Illawarra Health and Medical Research Institute | Faculty of Science, Medicine and Health

2018

Exploring Data Quality Management within Clinical Trials

Lauren Houston
University of Wollongong, lah993@uowmail.edu.au

Yasmine Probst
University of Wollongong, yasmine@uow.edu.au

Ping Yu
University of Wollongong, ping@uow.edu.au

Allison Martin
University of Wollongong, allisonh@uow.edu.au

Exploring Data Quality Management within Clinical Trials

Abstract

Background Clinical trials are an important research method for improving medical knowledge and patient care. Multiple international and national guidelines stipulate the need for data quality and assurance. Many strategies and interventions are developed to reduce error in trials, including standard operating procedures, personnel training, data monitoring, and design of case report forms. However, guidelines are nonspecific in the nature and extent of necessary methods.

Objective This article gathers information about current data quality tools and procedures used within Australian clinical trial sites, with the aim to develop standard data quality monitoring procedures to ensure data integrity.

Methods Relevant information about data quality management methods and procedures, error levels, data monitoring, staff training, and development were collected. Staff members from 142 clinical trials listed on the National Health and Medical Research Council (NHMRC) clinical trials Web site were invited to complete a short self-reported semiquantitative anonymous online survey.

Results Twenty (14%) clinical trials completed the survey. Results from the survey indicate that procedures to ensure data quality varies among clinical trial sites. Centralized monitoring (65%) was the most common procedure to ensure high-quality data. Ten (50%) trials reported having a data management plan in place and two sites utilized an error acceptance level to minimize discrepancy, set at <5% and 5 to 10%, respectively. The quantity of data variables checked (10–100%), the frequency of visits (once-a-month to annually), and types of variables (100%, critical data or critical and noncritical data audits) for data monitoring varied among respondents. The average time spent on staff training per person was 11.58 hours over a 12-month period and the type of training was diverse.

Conclusion Clinical trial sites are implementing ad hoc methods pragmatically to ensure data quality. Findings highlight the necessity for further research into "standard practice" focusing on developing and implementing publicly available data quality monitoring procedures.

Research Article

Exploring data quality management within clinical trials

Corresponding author
Miss Lauren Houston
228/41 Northfields Avenue
School of Medicine, Faculty of Science, Medicine and health,
University of Wollongong, Wollongong NSW 2522, Australia

E: lah993@uowmail.edu.au
M: +61 430 371 929

Author and co-authors

Ms Lauren Houston[1,2], Dr Yasmine Probst[1,2], A/Prof Ping Yu[3] &
Dr Allison Martin[1,2].

[1]School of Medicine, Faculty of Science, Medicine and Health, University of
Wollongong, Wollongong NSW 2522, Australia.

[2]Illawarra Health and Medical research institute, University of Wollongong, Wollongong
NSW 2522, Australia,

[3]School of Computing and Information Technology, Faculty of Engineering and
Information Sciences, University of Wollongong, Wollongong NSW 2522, Australia.

Abstract

Background: Clinical trials are an important research method for improving medical knowledge and patient care. Multiple international and national guidelines stipulate the need for data quality and assurance. Many strategies and interventions are developed to reduce error in trials, including standard operating procedures, personnel training, data monitoring and design of case report forms. However, guidelines are non-specific in the nature and extent of necessary methods. **Objective:** To gather information about current data quality tools and procedures used within Australian clinical trial sites, with the aim to develop standard data quality monitoring procedures to ensure data integrity. **Methods:** Relevant information about data quality management methods and procedures, error levels, data monitoring, staff training, and development were collected. Staff members from 142 clinical trials listed on the National Health and Medical Research Council (NHMRC) clinical trials website were invited to complete a short self-reported semi-quantitative anonymous online survey. **Results:** Twenty (14%) clinical trials completed the survey. Results from the survey indicate that procedures to ensure data quality varies among clinical trial sites. Centralized monitoring (65%) was the most common procedure to ensure high-quality data. Ten (50%) trials reported having a data management plan in place and two sites utilized an error acceptance level to minimize discrepancy, set at <5% and 5-10%, respectively. The quantity of data variables checked (10-100%), the frequency of visits (once-a-month to annually), and types of variables (100%, critical data or critical and non-critical data audits) for data monitoring varied among respondents. The average time spent on staff training per person was 11.58 hours over a 12-month period and the type of training was diverse. **Conclusion:** Clinical trial sites are implementing ad-hoc methods pragmatically to ensure data quality. Findings highlight the necessity for further research into "standard practice" focusing on developing and implementing publically available data quality monitoring procedures.

Keywords

1. Background and Significance

Clinical trials are an important research method for improving medical knowledge and patient care. Evidence has linked poor data quality to incorrect conclusions and recommendations [1-4]; therefore, data quality is of paramount importance for acquiring reliable research findings from clinical trials [5, 6]. As poor data quality may stem from error; therefore, preventing data error is just as important as the development, design, and collection of clinical trial data [7]. Assessment of all possible sources of error, including data recording, abstraction, transcription, entry, coding and/or cleaning processes, contributes to improving data quality for clinical trials [8].

Many strategies and interventions have been developed aiming at reducing error in clinical trials, including standard operating procedures (SOPs), personnel training, data monitoring and design of case report forms (CRF). Additionally, multiple international and national guidelines stipulate the need for data quality and assurance, however, they are non-specific in the nature and extent of the necessary methods. These guidelines include the International Conference on Harmonisation Good Clinical Practice (ICHGCP) guideline E6 (1996)[9] updated in 2015, guideline E6(R2)[10]; the International Standards Organisation (ISO) Clinical Investigation of Medical Devices for Human Subjects – GCP (2011) (ISO 14155:2011)[11]; the Food and Drug Administration (FDA) Guidelines for Monitoring of Clinical Investigations (1998) updated in 2013[12]; the European Union's Clinical Trial Directive 2001/20/EC (2001)[13], updated in 2009 and the Australian Government's National Statement on Ethical Conduct in Human Research, (1999), updated, in 2007[14].

According to ICHGCP, clinical trials must protect the rights and safety of all patients and ensure trial results are legible and valid. On-site monitoring is important to achieve high data quality and to ensure the method of source data verification[1] (SDV) fulfils the original objectives [15].To optimize the efficacy of monitoring there has been an emphasis on data audits and reducing on-site monitoring [10, 11]. However, the ICHGCP guidelines are flexible in interpretation and do not provide specific details on how and when to conduct audits, or how much or how little monitoring is required to maintain data integrity [16, 17]. There is a lack of evidence to support intensive monitoring for data quality, in fact, updated guidelines promote alternative monitoring methods, such as risk-based approaches[2], centralized[3] and remote monitoring[4], that complement trial procedures by improving the use of resources available [12, 18-20]. Although a reduction in on-site monitoring is suggested (updated ICHGCP guidelines re [21, 22]), the majority of clinical trials continue to conduct traditional 100% SDV[23]. Lack of clear guidance on which monitoring method is valid and cost-effective to ensure data integrity creates confusion within the clinical research community. A reduction in on-site SDV and the risk of missing critical issues are tradeoffs that more efficient, modern monitoring approaches need to consider.

---

[1] Source data verification (SDV) defines the process of comparing data collected on original source documents to data recorded on a case report form (CRF) either on paper or electronic records

[2] Risk-based approach includes a mixed method approach focused on the critical data points an processes that are identified to have the most risk via a targeted or triggered assessment.

[3] Centralized monitoring is a remote evaluation carried out by the sponsor personnel or representatives at a location other than the sites at which the clinical investigation is being carried out on a real time basis.

[4] Remote monitoring is off-site monitoring of activities previously conducted on-site. Documents are delivered via email, fax or snail mail to a clinical research associate (CRA) to conduct SDV and satisfy queries.

Due to a growing concern about the effectiveness and efficiency of monitoring procedures, Brosteanu et al. (2017) conducted a clustered randomised study comparing intensive on-site monitoring and risk-adapted monitoring. Results found the benefit of intensive on-site SDV to be small (8.2%) when compared to risk-based monitoring, which utilized less than 50% of resources whilst ensuring the same level of GCP compliance [18]. It is evident that standardized approach needs to be adopted for monitoring of data quality in clinical research. In support of this notion, a survey conducted by the Clinical Trials Transformation Initiative (CTTI) investigated the intensity, focus, and methodology of monitoring practices by clinical research sponsors over a range of trial settings [24]. It was found that there is heterogeneity within and between organizational types including academic/government, clinical research organizations, and industry.

Evidence of the effect of non-standardized data quality checks within clinical trials is the online blog of publication retractions due to fraudulent data [25]. This website lists 484 publications in 2017 alone that had to be withdrawn due to incorrect data/analysis. There appears to be a lack of knowledge about systematic methods and procedures for data quality assessment in clinical trials [1].To ensure data integrity in clinical research it is imperative to introduce a 'gold standard methodology' so that manuscripts can be published referencing their methods employed and the broader research community can be reassured the data was valid.

2. Objectives

The objective of this feasibility study was to gather information about current data quality tools and procedures used within Australian clinical trial sites, with the aim to develop a standard data quality monitoring procedures to ensure data integrity.

3. Methods

3.1. Setting

Clinical trial sites listed on the Australian Government National Health and Medical Research Council (NHMRC) Australian Clinical Trial site list were invited to participate in this survey (n=148) [5, 26]. Employees targeted to complete the survey included the manager/employee in charge of trial-related data quality assurance processes. The NHMRC clinical trial list was determined as a representative sample of Australian clinical trials including all phases (I-IV) and types (treatment, diagnostic/screening, and prevention) of clinical trials.

Clinical trial sites that were identified to have an affiliation with the University of Wollongong (UoW), the organization that the researchers were employed, were excluded from the study to avoid the potential risk of bias. Any overlapped sites which, may cause duplication, were also excluded. Several clinical trial networks on the list responded that they did not run clinical trials independently. In this case, permission was given for these networks to forward the survey to their collaborating organizations that run clinical trials. Informed tacit consent was obtained by completion and return of the online questionnaire survey form.

3.2. Development of the online questionnaire survey form

Eleven survey questions were adapted in short form from the published and validated survey questions [24, 27]. Information gathered from the survey included data quality management methods and procedures, error levels, data monitoring, staff training, and development (see Appendix A).

Construct validation was completed by a convenience sample of ten UoW researchers who reviewed survey questions to ensure that the intended concept was assessed. Participants were also asked to comment on any procedural, usability and transparency issues faced in completing the survey. Expert advice from a data management manager was sought to ensure content validity, question clarity and answers fully address the research questions. The online questionnaire survey was designed using the Research Electronic Data Capture (REDCap) tool hosted at the University of Wollongong [28].

3.3. Survey administration

A cross-sectional study design was applied to get an overview of the current quality tools and practices implemented in Australian clinical trials. Invitations for participation were sent to the identified contact person for each clinical trial site via email. The email contained a brief introduction and a direct link to the survey. Each clinical trial site was provided with an individual identification code and three email reminders were sent over a four-month period to non-respondents. Clinical trial sites were asked to forward the survey to their collaborating sites using the same individual identification code if they were identified to be clinical institute network.

3.4. Data analysis

Questionnaire responses were standardized into categorical options and numerically coded for analysis. Statistical analyses were conducted using IMB SPSS software (Version 22, IMB Australia, Lane Cove, NSW, Australia). Data was explored via descriptive statistical analyses. Free text responses were analyzed using the six phases of thematic analysis [29] and conducted using NVivo qualitative data analysis Software (QRS International Pty Ltd. Version 10, 2012). Ethics approval was obtained from the University of Wollongong Human Research Ethics Committee (HE16/131).

4. Results

Of the 148 clinical trial sites identified, 142 initial invitations were sent out, see Figure 1. A total of 34 clinical trial sites consented to participate in the online survey, yielding a response rate of 24%. Of the 34 responses, 14 were excluded from the analyses due to missing data for >25% of survey questions. Three clinical networks asked to forward the survey invitation email to respondents that were more appropriate as well as to their collaborating sites. Finally, 20 clinical trial sites completed the survey in full and were included in data analysis. Table 1 listed the number and types of clinical trial sites. At each site, more than one type of clinical trial was conducted at any point in time.

4.1. Data management and monitoring

At the time of the survey, ten sites (50%) reported having a clinical data management plan in place and the majority (n=19) had implemented at least one or more procedures to ensure data quality (see Table 2). Only two sites set an error acceptance

level, <5% and 5-10%, respectively, both responding that no follow-up or further monitoring was conducted if the error rate was found to be higher than the error acceptance level.

The structure of data monitoring was reported in terms of variables to be selected, its coverage and amount, and time of execution. Monitoring 100% of the data points was the most common (n=7) response, although the procedures implemented varied greatly, and the amount of data included in monitoring ranging from 10-100%. The timing of data monitoring varied and was specific to the clinical trial and study design. The frequency of data monitoring varied among the six sites (30%) from monthly to annually. The variables included in data monitoring were completed on all (100%) data points (n=5), only critical data points (n=1), critical and non-critical data points defined by each study (n=3) or were dependent on the clinical trial (n=3). When asked about how their data was monitored, seven sites (35%) reported that they implemented at least one or more sampling techniques to extract data points, seven sites (35%) did not know and one site (5%) did not implement sampling techniques at all, see table 2 for further details.

4.2. Training and development

A certain type of staff training and development devoted to data quality were conducted at all clinical trial sites (100%, see Table 3). The average amount of time spent on staff training and development per person, per clinical trial was 11.58 ± 9.01 hours, (range 2 to 30) over a 12-month period.

The personnel responsible for reviewing the reports of data quality and consistency varied from chief investigators (65%), auditor/monitor (60%), the data

manager (55%) and sponsor (50%). In total, 75% of respondents answered that more than one person reviewed the reports.

5. Discussion

This feasibility study highlights the heterogeneity of data quality management practices within Australian clinical trials. Only 50% of the respondent clinical trial sites currently had a clinical data management plan in place, confirming our proposition that developing and maintaining a data management system is a challenge for clinical trials [30]. This is also in accordance with a recently published survey (26). This survey reported considerable variation in data management, with over 50% of clinical research centres having a data management system but many did not comply with guidelines and legal requirements (GCP and FDA)[27]. There are many reasons for this, such as individual clinical trials implementing different procedures dictated by the sponsor, or monetary constraints in academic versus pharmaceutical clinical trials [31].

Centralised and remote monitoring were found to be the most common data monitoring methods utilized although there appeared to be a lack of credible literature to suggest the advantage of these 'newer' methods over the more traditional approaches [20]. This study identified that 50% of sites still use traditional data monitoring methods such as 100% on-site SDV, which is an expensive, labor-intensive activity [32] and does not guarantee error-free results [33]. Andersen et al. [34] compared the effect of partial SDV and traditional 100% SDV using post hoc analyses of three-phase III randomised control trials. Because completing traditional 100% SDV monitoring only reduced error

marginally (0.26%) compared to partial SDV, the authors challenged the belief that a 0%

error rate is not an achievable goal. Only 2 out of 20 trial sites in our survey reported

having a set error acceptance level, being ≤10%, which is in line with published literature

[8, 35, 36]. One of the two clinical trial sites stated that they implemented a 5-10%

threshold range, however, no further comment on why or when a different threshold for

data validation was adhered to. As the survey was administered to clinical trial sites, the

researchers have assumed that it might be possible that different clinical trial types have a

tighter threshold than others; for example, a 5% threshold for a phase IV trials compared

to 10% for epidemiological trials. Future research is required to explore the rationale for

different levels of error acceptance within clinical trials.

The major quality assurance activity reported to 'prevent' data errors was regular

education and training of data collectors throughout the clinical trial. Although the

majority of survey respondents reported that staff training and development was

undertaken, the amount of training time varied greatly. Many researchers receive little to

no training in regards to best practice for attaining, evaluating and controlling the quality

of data collected. This is in line with the literature that reported due to the limitation of

time and resources, not all research trials implement all the necessary data quality

management tools and procedures [37].

Within the pharmaceutical/private industries [38] and information sciences

literature, data quality tools and procedures are well developed in which many

frameworks acknowledge the multiple dimensions of data quality [39-44]. However, only

a small body of clinical and health researchers have described the use of data quality

frameworks [38, 45-48], and fewer have identified appropriate methods to quantify the

quality of data [8]. Although many data quality dimensions and attributes have been determined within the clinical and health literature, the majority provides no usable definitions. Public sharing of this knowledge is crucial in developing a standardized approach that can be implemented across the clinical and broader research community to improve the rigor of clinical trials.

5.1 Study limitations

The results of this feasibility study are limited to clinical trials listed on the Australian NHMRC clinical trial site list. The survey results are subject to potential bias in a positive direction as the employee who completed the survey may be more knowledgeable about their organization's data quality management procedures than those who chose to not participate. As clinical trial sites were recruited as an organization it is impossible for us to track if the person who completed the survey was best positioned to do so in the organization. In addition, the tools and procedures differed among clinical trials, which were influenced by overarching site policies. All these cause difficulty in interpreting results. The reason for low response rate might be clinical data audits and data management procedures are usually considered highly confidential by many research organizations and kept in-house [49]. As it is impossible to collect data about how many forwarding e-mails were sent by the clinical trial sites listed on the NHMRC website, this data was not accounted for in calculating a response rate. The low response rate means that the results of this survey should be used with caution. They may not be generalizable as a representative sample of clinical trial sites. Future research should include qualitative analysis through key-informant interviews with the provision of SOPs. This feasibility study was not designed to assess which data management tool is more effective, but

rather to gather information about current data quality tools and procedures used within Australian clinical trials. Further research is required to fully examine the best method of monitoring data quality to assure and control data integrity in clinical research.research.

At the conclusion of three studies, OPTI-misation of MONitoring (OPTIMON) [50], Strategic Timing of AntiRetrovial Treatment (START) trial Monitoring Substudy [51] and TargetEd Monitoring: Prospective Evaluation and refinement (TEMPER) study [52] the scientific community will have a better understanding of effective monitoring strategies. At the conclusion of all three studies, empirical evidence will be provided and aid in improving the currently limited published procedures, as all three studies have different aims and designs in developing audit methodology.

This research is part of a collaborative project and an important opportunity for clinical trial sites to bring together their existing experience to improve data quality management systems. It is recommended to further compare what is described in the literature and what is currently happening at a site level to identify the gaps, the facilitators and the barriers to implementing data quality management systems. To achieve the objective of informing the data quality improvement initiative over a broad spectrum of clinical trials, this critical information needs to be made freely available in the published literature.

6. Conclusions

This is the first survey gathering information about current data quality tools and procedures within Australian clinical trial sites. This survey found that clinical trial sites

were implementing newer approaches such as centralized and remote monitoring despite the majority were still completing 100% SDV, a labor-intensive and cost-inefficient method. It is clear that data quality management procedures vary greatly between clinical trials sites, with only 50% of the trial sites with a data management plan in place. Further research is required to assess differences between data management tools and procedures between clinical trials within a clinical trial site. This will allow researchers to investigate what is "standard practice" and focus on developing and implementing publically available data quality monitoring procedures to ensure data integrity. Data quality is essential for the reliability of scientific findings generated from the investment in clinical trials, adequate infrastructure, staff skills, management support and resources need to be in place to ensure data is effectively managed. It is time that quality assurance and quality control tools and procedures implemented in clinical trials are cited in all publications.

Clinical Relevance Statement

It is vital to ensure the scientific rigor of clinical trials to evaluate data quality management procedures and to assure the accuracy of findings and to reduce error. This survey highlights the heterogeneity of clinical trial's data quality management practices in Australia. It thus suggests that 'best practices' need to be made freely available in the published literature. Adequate infrastructure, staff skills, management support and resources need to be considered and implemented to ensure high quality data and research are produced and managed effectively.

Multiple Choice Question

Question

1. Which of the following data monitoring methods is the traditional approach that fulfills the International Conference on Harmonisation Good Clinical Practice (ICH-GCP) E6 guideline (1996)?

    a. Remote monitoring

    b. Source data verification

    c. Central monitoring

    d. Risk-based approach

Answer – B: Source data verification

Explanation: In 1996, the International Conference on Harmonisation (ICH) guideline E6 on Good Clinical Practice (GCP) reported there is a need for on-site monitoring before, during and after a clinical trial. The practice of source data verification (SDV) fulfils the ICH-GCP requirements and is a process of comparing data collected on original source documents to data recorded on a case report form (CRF) or electronic record. Source documents are considered the "gold standard" from which data is obtained in clinical trials. Therefore, SDV is considered the traditional approach to monitoring data utilized by clinical trial auditors.

Conflict of Interest

The author(s) declare that they have no conflict of interests.

Declarations/Author contributions

Acknowledgements

Protection of Human subjects
The study was performed in compliance with the World Medical Association Declaration of Helsinki on Ethical Principles for Medical Research Involving Human Subjects and was approved by the University of Wollongong Human Research Ethics Committee (HE16/131).

References

1. Chen H, Hailey D, Wang N, Yu P. A review of data quality assessment methods for public health information systems. International Journal Of Environmental Research And Public Health. 2014;11(5):5170-207.
2. Hasan S, Padman R. Analyzing the effect of data quality on the accuracy of clinical decision support systems: a computer simulation approach. AMIA Annual Symposium Proceedings / AMIA Symposium AMIA Symposium. 2006:324-8.
3. Huneycutt BJ, Illes J, Boone C, Jackson K, Woolett G, Health A. In the Patient's Interest: Improving Access to Clinical Trial Data. 2014 [cited 06 July 2016]. Available from: http://www.fdanews.com/ext/resources/files/09/09-03-2014-Avalerereport.pdf.
4. Kahn MG, Brown JS, Chun AT, Davidson BN, Meeker D, Ryan PB, et al. Transparent reporting of data quality in distributed data networks. EGEMS (Washington, DC). 2015;3(1):1052-.
5. Bhatt A. Quality of clinical trials: A moving target. Perspectives in Clinical Research. 2011;2(4):124-8.
6. Bowman S. Impact of Electronic Health Record Systems on Information Integrity: Quality and Safety Implications. Perspectives in Health Information Management. 2013:1.
7. Society for Clinical Data Management Good Clinical Data Management Practices. 2013.
8. Nahm ML, Pieper CF, Cunningham MM. Quantifying Data Quality for Clinical Trials Using Electronic Data Capture. PLoS ONE. 2008;3(8):1-8.
9. International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use 1996, ICH Harmonized Tripartite Guidelines: Guideline for Good Clinical Practice E6(R1) 1996 [Accessed 12 October 2014]. Available from: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R1_Guideline.pdf.
10. International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use 2015, E6(R2) Good Clinical Practice 2015 [Accessed 30 June 2016]. Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM464506.pdf.
11. International Standards Organisation, Clinical investigation of medical devices for human subjects - Good clinical practice ISO14155:2011. 2011. Available from: https://www.iso.org/obp/ui/#iso:std:iso:14155:ed-2:v1:en.
12. Administration FaD. FDA, Guidance for Industry, oversight of clinical investigations - a risk-based approach to monitoring 2013 [Accessed 06 July 2016]. Available from: http://www.fda.gov/downloads/Drugs/.../Guidances/UCM269919.pdf.
13. European Commisson, Clinical trials - Directive 2001/20/EC
Public Health Medical Products for Human Use.2001. Available from: https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-1/dir_2001_20/dir_2001_20_en.pdf.
14. Australian Government National Health and Medical Research Council (NHMRC), National Statement on Ethical Conduct in Human Reserach 2007.
15. Bakobaki JM, Rauchenberger M, Joffe N, McCormack S, Stenning S, Meredith S. The potential for central monitoring techniques to replace on-site monitoring: Findings from an international multi-centre clinical trial. Clin Trials. 2012;9(2):257-64.
16. Sourabh D. Hybrid approaches to clinical trial monitoring: Practical alternatives to 100% source data verification. Perspectives in Clinical Research, Vol 2, Iss 3, Pp 100-104 (2011). 2011(3):100.
17. Macefield RC, Beswick AD, Blazeby JM, Lane JA. A systematic review of on-site monitoring methods for health-care randomised controlled trials. Clin Trials. 2013;10(1):104-24.
18. Brosteanu O, Schwarz G, Houben P, Paulus U, Strenge-Hesse A, Zettelmeyer U, et al. Risk-adapted monitoring is not inferior to extensive on-site monitoring: Results of the ADAMON cluster-randomised study. Clin Trials. 2017:1740774517724165.
19. Tantsyura V, Dunn IM, Waters J, Fendt K, Kim YJ, Viola D, et al. Extended Risk-Based Monitoring Model, On-Demand Query-Driven Source Data Verification, and Their Economic Impact on Clinical Trial Operations. Therapeutic Innovation & Regulatory Science. 2015;50(1):115-22.
20. Olsen R, Bihlet AR, Kalakou F, Andersen JR. The impact of clinical trial monitoring approaches on data integrity and cost - A review of current literature. European Journal of Clinical Pharmacology. 2016;72(4):399-412.
21. Mealer M, Kittelson J, Thompson BT, Wheeler AP, Magee JC, Sokol RJ, et al. Remote source document verification in two national clinical trials networks: A pilot study. PLoS ONE. 2013;8(12):E81890-.
22. Tudur Smith C, Stocken DD, Dunn J, Cox T, Ghaneh P, Cunningham D, et al. The Value of Source Data Verification in a Cancer Clinical Trial. PLoS ONE. 2012;7(12).
23. Journot V, Pignon JP, Gaultier C, Daurat V, Bouxin-Métro A, Giraudeau B, et al. Validation of a risk-assessment scale and a risk-adapted monitoring plan for academic clinical research studies - The Pre-Optimon study. Contemp Clin Trials. 2011;32(1):16-24.
24. Morrison BW, Cochran CJ, White JG, Harley J, Kleppinger CF, Liu A, et al. Monitoring the quality of conduct of clinical trials: a survey of current practices. Clin Trials. 2011;8(3):342-9.
25. Center for Scientific Integrity, Retraction Watch 2015 [cited 2017 13 Oct 2017]. Available from: http://retractionwatch.com/.
26. Australian Government, National Health and Medical Research Council (NHMRC), Australia Clinical Trials, Trial Sites. Department of Industry Innovation and Science.2016 [cited 10 Aug 2016]. Available from: https://www.australianclinicaltrials.gov.au/trial-websites.
27. Kuchinke W, Ohmann C, Yang Q, Salas N, Lauritsen J, Gueyffier F, et al. Heterogeneity prevails: the state of clinical trial data management in Europe - results of a survey of ECRIN centres. Trials. 2010(1):79.
28. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42(2):337-81.
29. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology. 2006;3(2):77-101.
30. Ohmann C, Kuchinke W, Canham S, Lauritsen J, Salas N, Schade-Brittinger C, et al. Standard requirements for GCP-compliant data management in multinational clinical trials. Trials. 2011;12(1):85-.
31. Krishnankutty B, Bellary S, Kumar BRN, Moodahadu LS. Data management in clinical research: An overview. Indian journal of pharmacology. 2012;44(2):168-72.
32. Califf RM, Karnash SL, Woodlief LH. Developing systems for cost-effective auditing of clinical trials. CONTROL CLIN TRIALS. 1997;18(6):651-60.

33. Eisenstein EL, Lemons Ii PW, Tardiff BE, Schulman KA, Jolly MK, Califf RM. Reducing the costs of phase III cardiovascular clinical trials. Am Heart J. 2005;149(3):482-8.
34. Andersen JR, Byrjalsen I, Bihlet A, Kalakou F, Hoeck HC, Hansen G, et al. Impact of source data verification on data quality in clinical trials: an empirical post hoc analysis of three phase 3 randomized clinical trials. British Journal Of Clinical Pharmacology. 2015;79(4):660-8.
35. Clarke D, Breen L, Jacobs M, Franklin R, Tobota Z, Maruszewski B, et al. Verification of data in congenital cardiac surgery. Cardiology in the Young. 2008;18(S2):177-87.
36. Houston L, Probst Y, Humphries A. Measuring Data Quality Through a Source Data Verification Audit in a Clinical Research Setting. Studies In Health Technology And Informatics. 2015;214:107-13.
37. Kleppinger CF, Ball LK. Building quality in clinical trials with use of a quality system approach. Cliniacl Infectious Diseases. 2010;51(Suppl 1):S111-6.
38. Arts DGT, De Keizer NF, Scheffer GJ. Defining and improving data quality in medical registries: A literature review, case study, and generic framework. J Am Med Informatics Assoc. 2002;9(6):600-11.
39. Scannapieco M, Catarci T. Data quality under the computer science perspectve. Ital J Arch Comput. 2002:1-12.
40. Batini C, Cappiello C, Francalanci C, Maurino A. Methodologies for data quality assessment and improvement. ACM computing surveys (CSUR). 2009;41(3):16.
41. Wang RY, Storey VC, Firth CP. A framework for analysis of data quality research. IEEE transactions on knowledge and data engineering. 1995;7(4):623-40.
42. Tayi GK, Ballou DP. Examining data quality. Communications of the ACM. 1998;41(2):54-7.
43. Pipino LL, Lee YW, Wang RY. Data quality assessment. Communications of the ACM. 2002;45(4):211-8.
44. Maydanchik A. Data quality assessment: Technics publications; 2007.
45. Davidson B, Lee YW, Wang R. Developing data production maps: meeting patient discharge data submission requirements. Int J Healthcare Technology and Management. 2004;6(2):223-40.
46. Riain C, Helfert M. An evaluation of data quallity related problem patterns in healthcare information systems. IADIS virtual multi conference on computer science and information systems; Lisbon, Portugal2005.
47. Mettler T, Rohner P, Baacke L. Improving Data Quality of Health Information Systems: A Holistic Design- Oriented Approach. European Conference on Information Systems (ECIS) proceedings Paper 1642008.
48. Kahn MG, Raebel MA, Glanz JM, Steiner JF, Riedlinger K. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. Medical Care. 2012;50(SUPPL. 1):S21-S9.
49. Rostami R, Nahm M, Pieper CF. What can we learn from a decade of database audits? The Duke Clinical Research Institute experience, 1997--2006. Clinical Trials (London, England). 2009;6(2):141-50.
50. Journot V, Pignon J-P, Gaultier C, Daurat V, Bouxin-Métro A, Giraudeau B, et al. Validation of a risk-assessment scale and a risk-adapted monitoring plan for academic clinical research studies — The Pre-Optimon study. Contemp Clin Trials. 2011;32(1):16-24.
51. Hullsiek KH, Kagan JM, Engen N, Grarup J, Hudson F, Denning ET, et al. Investigating the efficacy of clinical trial monitoring strategies: Design and Implementation of the Cluster Randomized START Monitoring Substudy. Therapeutic innovation & regulatory science. 2015;49(2):225-33.
52. Stenning S, Joffe N, Batra P, Jones C, Montana CD, Tappenden N, et al. Update on the temper study: targeted monitoring, prospective evaluation and refinement. Trials. 2013;14(Suppl 1):P138-P.

Appendix A: Online survey questions

1. Does your institute currently have a clinical data management plan in place?

2. What type of clinical research does your institute conduct?

3. Does your institute have any of the following procedures in place to ensure high-quality data is produced?

4. What percentage of your data is monitored?

5. Does your institute have an error acceptance level?

6. Following on from Question 5, if the error rate is found to be higher than the approved acceptance level, does your institute implement further follow-up monitoring?

7. How often does your institute conduct internal data monitoring?

8. What variables are included in data monitoring?

9. Does your institute use any of the following sampling techniques to select what data points are monitored?

10. Please specify the type of staff training/development you conduct that is devoted to data quality for clinical trials.

11. Who reviews the reports of data quality and consistency?

Table 1: Demographic characteristics of clinical trial site employees

Variable	n(%)
Gender	
Male	0(0)
Female	20(100)
Highest level of education	
College/TAFE course	1(5)
Bachelor degree	12(60)
Doctoral degree	7 (35)
Duration of current employment (years)	
0 – 4	9(45)
5 – 9	7(35)
10 – 15	3(15)
+ 15	1(5)
Appointment (current job or position)	
Continuing employment (no specified end date)	9(45)
Fixed-term contract (specified time or ascertainable period) (years)	11(55)
< 1	6(45)
2	2(18)
>3	3(27)
Current job title (n=19)[1]	
Research Fellow	2(10)
Research Governance Officer	1(5)
Manager	10(50)
Clinical Trial/Study Coordinator	4(20)
Chief Operating Officer	1(5)
Chief Executive Director	1(5)

---

[1] One clinical trial employee did not enter the current job title.

Table 2: Type of clinical trial conducted at each site

Types of Clinical Trial	Number of Trial Sites
Treatment	16
Phase II	14
Phase III	14
Quality of life	12
Prevention	11
Epidemiology	11
Phase 1	10
Phase IV	10
Screening	9
Diagnosis	7
Genetic screening	6

Table 3: An outline of survey responses

	Yes n(%)	No n(%)	Don't know n(%)	Not applicable n(%)
Currently have a clinical data management plan	10(50)	7(35)	3(15)	-
Procedures to ensure high-quality data	19(95)	-	1(5)	-
Centralized monitoring	13(65)			
Remote monitoring	11(55)			
Logic, range and consistency checks	11(55)			
On-site source data verification	10(50)			
Statistical techniques	7(35)			
Risk-based targeted monitoring	5(25)			
Risk-based triggered monitoring	3(15)			
Currently have an error acceptance level	2(10)	11(55)	7(35)	-
Percent of data monitored	14(70)	-	5(25)	1(5)
10%	2(10)			
20%	2(10)			
75%	2(10)			
100%	7(35)			
Amount of data monitored depends on data point/outcome measured	1(5)			
Frequency of internal data monitoring	13(65)	-	3(15)	5(25)
Every month	2(10)			
Every 9 months	1(5)			
Annually	3(15)			
Monitoring completed when data points are identified with issues of poor quality	1(5)			
Varies between projects and study design	5(25)			
Variables included in data monitoring	12(60)	-	4(20)	4(20)
Critical data points	1(5)			
Critical and non-critical data points	3(15)			
All (100%) data points	5(25)			
Varies between projects and study design	3(15)			
Sampling techniques to select data points[1]	7(35)	1(5)	5(25)	7(35)
Simple random sampling	4(20)			
Systematic sampling	2(10)			
Stratified sampling	2(10)			
Cluster sampling	1(5)			
Varies between projects and study design	1(5)			

Centralized monitoring: Data collected through an electronic data capture and queries identified by monitor that may need further attention to alleviate problems

Remote monitoring: Data monitored off-site, includes delivering documents via email, fax or snail mail to monitors to conduct source data verification.

Logic, range, and consistency check: Logic check, flag indicator results that fail a common-sense comparison to other indicator or other disaggregation; Range check, check the value of data to see if it is within a certain range; Consistency check, performed to determine if the data has an internal conflict and data field correspond

On-site SDV: At the site comparing source data (original or certified copy) document to data recorded or entered to a case report form or electronic record or database.

Statistic technique: For example cluster and outlier analysis.

Risk-based targeted monitoring: Focus on a certain data point that has been identified to have the most risk.

Risk-based triggered monitoring: After a certain event like a large number of adverse events or deviations further detailed monitoring occurs.

---

[1] Sampling technique refers to the way data points are selected to be a sample for data monitoring.

Table 4: Type of staff training and development devoted to data quality

Type of training/development	n(%)
Education throughout clinical trial (as needed)	13(65)
ICH-GCP training	12(60)
Group education and training	11(55)
SOP training	11(55)
Education prior to research	10(50)
Skills training and development	7(35)
One-on-one education and training	6(30)
Not applicable	1(5)
Other	1(5)

ICH-GCP: International Conference on Harmonisation and Good Clinical Practice
SOP: Standard Operating Procedure