c(1, 2, 3)**

**labels = NA, ### Labels for the values; NA or e.g. c("no", "maybe", "sometimes")**

**diagnosisMethod = "IOM'96", ### NA or a value from the Google Document at https://docs.google.com/spreadsheets/d/1ogAbOQbH4hkBWSH-fgMsZ3hz-QBFcR8F-Ln5FOYi3qU/edit#gid=0**

**diagnosisCutoff = NA, ### NA or description of cut-off, as text**

**diagnosisProvider = "not reported", ### NA or a value from the Google Document at https://docs.google.com/spreadsheets/d/1ogAbOQbH4hkBWSH-fgMsZ3hz-QBFcR8F-Ln5FOYi3qU/edit#gid=0**

**diagnosticTeam = "dysmorphologist, geneticist, developmental psychologist",**

**syndromeCategory = 1, ### 1 (FAS), 2 (pFAS), 3 (ARBD), or 4 (ARND)**

**IOM\_qual\_1996\_A = 4,**

**IOM\_qual\_1996\_B = 1,**

**IOM\_qual\_1996\_Ci = 2,**

**IOM\_qual\_1996\_Cii = 0,**

**IOM\_qual\_1996\_Ciii = 0,**

**IOM\_qual\_1996\_Di = 0,**

**IOM\_qual\_1996\_Dii = 0,**

**IOM\_qual\_1996\_Diii = 2,**

**IOM\_qual\_2005\_A = 4,**

**IOM\_qual\_2005\_Bi = 2,**

**IOM\_qual\_2005\_Bii = 2,**

**IOM\_qual\_2005\_Biii = 2,**

**IOM\_qual\_2005\_Ci = 1,**

**IOM\_qual\_2005\_Di = 0,**

**IOM\_qual\_2005\_Dii = 1,**

**prevalence = TRUE, ### FALSE or TRUE - TRUE if this concerns a prevalence**

**parent = NA, ### Name of parent variable into which this one is aggregated**

**description = NA, ### Description, as text; or NA**

**extrapolated = FALSE, ### FALSE or TRUE - TRUE is this concerns extrapolation of prevalence**

**extrapolationComment = NA,**

**comment = NA ### Comment, as text; or NA**

**);**

**##########################################################################**

## Summary

If at this point anything remains unclear, please contact the principal investigator. If you’re good to go, here is a brief summary of the steps/procedure:

1. Open [this link](https://docs.google.com/spreadsheets/d/1ogAbOQbH4hkBWSH-fgMsZ3hz-QBFcR8F-Ln5FOYi3qU/edit?usp=sharing), select a study, and enter your name in the ‘E’ column; (Only one extractor per article)
2. Open the PDF of the selected study; See folder: \Dropbox\Gjalt-Jorn & Sylvia\extraction\Articles
3. Open the ‘fasd – extraction.r’ template, and enter the BibTex key of the selected study; see folder: Dropbox\Gjalt-Jorn & Sylvia\extraction
4. Save the extraction template file, using the BibTex key as shown in the Google Drive document as the filename. Please copy this BibTex key precisely and save the document in the folder \Dropbox\Gjalt-Jorn & Sylvia\extraction\Extracted Articles;
5. Scan the Methods section (or potentially the Results or even the Introduction sections), and:
   1. extract and enter the sample/subsample information;
   2. extract and enter the methodological information;
6. Scan the results section and:
   1. extract and enter all reported associations between relevant variables; (NA for Q1)
   2. extract and enter all univariate results for relevant variables;
7. Scan the methodology section (or potentially the Results or even the Introduction sections), and:
   1. extract and enter all information about how the variables involved in associations or univariate results were measured;
8. *Point 5-7:* Heterogeneous things we use qualitatively
9. *Point 5-7:* Homogeneous things, we want to use as variables and analyse. In general, these categories are all same among studies.
   1. Appoint subsample to variable in univariate section
   2. Add subsample in list of subsamples
   3. Add features in measurement, see the example about ethnicity
   4. Add univariate result for subsample in univariate section for this attribute (so: ‘value’, a number, which corresponds with the label of the operationalization)
10. Save the extraction file;
11. If in any of the previous steps, you discovered that the article should be excluded, enter this in the ‘D’ column of the online form (with the reason for exclusion); Note Q2 if it is an article included for any other Research Question
12. In the online form, indicate that you finished processing the article in the ‘G’ column

1. The degrees of freedom for an independent sample t-test is the total sample size -2; for a matched pairs t-test, the sample size -1 (i.e. half the number of observations or data points -1); for a Pearson correlation, the sample size -2 (i.e. half the observations or data points -2); for an analysis of variance, the degrees of freedom for the numerator (df1) is the number of categories comprising the relevant variable (i.e. the number of groups) -1, and the degrees of freedom for the denominator (df2) is the number of observations - df1 –2; and for χ2, the product of the number of categories of each variable -1 (i.e. a table of 5x3 has (5-1)x(3-1) = 4x2=8 degrees of freedom). [↑](#footnote-ref-1)
2. The problem with multivariate associations is that variance in the dependent variable that is explained by more than one predictor (i.e. non-uniquely explaind variance) is entirely removed from the model. In addition, effect sizes from multivariate models often compare explained variance to the error variance, which, in multivariate models, is smaller than for bivariate associations, because the explained variance of all predictors has been removed from the dependent variable (whereas for bivariate associations, only the variance explained by the sole predictor has been removed from the variance in the dependent variable to yield the error variance). Therefore, in an integration with effect sizes from bivariate associations, such associations would bias the outcome. [↑](#footnote-ref-2)