Improving the quality of drug research or simply increasing its cost? An evidence-based study of the cost for data monitoring in clinical trials

Esther Pronker,¹ Bart F. Geerts,² Adam Cohen³ & Herman Pieterse⁴

¹Vacceleron, Leiden, ²Department of Anaesthesiology, Leiden University Medical Centre, Leiden, ³Centre for Human Drug Research, Leiden and ⁴Profess Medical Consultancy BV, Heerhugowaard, The Netherlands

WHAT IS ALREADY KNOWN ABOUT THIS SUBJECT

- Current knowledge is based on practical experience. This paper presents the first novel, quantitative and qualitative view on the efficacy of the sponsor query system widely applied in the pharmaceutical industry, used for quality control purposes by data management for clinical data.

WHAT THIS STUDY ADDS

- This study presents a structured view on the process of clinical data management and sponsor queries, the efficacy and cost effectiveness of sponsor queries and dual data entry and the conclusion to advocate a more evidence-based approach to clinical data validation.

Correspondence

Miss Esther Pronker MSc, Centre for Human Drug Research, Zernikedreef 10, 2300 CL Leiden, The Netherlands.
Tel.: + 0(031) 7152 6400
Fax: + 0(031) 7152 6499
E-mail: e.s.pronker@gmail.com

Keywords

clinical data management, clinical research and sponsor query/queries, contract research organization

Received

8 December 2009

Accepted

20 October 2010

AIM

Procedures for verification of data from clinical studies are intended to maintain reliability for clinical trial results. Guidelines or legislations relating to clinical data management are of limited value and no study has yet demonstrated its effectiveness.

METHOD

Sponsor queries and dual entry procedures from one CRO on three different phase I trials are analysed on content, impact and cost.

RESULT

In this study, sponsor queries and dual entry procedures proved time and cost inefficient in detecting data discrepancies.

CONCLUSION

We advocate a more evidence-based approach for enhancing data integrity throughout the process of clinical data management.

Introduction

The price of drugs has increased and consequently the pharmaceutical value chain has come under scrutiny [1]. Drug pipelines are dwindling, leading to increasing cost for drug development [2], and patients are charged more for their medications [3, 4]. Clinical phases of drug develop-

ment require significant resources to assure patient safety and data integrity, the latter being endorsed by clinical data management departments (CDM) [3, 4]. In view of the high investment in clinical development, CDM could be a suitable candidate for cost savings.

Regulatory bodies have intensified administrative specifications for every process within clinical trials, aimed

at preventing fraud and medical incidents [3]. These developments are principally useful but have strongly added to the operational difficulties of performing clinical trials. International Organization for Standardization and International Conference on Harmonization documents only specify that data should be accurate, complete, legible and timely [5]. With no instruction on how to attain or achieve this, trial organizers create individualized solutions.

As a result, industry dogma defines how data integrity and quality can only be guaranteed through 100% data validation. The current best practice for assuring data integrity entails several procedures at each step of the study life cycle [6], including dual-data entry, external audits, regular external and internal monitoring, monitor queries, sponsor queries and governmental audits¹. Each audit and monitoring procedure applies similar data validation tools [7].

In this study we examined the cost and potential efficacy of commonly used methods of assuring data integrity, focusing on sponsor queries and dual data entry methodologies. We argue that the sum of all procedures are costly and labour intensive, but may be excessive as the final objective is that trial results are reliable. We find that not all individual data points have to be correct, as randomization assures absence of bias. Moreover, there seems to be no distinction between trivial and critical issues, seeing that not all data points are equally important for statistical evaluation [8]. We suggest an evidence-based approach to select actively critical parameters for assuring data integrity.

Methods

First, we assessed the efficacy of the sponsor query procedure². The study was performed at the Centre for Human Drug Research (CHDR), an academic oriented CRO with qualified personnel and clearly defined standards of operation (SOP) with an autonomous quality-assurance officer. For this analysis three phase I studies were selected, each having similar data processing routes, but differing in sponsors, number of participating volunteers, duration and time period in which they were performed. One thousand three hundred and ninety-five sponsor queries from the three studies were reviewed.

Sponsor queries were assessed on two facets, keeping in mind that one query addresses one data point. A query was evaluated on content (scores are based on mutually

¹The innovation of direct data entry (recording the data directly into an electronic system, eliminating the use of paper CRFs) has not been taken into account in this research.

²There are three types of queries based on who identified the query: CRO internal queries, monitor queries and sponsor queries. Sponsor-originated queries are the only formally recorded queries that are communicated from the sponsor back to the CRO. CRO internal queries or monitor queries are usually reported by word-of-mouth or e-mail. However a full accurate report of these types of queries is not available.

Table 1
Impact of sponsor queries

PART A		
Question	Value/Quantity	%
Impact Q1: Was data changed?		
No	1003	71.90
Yes	392	28.10
Impact Q2: Was confirmation asked for a data point?		
No	199	14.27
Yes	1196	85.73
Impact Q3: Did the query concern an endpoint?		
No	298	21.39
Primary	570	40.92
Secondary	382	27.42
Other	143	10.27
Impact Q4: Was the change significant for the specific data point?		
No	1326	95.05
Yes	69	4.95
Impact Q5: Could the change have changed the results of the trial?		
No	1389	99.57
Yes	6	0.43
PART B		
Impact Q5: Explained	Sponsor query	
A	On the <i>General Physical Examination Page</i> a very long abnormality for psychiatric behaviour is recorded. This term is too long to enter on this page, therefore DE wrote a comment	
B	The date and time performed should respect the theoretical time. Please verify and/or confirm	
C	Actual does should be equal to planned dose. Please correct	
D	First inhalation of product has a comment, indicating that a leak existed, but full actual dose is reported. Please correct or explain.	
E	The estimated date is prior to the demography date. Please clarify	
F	An AE took place, and is recorded with <i>action taken C.O.</i> The corresponding CNP page is empty. Please provide treatment/procedures with relevant information to be recorded on the CNP, or verify that the action O can be removed for the AE.	

Table showing impact of sponsor queries based on the five question impact measure (per cent accurate to 2 decimal places); Impact Question 1 refers to whether the data point is adjusted as a direct consequence of the query. Impact Question 2 indicates whether the query challenges the credibility of the data point, by stating 'please confirm'. Impact Question 3 refers to whether the data point queried is related to a clinical endpoint. Impact Question 4 refers to the empirical judgement of whether the data point has potential statistical impact on its specific parameter. Impact Question 5 asks: Would the discrepant data, if left unnoticed by the sponsor query, have any influence on the outcome of the clinical trial? Below Impact Question 5 is a description of the six sponsor queries that could potentially influence the statistics of the clinical trial.

exclusive categories, data point content, data point characteristic, query destination for answering, query topic and re-queries) and impact (scores are based on a five-question relative impact model, Table 1). All queries were scored by two arbiters in a double-blind fashion, and conflicting results were reviewed and resolved by an independent CDM consultant.

The data were analyzed using statistical descriptives, followed by Fisher-exact and Pearson's Chi-Squared to assess the relative association between categories. We studied associations between data point content parameters and the five impact questions individually, and refuted the null hypothesis each time.

Second we reviewed the dual-data entry procedure. We used information on dual-entered study data as conducted at CHDR over the last 5 years in their trial database (Promasys BV Leiden, the Netherlands). Numerical data from first and second entry were compared and if a change was made on the second entry (assumed to indicate a discrepant first entry) the percentage difference between the two entries was recorded ignoring the positive or negative value. If a text value was changed during second entry the difference in number of characters was expressed as a percentage of the original entry.

Ultimately; the cost involved was estimated based on a procedural flow-chart and real-time recording of activities at the CRO.

Results

Sponsor query assessment

When assessing the query content parameter, 70% of queries addressed administrative qualities of the data point, for example an unclear checked box, whereas 12% of the queries addressed a medical data point. For query impact, 80% of the queries required a confirmation of the data point. The majority of the data points queried were related to a clinical endpoint (68%).

There were only six queries (0.4% from 1395 queries and 0.001% of the combined 599 154 data points) that might have influenced the results of the clinical trials if the discrepancy had not been revealed. This leads to a number needed to treat of 10 000 data points in order to find a possible significant error.

The assessment was conservative as the six queries concerned a discrepancy in the coding of a non-serious adverse event. Five of the six queries were related to an administrative parameter but referred to a critical data point that could potentially influence the trial results.

The cost of the sponsor-CRO query procedure cannot be accurately defined. However, if we assume that the handling of a single query by staff at the trial site and the sponsor takes about 1 h combined, the cost can be conservatively estimated at about €150. This means that for the three trials about €200 000 was spent for the correction of a minute amount of erroneous data.

The dual data-entry procedure

We evaluated the efficacy of this procedure to detect significant errors that might influence the outcome or conclusions of a trial (Table 2). Of the total number of dual entries ($n = 1\,605\,682$) 1.8% were changed during dual entry and

Table 2

CHDR database inspection of single and dual entry

	Number of data points	(%)
Total number of data points	2 806 797	–
Single entry	1 230 738	–
Dual entry	1 576 059	–
Number changed after single entry	22 533	1.83
% change	–	155.8
Number changed after dual entry	1 242	5.5

The table shows the descriptive results of the data points changed after dual entry. % value is accurate to one decimal place.

Figure 1

Histogram showing magnitude of differences between first and second data entry

the average change amounted to 156% of the primary entered original value. The magnitude of most of these changes was within 0–150%, with outliers of over 500% (Figure 1). The parameters tested were measured at least on 10 occasions in at least 10 studies.

If these changes were all in one direction this would lead to a maximum theoretical difference in the average value of a data set of 1.7% compared with the situation in which the errors were not detected. The probability of such a difference leading to an important change in statistical inference is low in view of the normal variability in biological data that generally exceeds 10%.

Dual entry of this number of data points (assuming 10 s per entry) approximately requires 2 man years at an all inclusive cost of approximately €200 000.

Discussion

Clinical trials are essential for evaluation of many interventions in health care making quality data indispensable.

During the evolution of the pharmaceutical value chain, procedures have been added [9]. We have demonstrated that traditional procedures need to be evaluated continuously to assure that they are cost effective. By performing this evidence-based audit we have demonstrated the resources involved with generating and solving sponsor queries. Additionally, we estimated the cost savings to be considerable, excluding knock-on effects on travel expenses and infrastructure.

To reduce CDM cost we propose small procedural alterations. First direct data-entry could replace dual-data entry. Second, sponsors could review queries before blindly sending them to the CRO, filtering out queries that relate to self-evident checks. This can be accomplished through updating the trial validation plan and strategy. Third, communication between sponsor and CRO could be improved to implement a feedback system on query type that allows for a learning curve; fewer sponsor queries nearer trial completion. Furthermore, giving more attention to the initial planning phases of a study may also affect data quality at later stages.

Last, we advocate a more evidence-based approach to clinical data management using the concept of 'resilience', the degree of flexibility for data point error. Currently there is no assessment to identify high risk data points that, if discrepant, could influence the results and consequently the conclusions drawn from a trial. We hypothesize that it is possible to pre-select these susceptible data points based on two criteria: its relation to a clinical endpoint and its flexibility for being discrepant. The latter criterion can be forecast using power-based calculations to identify high-risk data points ($P = 0.95$ for example), that have the potential to influence the statistics if discrepant. This will reduce the resources involved as instigated by industry dogma and increase CDM efficiency.

Considerations

Several restrictions apply to this study. First, the sample consisted of only three studies. Although selection was actively performed to prevent structural bias and spread the variables, it was only a small sample compared with the hundreds of drug trials performed on a global scale. Additionally, the extremely low detection rate of erroneous values may not be representative for other situations. They were obtained in a GCP-regulated and professionally staffed unit, making error rates (and also cost per error) quite different from other environments. The value of different quality assurance procedures therefore might be tested for that particular environment. Our data and methodology may assist with this.

This article presents the first empirical study on the topic of sponsor queries and the CDM system. This pilot study can benefit scientific organizations and pharmaceutical companies by starting to rethink the concept of data validation and current procedures to achieve this.

Competing interests

Statements, permissions and conflict of interest

All authors hereby declare to have participated in this study (i.e. conception, execution and drafting of the paper) and have seen and approved this final manuscript. All authors declare there is no conflict of interest of any kind as defined by <http://mc.manuscriptcentral.com/societyimages/bcp/Declaration.pdf>.

Submission of this article implies that the work described has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, without the written consent of the copyright-holder. As defined by http://www.blackwellpublishing.com/pdf/BJCP_ELF.pdf.

The study was supported solely by institutional grants. We would like to thank the Promasys team, Frank Stap and Marieke de Kam for their assistance.

REFERENCES

- 1 Shuchman M. Commercializing clinical trials – risks and benefits of the CRO boom. *N Engl J Med* 2007; 357: 1365–8.
- 2 Hughes B. 2007 FDA drug approvals: a year of flux. *Nat Rev Drug Discov* 2008; 7: 107–9.
- 3 Califf RM. Clinical trials bureaucracy: unintended consequences of well-intentioned policy. *Clin Trials* 2006; 3: 496–502.
- 4 Rondel RK, Varley SA, Webb CF. *Clinical Data Management*. Chichester: John Wiley and Sons, 2000.
- 5 ICH Topic E 6 (R1) Guideline for Good Clinical Practice: CPMP/ICH/135/95. European Medicines Agency. July 2002. Available at http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002874.pdf (last accessed 19 November 2010).
- 6 Promasys BV. *Promasys for Windows; User Manual*. Leiden Bio Science Park: CHDR, 2007.
- 7 Prokscha S. *Practical Guide to Clinical Data Management*. 2nd edn. FL, USA: Taylor and Francis group, 2007.
- 8 Pieterse H. *The Future of Computerization at the Investigator Site*. Amsterdam: Health Information Initiatives in the Netherlands, 1998.
- 9 Ratti E, Trist D. The continuing evolution of the drug discovery process in the pharmaceutical industry. *Farmaco* 2001; 56: 13–9.