# A Pilot Study on the Quality of Data Management in a Cancer Clinical Trial


## E. van der Putten, J. W. van der Velden, A. Siers, and E. A. M. Hamersma, for the Cooperative Study Group of Dutch Datamanagers

*Comprehensive Cancer Center, Amsterdam (E.v.d.P.); Daniël den Hoed Cancer Center, Rotterdam (J.W.v.d.V.); Comprehensive Cancer Center, Utrecht (A.S.); The Netherlands Cancer Institute, Amsterdam (E.A.M.H.)*



**ABSTRACT:** Twelve institutional data managers were asked to independently code the data from a patient chart of one patient in an ovarian cancer trial. They abstracted data from the medical record and filled out three types of trial forms (on-study, chemotherapy, and summary forms). The analysis of the processed data revealed that the median rate of errors was 13% for the 12 data managers. The error rate differed among the types of trial forms. The factors causing these errors were mistakes in interpretation, documentation, and coding. The level of experience of the data managers proved to be an important factor. It became clear that the documentation in the medical record was inadequate. We conclude that data managers as well as physicians involved in cancer clinical trials need specific training and that the quality of data management in cancer clinical trials is an important issue for further investigation.



KEY WORDS: *cancer clinical trial, data management, quality control*


## INTRODUCTION

Results of cancer clinical trials are often difficult to interpret, due to differences in selection criteria, treatment schedules, and response criteria [1]. For most tumor types and/or treatment modalities a large number of patients in phase III trials is needed to guarantee reliable study outcomes, and multicenter studies are often mandatory. A complicated trial design with long-lasting treatment schedules is one of the problems that occur when clinical trials have to be analyzed. It is clear that the interpretation of results of clinical studies is a difficult task; however, few studies investigate the quality of study data on which the trial results are based.


*Address reprint requests to: E. van der Putten, Comprehensive Cancer Center Amsterdam, Plesmanlaan 125, 1066 CX Amsterdam, The Netherlands.*


*Received February 10, 1986; revised November 4, 1986.*




0197-2456/87/$3.50

Controlled Clinical Trials 8:96–100 (1987)



52 Vanderbilt Ave., New York, New York 10017

A study by Sylvester et al. [2] showed, in a multicenter study design, difference in the quality of data, favoring "major" participants (with an average of more than five patients per trial) as opposed to "minor" participants (with an average of less than five patients per trial). In contrast, in another study by the Eastern Cooperative Oncology Group [3] the results do not indicate that institutions entering a small number of patients in a clinical trial have data of less quality, provided that a general instruction for filling out trial forms is available.

Data managers play a central role in the conduct of cancer clinical trials within the European Organization for Research on Treatment of Cancer (EORTC). Two types of data managers are characterized. First of all there are the data managers at the coordinating center. These data managers are responsible for managing the data base of a specific trial and for analyzing the incoming flow of study data in close cooperation with the study coordinator (principal investigator who has the overall responsibility for the study). The second type of data managers have the task of abstracting the study data from the medical record (or patient chart) within the participating institutions. Studies that examine the influence on the quality of data supplied by the institutional data managers are even more scarce [4,5]. The aim of our study was to determine whether the quality of trial data can be influenced by the skill and experience of a institutional data manager.

## METHODS

We copied the patient chart of a patient treated according to an ovarian cancer trial. The copies were distributed among 12 institutional data managers, who were asked to independently code the data of the trial patient using the appropriate trial forms. One copy of the patient record plus the trial forms was handed to the study coordinator of the trial in order to standardize the data. The forms filled in by the study coordinator were used as the standard of comparison. All 12 data managers were experienced in processing data from clinical trials and participated voluntary in this study. The patient on which this comparative study is based had been treated in a phase III EORTC study with chemotherapy and surgery. In this trial patients with ovarian cancer, FIGO stage $II_{b,c}$, III, IV were randomized to receive either a standard cisplatinum-containing regimen or an experimental carboplatinum combination. The forms used for this trial were as follows: on-study form, chemotherapy form, and summary form. The trial forms contained questions on prognostic factors (performance score, type of histology, etc.), side effects due to chemotherapeutic treatment, and treatment evaluation.

The data supplied by the study coordinator were carefully checked by a small working party (specialists in the field of trial oncology) and with minor corrections these data were accepted as "standard data." The study coordinator had also actually treated this particular patient in his own clinic. The trial forms of the data managers were collected and compared with the standardized data. The number of errors per data manager were counted and the reasons for these errors were analyzed by the above-mentioned working party.

## RESULTS

The total number of therapy related questions, the so-called items, on the three types of trials forms was 84. The median number of mistakes made by the twelve data managers was 11 (13%), ranging from 6 to 17 (7%–20%) over all items on the forms. We examined the distribution of the errors over the different types of study forms. The number of items on the on-study, chemotherapy, and summary forms were 15, 54, and 15, respectively. As is shown in Table 1, the median number of mistakes were not equally divided over the different trial forms.

The errors made by the data managers were classified into three categories:

*Interpretation errors:* This type of error was considered to be the most important one. Occasionally it occurred that although the answer to a question on the trial form was available in the patient chart an incorrect answer was given. Most of the time the line of questioning lacked specificity; it was possible to give more than one answer to the same question. For example, on the on-study form "initial" laboratory values were demanded. The word "initial" was not identically interpreted; in fact seven different values for the "initial" serum creatinine were given. Also the "date of diagnosis" caused some confusion. Four different options were mentioned: operation date, date the pathology laboratory received the tumor specimen, date of the pathology report, and the date the pathology outcome was received by the clinician.

*Documentation errors:* Unlike the interpretation errors, which were due to problems in perceiving the exact meaning of questions on the forms, the documentation errors originated from inadequate information in the patient chart. In some cases essential information could not be directly found in the chart. At the third chemotherapy treatment cycle the data managers could only guess if one of the cytostatic drugs was reduced for 1 or 2 weeks. This resulted in a dosage in the drug cyclophosphamide that varied in the same treatment cycle from 525 up to 2100 mg.

*Coding errors:* These mistakes do not fit into the category of interpretation and documentation errors. In other words, the question on the trial form was explicit and the required answer could have been directly abstracted from the chart. Most frequently these coding errors resulted in incorrect dates. This is clearly illustrated by the fact that some data managers gave an

**Table 1** Distribution of Errors by Trial Form

| Form | | Median Number of Errors (%) | Range (%) |
|---|---|---|---|
| On-study | ($n$ = 15)[a] | 3 (20) | 1–6 (7–40) |
| Chemotherapy | ($n$ = 54) | 6 (11) | 2–8 (4–15) |
| Summary | ($n$ = 15) | 2 (13) | 0–3 (0–20) |

[a] $n$ = number of questions on the form.

**Table 2** Categories of Errors by Trial Forms

| Form | | Median Number of Errors (%) Interpretation | Documentation | Coding |
|---|---|---|---|---|
| On-study | ($n$ = 15)[a] | 3 (20) | 0 (0) | 0 (0) |
| Chemotherapy | ($n$ = 54) | 2 (4) | 0 (0) | 4 (7) |
| Summary | ($n$ = 15) | 0 (0) | 1 (7) | 1 (7) |
| Total | ($n$ = 84) | 5 (6) | 1 (1) | 5 (6) |

[a] $n$ = number of questions on the form.

incorrect date of birth for the patient. The median number of mistakes sorted by the three categories and study form are given in Table 2.

## DISCUSSION

The data managers had had experience in this specific field of trial work ranging from 1 to 8 years, with a mean of 3 years. The data managers had had different types of (para)medical training. Five data managers were nurses, three were physicians, and four had miscellaneous types of schooling (biology, psychology, laboratorian technician, and physical therapy). The time data managers spent on filling out the forms varied from 1 to 8 hours. This did not seem to influence the quality of data, in contrast to experience level of the data manager, which proved to be an important factor, as experienced data managers made fewer errors. It should be mentioned that the data managers were unable to consult the physician who had treated the patient, something they normally do when difficulties in treatment evaluation arise.

We can look at the causes of the three types of errors (interpretation, documentation, coding) from two different angles. From a clinical point of view it can be stated that the documentation in the patient chart was inadequate and this is an indication that clinicians should work strictly according to protocol. A checklist would help to ensure that a patient's treatment or follow-up takes place at the appropriate time.

On the other hand it is clear that data managers need specific training in this line of work, and that close cooperation between data manager and local investigator is very important. In addition, as problems were encountered in the patient chart itself, better training of physicians participating in clinical trials would be of great interest. The only type of training so far for institutional data managers is doing practical work in a experienced trial department. There is need for some structured schooling in general oncology as well as in specific trial oncology. On the latter part, the trial oncology, one could also think of workshops on typical data management problems. Workshops would provide an excellent platform for institutional and coordinating data managers for exchanging their experiences. In this setting the possibility of making manuals with item explanations could be studied. The results in this limited study suggest that there is field for further investigation on the quality of data

supplied for cancer clinical trials. Larger studies with more patients and less complicated questions are required to give an indication of the impact of the quality of data management on study results in cancer clinical trials.

We thank the data managers for providing the study material for our investigation and W. W. ten Bokkel Huinink, who not only supplied the standardized data but who was also a great stimulus for the whole study.

## REFERENCES

1. Pocock SJ: Clinical Trials: A Practical Approach. New York: Wiley, 1983
2. Sylvester RJ, Pinedo HM, de Pauw M, Staquet MI, Buyse ME, Renard J, Bonadonna G: Quality of institutional participation in multicenter clinical trials. N Engl J Med 305: 852–855, 1981
3. Eastern Cooperative Oncology Group: Participation of community hospitals in clinical trials: Analysis of five years of experience in the Eastern Cooperative Oncology Group. N Engl J Med 306: 1076–1080, 1982
4. Jasperse DM, Ahmed SW: Comparison of two data collection methods. Controlled Clin Trials 6: 241, 1985
5. v.d. Velden JW, Hamersma EAM, v.d. Putten E: Quality control, a special study by Dutch data managers. Controlled Clin Trials 3: 48, 1985