NHS
University Hospitals Plymouth
NHS Trust

STANDARD OPERATING PROCEDURE

DO NOT USE THIS SOP IN PRINTED FORM WITHOUT FIRST CHECKING IT IS THE LATEST VERSION

The definitive versions of all UHPNT RD&I Dept SOPs appear online, not in printed form, to ensure that up to date versions are used. If you are reading this in printed form check that the version number and date below is the most recent one as shown on the Trust's website: https://www.plymouthhospitals.nhs.uk/researchers

Database management, security, design and validation

SOP No:	P13
Version No:	2.2
Effective Date:	Jan 2019
Supersedes:	Version 2.2, Sep 2017
Page:	**1** of **11**

Last Review Date:	Jan 2019	Next review date:	Jan 2022

	APPROVED BY
Name:	Chris Rollinson
Job Title:	Research Governance Manager
Signature:	
Date:	21st Jan 2019

1 Purpose and Scope

The purpose of this Standard Operating Procedure (SOP) is to describe the standard procedures to be followed when a recognised database for storage of clinical trial data is not in place. Research data should be collected, recorded and managed in accordance with the principles of ICH Guidelines for Good Clinical Practice (GCP), the Data Protection Act 2018 and the appropriate University Hospitals Plymouth NHS Trust (UHPNT) policies.

All personal data and sensitive personal data must be managed in accordance with the principles of the Data Protection Act 2018. In particular, all personal data should be kept securely and not transmitted in a way that could cause loss of data or allow interception by unauthorised parties. All persons dealing with personal data are responsible for ensuring the security and safety of these data.

GCP requires that an electronic database should be fit for purpose; provide a clear audit trail; have security of access and be protected against deletions. In a certain proportion of small-scale clinical trials, owing to their size and constraints on funding, there is little technical support for researchers in the design of databases. The aim of this document is to set out a method of optimising compliance with GCP when data are entered from Case Report Forms (CRF) into spreadsheets such as Microsoft Excel.

In Scope: this SOP applies to clinical trials where University Hospitals Plymouth NHS Trust (UHPNT) has accepted the role of ‘Sponsor’

This SOP applies to small-scale clinical trials, where there is no provision for software to enable design of a database in which to store clinical data. These trials will typically be trials where a small number of subjects are being recruited.

This SOP does not apply to commercially funded research or research sponsored by an external non-commercial organisations.

Definitions

PI	Principal Investigators
CI	Chief Investigator
CTIMP	Clinical Trial of an Investigational Medicinal Product
GCP	Good Clinical Practice
HCA	Health Care Assistants
HRA	Health Research Authority

MHRA	Medicines and Healthcare products Regulatory Agency
REC	Research Ethics Committee
RD&I	Research Development & Innovation
RO	Research Office
SOP	Standard Operating Procedure
UHPNT	University Hospitals Plymouth NHS Trust

2 Who should read this document?

All staff involved in setting up a research database e.g. Chief Investigators (CI), Principal Investigators (PI), Trial Co-coordinators / managers, RD&I Managers.

Ultimately the Chief Investigator or delegate is responsible for the design and development of the database. The CI should ensure that all persons with access to the database have full understanding of its content, use and data protection requirements.

3 Procedure to Follow

3.1. Management of electronic data

I. The study protocol, or other study document, should clearly specify the data to be stored electronically and which data elements are to be retained upon archiving.

II. Electronic data should be accurate and reliable.

III. The computerised system should safeguard study blinding where this exists. Blinding should not be broken through the day-to-day use of the computerised system. The system should provide protection against unintentional unblinding but support, where appropriate, any unblinding procedures described in the study protocol.

IV. There will be an unambiguous and unique identifying ID for each participant. The code or file linking participants' names with their IDs should be kept secure and separate from the data used for trial analysis.

V. Identifiable data should not be retained for longer than is necessary to meet the requirements described in the study protocol and to meet requirements set by the grant-awarding body, Research Ethics Committee and regulatory authorities. For the latter, this is at least five years after the conclusion of the trial. Deleting data may require the physical destruction of digital media.

VI. Electronic data should be archived in a way that is secure and which uses media with the appropriate longevity for the required storage time. If application software is required to read the archived data, this software should be archived together with the data.

3.2. Security

I. Electronic data should only be stored on devices that that are backed up in a secure and timely manner. Data should not be held on devices that do not participate in a backup regime. In practical terms this will generally mean:

- liaising with central IT services to confirm that the server upon which the CTIMP database is stored is regularly (e.g. daily) backed-up onto remote and/or removable disk storage, the latter of which are then placed into a secure fire safe.
- not using systems where the data are stored only on the hard-disk of a PC or laptop.

II. Personal data must not be stored or transmitted on removable media or laptops without encryption (UHPNT support the use of encrypted pen drives which can be obtained through the Trust).

III. Any machines used to enter or access trial data should have up-to-date operating system patches installed together with appropriate security software (e.g. antivirus, anti-spyware and firewall).

IV. Access to electronic data must require a unique password.

V. Data used for trial analysis should be anonymised. Data used for trial management should use the minimum number of personal identifiers necessary for study conduct and ensuring patient safety.

VI. Access to the data should be limited to authorised personnel and each user of the system should have an individual account. Accounts must never be shared, nor should there be ‘guest’ accounts.

VII. Individual users should login with their own usernames and passwords.

VIII. Individual users should not login so as to provide access to another user.

IX. A record should be kept of authorised users and the access levels that apply to each user. The list of authorised users should be kept in the Trial Master File.

X. Datasets should be encrypted prior to transmission (if not using NHS to NHS e-mail or other secure method), or transfer (e.g. on CD), using an encryption software such as Truecrypt (https://www.truecrypt71a.com/).

XI. The study database should ideally have an audit trail for any changes made to electronic data after initial entry. For example, if a laboratory result is changed, the original value ought to remain in the database together with the date of the change, a brief explanation of why the change was made and who made the change (this may not be possible for if using Excel, however, any changes need to be reflected by the source data).

XII. At the end of a trial, data should be locked to prevent further additions or changes to the data using the data management system's file-locking facility if such a facility exists, otherwise a snapshot of the data should be taken for analysis and access to the data denied to anyone except the data manager.

3.3. DATABASE DESIGN

I. Creation of data fields

- A database should be designed to ensure that it captures all the information that is required according to the protocol. It should directly reflect the content of the Case Report Form (CRF).

II. Sheet identification

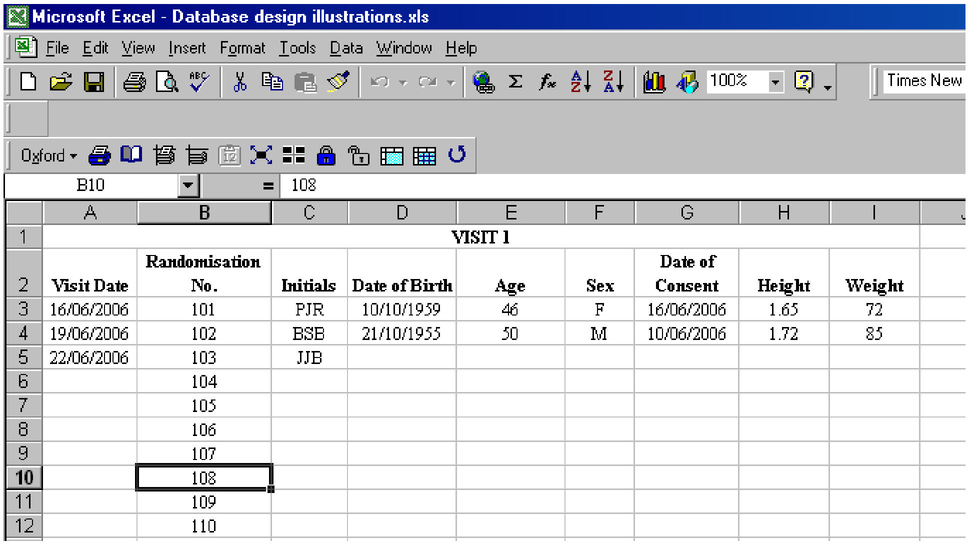
- Create sufficient sheets within the Excel file to cover each visit. Enter the name of the visit on the sheet tabs.
- Create additional sheets to cover adverse event collection, concomitant medications etc. as applicable to the study.

III. Data field identification

- Within each visit sheet, and through reference to the CRF, enter a heading for each data entry as it appears on the CRF visit page. Each column should represent one data point.

IV. Subject identification

- Each row should represent one subject.

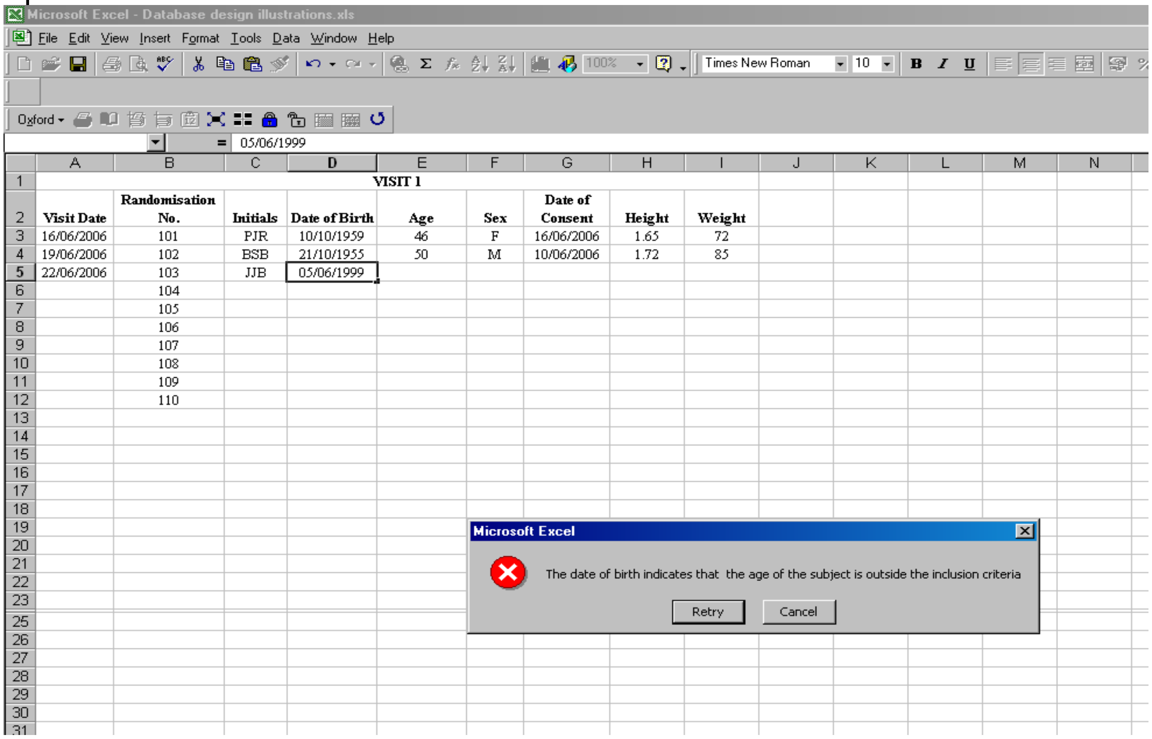

Microsoft Excel - Database design illustrations.xls
File Edit View Insert Format Tools Data Window Help
100%
Times New
Oxford
B10
= 108
A
B
C
D
E
F
G
H
I

	VISIT 1							
Visit Date	Randomisation No.	Initials	Date of Birth	Age	Sex	Date of Consent	Height	Weight
16/06/2006	101	PJR	10/10/1959	46	F	16/06/2006	1.65	72
19/06/2006	102	BSB	21/10/1955	50	M	10/06/2006	1.72	85
22/06/2006	103	JJB						
	104							
	105							
	106							
	107							
	108							
	109							
	110							

V. Optional advanced settings

- To ensure that data added is accurate; levels of validation can be added to the spreadsheet. Within certain fields, it is possible to restrict the type of data that can be added to a cell.
- For example, if the column on the spreadsheet represents date of birth, selecting the date option from the validation criteria will only allow dates to be entered into that column. Furthermore, if the inclusion criteria of the study have a restriction on age range, a start and end date can be added. If the value is outside the settings, an error alert can be generated.
- To set validation criteria, select the cells to be affected; choose Data, Validation and set your criteria under Input, Settings and Error Alert (Please note, this might slightly different dependent on the version of Office used).

Example



Microsoft Excel - Database design illustrations.xls

= 05/06/1999

	A	B	C	D	E	F	G	H	I
1				VISIT 1					
2	Visit Date	Randomisation No.	Initials	Date of Birth	Age	Sex	Date of Consent	Height	Weight
3	16/06/2006	101	PJR	10/10/1959	46	F	16/06/2006	1.65	72
4	19/06/2006	102	BSB	21/10/1955	50	M	10/06/2006	1.72	85
5	22/06/2006	103	JJB	05/06/1999					
6		104							
7		105							
8		106							
9		107							
10		108							
11		109							
12		110							

Microsoft Excel

The date of birth indicates that the age of the subject is outside the inclusion criteria

VI. Validation of Spreadsheet

- Once the spreadsheet design is believed to be complete, a validation should be undertaken to ensure that it is ‘fit for purpose’, as below.

VII. Create a sample CRF

- Enter data for a hypothetical participant, ensuring that all data boxes are complete. Include data entry on Adverse Events and Concomitant Medications pages.

VIII. Transfer data to database.

- Using the completed sample CRF, transfer all the data onto the database spreadsheet. Make a note of any difficulties involved in this process. Review each database page for omissions.

IX. Document validation

- Once the spreadsheet is considered to be ‘fit for purpose’, the sample CRF and printed spreadsheet should be filed as documentation of this process.

3.4. Creating and Using an Audit Trail

I. Preparing ‘Read Only’ access.

- Once development of the spreadsheet is complete, close the file. Right click on the file. Select ***Properties.*** Select ***General*** tab. Tick ‘Read Only’ box and select ‘OK’.

II. Creating the trail

- Once data entry commences, the file will require saving as a new spreadsheet on every occasion it is accessed. To save spreadsheets chronologically and record the identity of the member of the team adding to data, save with the date reversed, the time and initials. This method will automatically file the workbooks with the most recent version at the bottom of the list.

  Example: **100624.1630.CR**

  100629.1115.CR

  100629.1530.CR

III. Restricting Access

- Access should be restricted to the minimum number of personnel. A folder should be established within which all the trial spreadsheets should be stored. The Information Technology (IT) Department who should advise you how to restrict access as this depends on what type of access you require for which staff members.

3.5. Quality Checks

I. **Double and Single Data Entry with Control Checks**

- During data entry by trained staff, an average of 5% of errors is expected to occur. Two methods can be used to reduce the risk of errors: Double Data Entry or Single Data Entry with control checks.

  1) **Double Data Entry**
     This method involves two people entering the same CRF data onto the database independently of each other. Depending on the software used, the data may be entered twice onto the database on two separate files, which are

then compared by the system for accuracy. If the two entries do not match this would be flagged up by the database. Alternatively when the second data entry person enters the data, if it differs from that entered by the first person, a message immediately appears on screen and the original data can be checked. This method depends on the availability of a technically capable database.

2) **Single Data Entry with Control Checks**
This method may be more suitable for smaller single centre studies with less staff available for data entry and/or less sophisticated database software. Once the data has been entered, a visual check can be done between what is recorded on the paper CRF, and what was entered on screen.

II. **Data Cleaning and Validation**

- An integral part of the data management process is validation; to ensure the most accurate 'clean' set of data is provided for the statistical analysis. Data validation can be carried out at three stages during the trial:

a. When CRFs are completed by the investigator
To improve accuracy at this stage all staff completing CRFs should be sufficiently trained in their completion. Validation should also be carried out as part of the on-going monitoring of the study; either by members of the research team or by independent monitors. Validation via monitoring is done through Source Data Verification (SDV). SDV involves checking the data entered into the CRFs against that in the original source records e.g. patient's hospital files for accuracy.

b. When data are entered in the database by data entry staff
During data entry the two methods for validation described above (6.34) can be used i.e. data entry checks or double data entry. Where data entry checks are used, if the study database has software enabled for automatic data entry checks, an Edit Check Specification (ECS) document should be put together by the clinicians/statisticians/data staff involved with the study. The ECS should provide full details of the data entry checks that have been set up, and all checks should be tested before the trial begins.

Depending on the database software it is also advisable to set up warnings to alert data entry staff when values are entered outside of the expected range, or if the type of value entered is incorrect e.g. a numeric value entered rather than text. It is also useful to set up alerts for missing values where possible.

c. When data have been entered and are available for the data manager
At this stage it is advisable to carry out systematic post-entry computer tests. Lists should then be created (either through automatic database software system, or manually) of the following data queries:

- All missing values will be listed
- All values outside of pre-defined range

Logical checks should also be performed to ensure consistent reporting between relevant fields and that there are no implausible difference between fields e.g. male and pregnant.

All checks should be defined before the study starts, and should be described in the Edit Check Specification document described previously. Data validation should continue until all missing values and inconsistencies are corrected or clarified.

4 Document Ratification Process

The review period for this document is set as ***default of three*** years from the date it was last ratified, or earlier if developments within or external to the Trust indicate the need for a significant revision to the procedures described.

This document will be approved by the ***RD&I Manager or their Deputy***.

Non-significant amendments to this document may be made, under delegated authority from ***a Senior RD&I manager***, by the nominated author. These must be ratified by ***a Senior RD&I manager***.

Significant reviews and revisions to this document will include a consultation with ***appropriately knowledgeable staff***. For non-significant amendments, informal consultation will be restricted to ***staff*** who are directly affected by the proposed changes.

Dissemination and implementation

4.1. Dissemination of this SOP

4.1.1. New SOPs and new versions of existing SOPs: The Research Governance Manager will be responsible for ensuring authorised SOPs are uploaded on the RD&I intranet site. Internal Trust Staff are expected to use the RD&I intranet site to access latest versions of SOPs and to check the website regularly for updates.

Notice of new or amended procedural documents that have undergone a major amendment will be given *via* the following routes:

- Inclusion in the Trust weekly e-bulletin Vital Signs
- Direct email to Trust Researchers and or teams

4.2. Training in this SOP

4.2.1. All staff whose activities are subject to this SOP should ensure that they read and understand the content of the SOP.

5 Reference material

Princeton University. Excel Working with Data Lists. Office of Information Technology.
http://web.princeton.edu/sites/oitdocs/Documentation/Training/Excel%20Working%20with%20Data%20Lists.pdf
UK Policy Framework for Health and Social Care Research (2017) https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/uk-policy-framework-health-social-care-research/

ICH Harmonised Tripartite Guideline for Good Clinical Practice.
http://www.ich.org

Data Protection Act 2018
http://www.legislation.gov.uk/ukpga/1998/29/contents

Caldicott Principles
https://www.plymouthhospitals.nhs.uk/rd-information-governance

The University of Oxford CTRG
http://www.admin.ox.ac.uk/researchsupport/ctrg/resources/

6 Amendment History

Version Number:	2.1
Date Of Amendment:	Jan 2019
Details Of Amendment:	Updated Trust and Dept. name. Reduce signature requirement to single senior RD&I Manager. Updated references to the Data Protection Act 2018 and the UK Policy Framework for Health and Social Care Research (2017).

Version Number:	2.0
Date Of Amendment:	Aug 2017
Details Of Amendment:	Updated SOP template and numbering system. SOP reviewed and updated.

Version Number:	1.1 (minor amendment)
Date of Amendment:	Mar 2012
Details of Amendment:	Cover page - Change of SOP location address.