SOP16
DATA CHECKING OF MANUALLY INPUTTED DATA

| VERSION NUMBER | 1 | DATE OF VERSION (dd/mm/yyyy) | 1 | 2 | / | 0 | 1 | / | 2 | 0 | 1 | 5 |

WRITTEN/REVIEWED BY	Print Name	Angela Attwood
	Position	Research Fellow
	Signature	
	Date (dd/mm/yyyy)	1 2 / 0 1 / 2 0 1 5

APPROVED BY	Print Name	Angela Attwood
	Position	Research Fellow
	Signature	
	Date (dd/mm/yyyy)	1 2 / 0 1 / 2 0 1 5

DATE OF NEXT SCHEDULED REVIEW (dd/mm/yyyy)	0 8 / 0 1 / 2 0 2 1	
REVIEWED BY	Print Name	Maddy Dyer
	Position	Research Associate
	Signature	Maddy Dyer
	Date (dd/mm/yyyy)	0 8 / 0 1 / 2 0 2 0
	Outcome of review:	Minor changes

Definitions/Abbreviations	
RA	Research Assistant
SOP	Standard Operating Procedure
TARG	Tobacco and Alcohol Research Group
WSL	Work Stream Lead

SOP16
DATA CHECKING OF MANUALLY INPUTTED DATA

1. PURPOSE:

- To provide step-by-step instructions to all persons needing to complete a 20% check of their manually inputted data. In addition to manually entered data, 100% of randomisation allocation should be checked. This should be completed by someone other than the study RA and verification of the result should be sent the WSL.

2. PERSONNEL REQUIRED AND LEVEL OF EXPERTISE:

- Researcher
- Secondary data inputter (independent of study)
- Data checker (usually WSL so referred to as WSL in text)

3. MATERIALS AND EQUIPMENT REQUIRED:

- Random number generating software (e.g., www.randomizer.org)
- Study CRFs
- Spreadsheet (excel or similar) software

4. PROCEDURES:

This section provides step-by-step guidance for data checking procedures of manually inputted data. This should be completed as part of the study closure and the exact final error rate should be logged on the TARG work stream research management log (excel sheet) and your study closure checklist. If error rates greater than 1% are obtained, the data requires full re-inputting. If error rates of 1% or less are obtained, errors should be rectified but no further action is required. As part of the data check procedure, a designated subset of the data (20%) must be re-entered. This should be completed by someone independent of the study. Data should never be checked by the original data inputter. Steps for the data checking process are listed below in order of action. The person who should take lead on each step is given in parenthesis.

1. *Identifying cases for checking (Researcher and WSL)*
   The researcher should contact the WSL and confirm the final number of participants. The WSL will then use random number generating software to identify 20% of cases. For example, if the final sample was 48, ten cases are required for checking (rounded up from 9.6). The WSL will use random number generating software to identify ten participant numbers between 1 and 48. The WSL will then inform the Researcher of the cases (i.e., participant numbers) that require checking.

2. *Producing data check excel sheet (Researcher)*
   When the cases are known, cut and paste the inputted data for each of the cases in to an excel sheet. Participant IDs should be in column 1. The sheet should only contain the data of the cases to be checked. Copy and paste the column headings into the first row of the excel sheet. Send this to the WSL.

   Send a truncated copy of this excel sheet to the Data checker. This version of the sheet should have participant IDs listed in column one, and headings in the first row but should NOT contain any data. The data checker should collect the CRFs for the selected cases and input the full data set. Instruct the data checker to forward this version of the sheet to the WSL once complete. The hardcopy CRFs will need to be passed to the WSL.

3. *Running the data check (WSL)*
   To run the data check, the WSL will need: original data in excel sheet from the researcher, re-inputted data from the independent researcher, and the hardcopy CRFs (for identified cases only).

SOP16
DATA CHECKING OF MANUALLY INPUTTED DATA

Paste the two data equivalent data sets into one excel sheet, so that one data set sits above the other. For ease, highlight the original set in a different colour (see Figure 1).
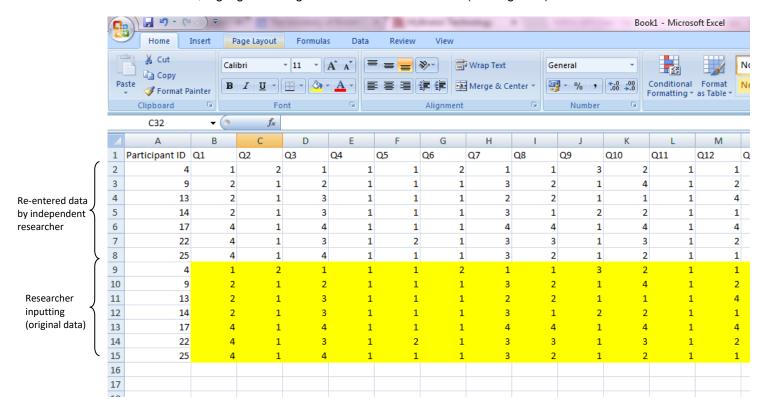

Figure 1

To run a data check subtract each cell from its equivalent cell, by typing =XX-XX into the first available clear cell (in the given example, this would be: =b2-b9). Press enter. You will only need to type in this formula for the first data pair. The formula can then be copied across all remaining cells. When inputted data pairs are the same, the result will equal zero (see Figure 2).

Figure 2.

SOP16
DATA CHECKING OF MANUALLY INPUTTED DATA

To calculate the error rate, you should count all of the cells which show a number other than zero*. For each error you should check the original CRF to verify the correct entry. Only errors in the original data set (Researcher's data) should be counted as errors. Error rate = error count / total number of cells.

*Note: for large data sets you can use conditional formatting to help highlight "non-zero" cells. To do this, select the third set of cells only (subtraction outputs) by clicking on the relevant rows. On the "Home" tab, click Conditional Formatting > Highlight cell rules > Greater than. In the "format cells that are greater than" input "0" and then choose a highlight type (e.g., light red fill with dark red text). Click OK. Repeat last sentence exactly with the exception that this time choose "less than" rather than "greater than". When done, all cells that do not equal zero (i.e., where Researcher and Data Checker data differ) will be highlighted (see Figure 3).
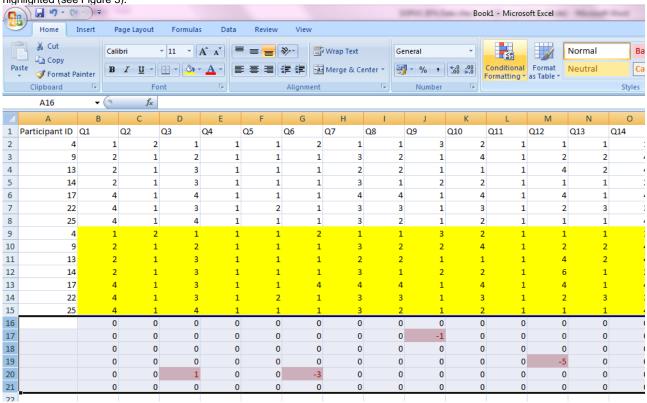

Figure 3

Email the Researcher the result of the data check (excel sheet and error rate). If error rate is greater than 1% a full re-entry of data is required. If error rate is 1% or less, the Researcher should amend the errors identified. Ask the Researcher to include these error rates on their closure checklist. Error rate should be entered into TARG work stream excel sheet by the WSL

4. *Following up the data check (Researcher)*
    a. If the error rate is greater than 1%: A full re-entry of data is required. Any disputes can be discussed with the WSL. For example, if the majority of errors occur on a single measure it may be acceptable to only re-enter data for one measure. This should only be the case if error rate is less than 1% when the problem measure is removed. This is at the discretion of the WSL.
    b. If error rate is less than 1%: The Researcher should amend any errors identified in the final data sheet. Retrieve CRFs and archive appropriately. No further action is required.

SOP16
DATA CHECKING OF MANUALLY INPUTTED DATA

5. TROUBLE SHOOTING:

Problem	Solution
Advice and guidance:	**Work stream lead** **Angela Attwood** **(**0117) 331 7450 internal 17450 Angela.Attwood@bristol.ac.uk