

Research article

## Quality control and data-handling in multicentre studies: the case of the Multicentre Project for Tuberculosis Research

Teresa Caloto<sup>1</sup>, Consuelo Huerta<sup>1</sup>, Teresa Moreno<sup>1</sup>, Dolores Guerra<sup>1</sup>, José Alcaide<sup>2</sup>, Concha Castells<sup>3</sup>, José I Cardenal<sup>4</sup>, Angela Domínguez<sup>2</sup>, Pilar Gayoso<sup>5</sup>, Gonzalo Gutiérrez<sup>6</sup>, Maria J López<sup>7</sup>, Francisco Muñoz<sup>8</sup>, Carmen Navarro<sup>9</sup>, Miguel Picó<sup>10</sup>, Francisco Pozo<sup>1</sup>, José R Quirós<sup>11</sup>, Francisco Robles<sup>12</sup>, José M Sánchez<sup>13</sup>, Hermelinda Vanaclocha<sup>14</sup>, Tomás Vega<sup>15</sup> and Mercedes Diez\*<sup>1</sup> for the MPTR Study Group

Address: <sup>1</sup>Unidad de Investigación en Tuberculosis, Instituto de Salud Carlos III, Sinesio Delgado, 6, 28029 Madrid, Spain; <sup>2</sup>Departamento de Sanidad y Seguridad Social, Dirección General de Salud Pública, Travessera de les Corts 131-159, 08028 Barcelona Spain; <sup>3</sup>Departamento de Sanidad, Delegación Territorial, María Díaz de Haro 60, 48010 Bilbao Spain; <sup>4</sup>Consejería de Bienestar Social, Servicio Territorial, Plaza de Alférces Provisionales s/n, 10071 Cáceres, Spain; <sup>5</sup>Complejo Hospitalario Cristal-Piñor, Ramón Puga 54, 32005 Orense, Spain; <sup>6</sup>Dirección General de Salud Pública, Avenida de Francia 4, 45071 Toledo, Spain; <sup>7</sup>Consejería de Salud Consumo y Bienestar Social, Dirección General de Salud y Consumo, Villamediana 17, 26071 Logroño, Spain; <sup>8</sup>Hospital Universitario de Valme, Carretera de Cádiz s/n, 41014 Sevilla, Spain; <sup>9</sup>Consejería de Sanidad y Política Social, Dirección General de Salud, Ronda de Levante 11, 30008 Murcia, Spain; <sup>10</sup>Dirección General de Salud Pública, Avenida de la Innovación s/n, Edificio Arena 1, 41020 Sevilla, Spain; <sup>11</sup>Dirección Regional de Salud Pública, General Elorza 32, 33001 Oviedo, Spain; <sup>12</sup>Dirección Provincial de Sanidad de Melilla, Plaza 1 de Mayo s/n, 29803 Melilla, Spain; <sup>13</sup>Dirección Provincial de Sanidad de Ceuta, Carretera San Amaro 12, 51001 Ceuta, Spain; <sup>14</sup>Dirección General de Salud Pública, Doctor Rodríguez Fornos 4, 46010 Valencia, Spain; and <sup>15</sup>Dirección General de Salud Pública y Asistencia, Avenida de Burgos 5, 47071 Valladolid, Spain

E-mail: Teresa Caloto - [Teresa\\_caloto@merck.com](mailto:Teresa_caloto@merck.com); Consuelo Huerta - [chuerta@ceife.es](mailto:chuerta@ceife.es); Teresa Moreno - [mmoreno@isciii.es](mailto:mmoreno@isciii.es); Dolores Guerra - [lola@enterprise.eui.upm.es](mailto:lola@enterprise.eui.upm.es); José Alcaide - [jalcaide@dsss.scs.es](mailto:jalcaide@dsss.scs.es); Concha Castells - [depidebi-san@eg-gv.es](mailto:depidebi-san@eg-gv.es); José I Cardenal - [jicardenal@bme.es](mailto:jicardenal@bme.es); Angela Domínguez - [angelad@dsss.scs.es](mailto:angelad@dsss.scs.es); Pilar Gayoso - [pgayoso@cristalp.es](mailto:pgayoso@cristalp.es); Gonzalo Gutiérrez - [ve@jccm.es](mailto:ve@jccm.es); Maria J López - [mjose.lopez@larioja.org](mailto:mjose.lopez@larioja.org); Francisco Muñoz - [med000552@nacom.es](mailto:med000552@nacom.es); Carmen Navarro - [Carmen.Navarro@carm.es](mailto:Carmen.Navarro@carm.es); Miguel Picó - [sito@arrakis.es](mailto:sito@arrakis.es); Francisco Pozo - [fpozo@h120.es](mailto:fpozo@h120.es); José R Quirós - [ramonqg@princast.es](mailto:ramonqg@princast.es); Francisco Robles - [froblesf@camelilla.es](mailto:froblesf@camelilla.es); José M Sánchez - [sanfer@teleline.es](mailto:sanfer@teleline.es); Hermelinda Vanaclocha - [vanaclocha\\_her@gva.es](mailto:vanaclocha_her@gva.es); Tomás Vega - [sybs.epi@dvnet.es](mailto:sybs.epi@dvnet.es); Mercedes Diez\* - [mdiez@isciii.es](mailto:mdiez@isciii.es)

\*Corresponding author

Published: 21 December 2001

Received: 9 October 2001

BMC Medical Research Methodology 2001, 1:14

Accepted: 21 December 2001

This article is available from: <http://www.biomedcentral.com/1471-2288/1/14>

© 2001 Caloto et al; licensee BioMed Central Ltd. Verbatim copying and redistribution of this article are permitted in any medium for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** The Multicentre Project for Tuberculosis Research (MPTR) was a clinical-epidemiological study on tuberculosis carried out in Spain from 1996 to 1998. In total, 96 centres scattered all over the country participated in the project, 19935 "possible cases" of tuberculosis were examined and 10053 finally included. Data-handling and quality control procedures implemented in the MPTR are described.

**Methods:** The study was divided in three phases: 1) preliminary phase, 2) field work 3) final phase. Quality control procedures during the three phases are described. Results: Preliminary phase: a) organisation of the research team; b) design of epidemiological tools; training of researchers. Field work: a) data collection; b) data computerisation; c) data transmission; d) data cleaning; e) quality control audits; f) confidentiality. Final phase: a) final data cleaning; b) final analysis.

**Conclusion:** The undertaking of a multicentre project implies the need to work with a heterogeneous research team and yet at the same time attain a common goal by following a homogeneous methodology. This demands an additional effort on quality control.

## Background

Multicentre studies call for additional logistic and methodological effort, yet this is offset by the advantages to be gained from obtaining a larger sample more quickly and improving the external validity of the results.

The Multicentre Project for Tuberculosis Research (MPTR) was a clinical-epidemiological study conducted into tuberculosis (TB) in Spain during the period, 1996–1998. For the purposes of the study a TB case was defined as anyone who fulfilled the following two conditions: a) microscopy and/or culture positive for *Mycobacterium tuberculosis complex*; and, b) therapy with at least two anti-TB drugs prescribed by a physician. Subjects who met the second but not the first condition were only included as cases if the prescription was still in place after three months. The field work in the MPTR comprised: a) identifying 19935 TB suspects by a monthly search of 14 databases for one year: a specific definition of TB suspect was established for each database; b) reviewing the respective clinical histories; and, c) collecting and computerising detailed information on the 10053 cases that met the case definition. These tasks were undertaken at a local level in all of the 96 participant public health areas (PHA), situated in 13 of Spain's Autonomous Regions (AR), namely, Andalusia, Principality of Asturias, Castile-La Mancha, Castile & Leon, Catalonia, Extremadura, Galicia, La Rioja, Murcia, Basque Country, Valencia, Ceuta and Melilla. Data were first aggregated at a regional level, and thereafter at the Tuberculosis Research Unit of the Carlos III Institute of Public Health, which acted as the Co-ordinating Centre (CC) and performed the necessary data-analysis. Several papers have been published based on the results from the MPTR [1–3].

## Methods

The study was divided into three phases, each one subdivided into different processes which are summed up in the following:

- 1) preliminary phase: organisation of the research team, design of epidemiological tools and training of researchers,
- 2) field work: data collection, data-computerisation and -transmission, data cleaning, quality control audits and confidentiality, and
- 3) final phase: data cleaning and final analysis.

The type of action taken at each phase of the process to ensure the reproducibility and validity of the information, along with the procedures implemented in order to measure quality (Figure 1), are described below.

## Results

### Preliminary phase

Data quality control is an aspect that has to be considered at the planning phase of any study, and particularly so in cases, such as multicentre studies, which necessarily involve researchers based at facilities that are far apart.

### Organisation of the research team

The MPTR was structured as a co-ordinated project having the above three levels of action, i.e., PHA, AR and CC, with specific tasks allocated to each. To monitor the validity of the results and resolve logistic or methodological problems, a Project Management Team (*Equipo Directivo del Proyecto*) was formed, made up of CC personnel and representatives from each of the Autonomous Regions. The Team met five times during this phase to decide methodological and organisational aspects. Teams with similar responsibilities were set up at both regional and local levels.

### Design of epidemiological tools

In order to standardise data collection on the 124 study variables, a structured questionnaire was designed and a detailed handbook drawn up, containing definitions for each variable.

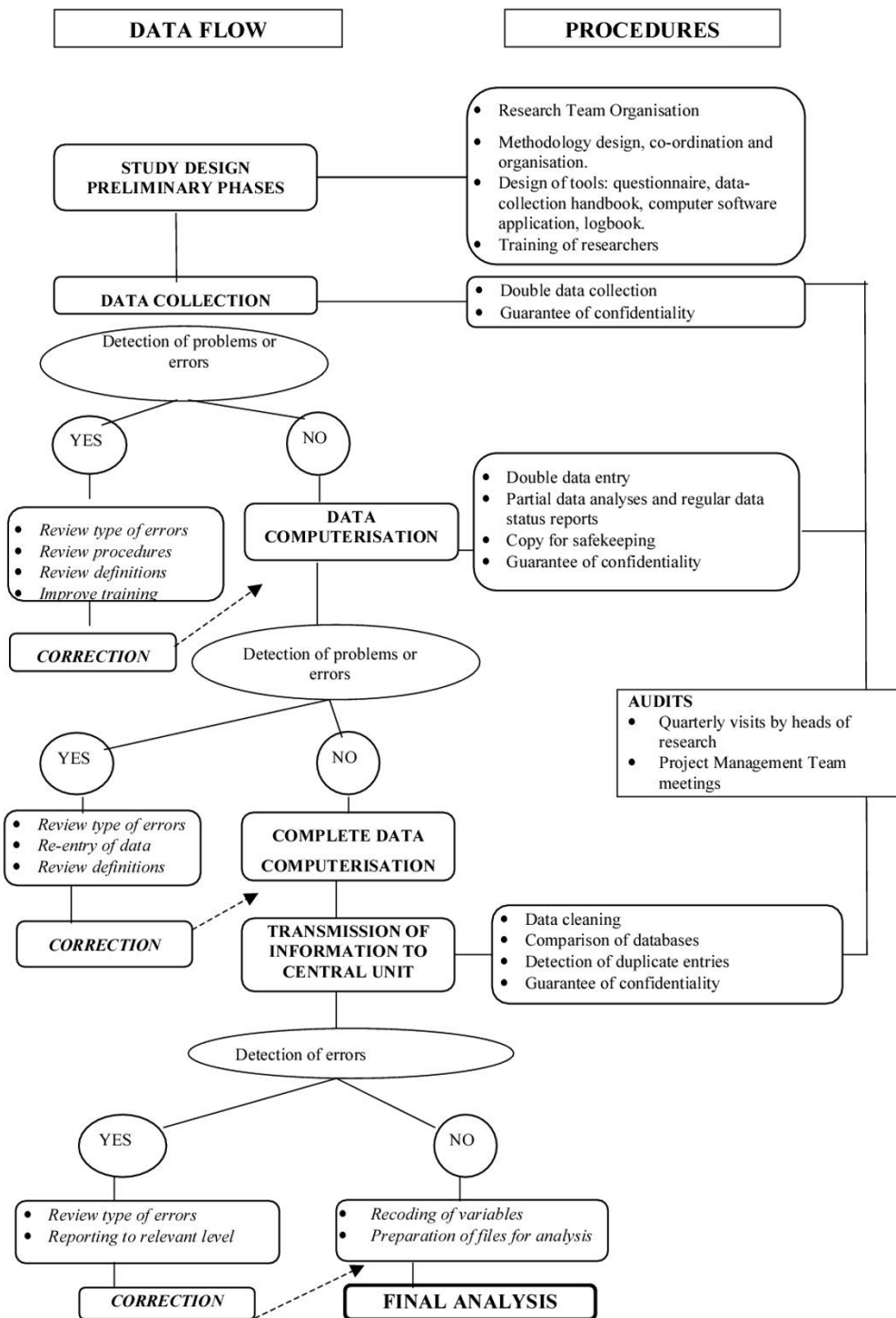
Two books were designed, namely, the Status Report (*Estadillo*) and "Log Book" (*Libro de Registro*). Whereas the former recorded the progress of the TB suspects from detection until confirmation, the latter systematically reflected the date, the name of the person doing the screening, the study procedures followed and any incidents arising at participant health-care facilities, AR and the CC.

A data-computerisation software application was purpose-designed, which in addition to all the standard functions, allowed for: a) data validation and monthly review of data consistency; b) detection and deletion of all duplicate entries; c) a breakdown of all cases pending information or confirmation; d) generation of random samples designed to check for data-entry errors; e) on-line personal data encryption.

Similarly, a second computer programme was specifically developed to detect inconsistencies, with all data being duly screened before onward transmission to higher levels.

### Training of researchers

All research staff tasked with case searching and data collection attended a two-day course, imparted in each AR by the same person.



**Figure 1**  
Data flow and procedures for quality control. MPTR

## Field work

### Data collection

The searching of the 14 databases used to identify TB suspects and reviewing of the clinical histories were both carried out in strict accordance with the study protocol. Under its terms, "TB suspects" were to be identified by means of a monthly search of all databases and duly recorded in the Status Report. The clinical histories of all such TB suspects were then reviewed: where the case was confirmed, the relevant information was recorded in the questionnaire and subsequently computerised; and where the TB suspect was not confirmed, a note of the reason for no confirmation (disease different from TB, out of the study period etc.) was entered into the Status Report, so that in each instance a judgement could be made on the appropriateness of not confirming the case.

### Data-computerisation and -transmission

Data were computerised and sent monthly from PHA to the AR. Here, after undergoing aggregation and quality control, these same data were dispatched within 10 days to the Tuberculosis Research Unit, where they were fed into the central database. Copies of all such databases and questionnaires remained at the various PHA and AR for the duration of the study. When the study had been concluded, all materials used (Log Books, Status Reports and questionnaires) were sent to the Tuberculosis Research Unit for filing, along with a final report confirming that the work had been done as per instructions.

To ensure data-entry quality, the Tuberculosis Research Unit typed in duplicate data for a random sample of 926 entries (approximately 10% of cases). There was an average of 15.3 errors per 10,000 characters, an error rate which is smaller than those of 22/10,000 and 23/10,000 found in studies where data-entry was performed locally as in the MPTR [4,5], but higher than the error rate of 9.5/10,000 or 3.8/10,000 found in studies where data entry was performed centrally [6,7].

### Data cleaning

All information forwarded by the AR underwent a monthly check for duplicates and errors at the Tuberculosis Research Unit, with any resulting flaws or discrepancies being recorded on monthly quality-control reports that were sent to the respective AR. In any instance where it became necessary for information to be checked and errors corrected, the AR instructed the pertinent PHA to carry out a new review of the relevant questionnaires or clinical histories; this continued procedure for quality control allowed for differences in quality of data collection between the 96 centres to be corrected by the end of the study. A record was kept of all amendments made. Quarterly analyses were run on the overall database to check for biases in data collection.

### Quality control audits

The head researcher at each participant health-care facility inspected the Logbooks and Status Reports once a month, to check whether the facts on record indeed corresponded to the procedures carried out. Moreover, to verify whether the information had been recorded accurately, a duplicate collection of data was made on the basis of the clinical histories of 5% of cases selected at random (520 cases overall).

Head researchers in the AR visited all the participant health-care facilities in the region once at the commencement, once at the end, and at quarterly intervals throughout the study. At three-monthly intervals, Tuberculosis Research Unit staff carried out an audit at a randomly selected facility in each of the AR. To standardise the auditing process and forestall omissions and oversights, the same quality control questionnaire was used for all visits, with detailed attention to all aspects to be monitored.

The results of these audits were recorded in the relevant Logbooks and *ad hoc* reports issued by the auditors to the Project Management Team. These reports were then discussed at the quarterly meetings, along with the partial analyses and any other items of interest.

### Confidentiality

In line with Spanish law governing data-protection, the following measures were adopted:

- a) database access was restricted, with each PHA allocated an installation code, as well as an access code subject to change every three months;
- b) questionnaires and diskettes were stored under lock and key; and,
- c) all identification data in the database were encrypted, and all such data in copies of questionnaires forwarded to AR and CC, deleted. In data sent via courier (with telephonic notification of dispatch and receipt), patients were solely identified by an eleven-digit code.

### Data cleaning and final analysis

On conclusion of the study, the CC unified the information proceeding from the three study levels by comparing the respective databases, carrying out the pertinent corrections and eliminating all duplicates. Each Region was furnished with a copy of its own final database.

For the purposes of analysis of tuberculosis incidence: cases were assigned to their respective health districts; vagrants were included in their Autonomous Region of residence; and patients who resided outside of the study area, were excluded.

## Discussion

Except in the case of clinical trials, published papers do not generally go far enough in providing the kind of detailed description demanded by quality-control methodology [8]. However, this is a matter of great practical importance that should be borne in mind in all phases of developing any project, and even more so in the case of a multicentre project.

The period preceding data collection is fundamental. It is in this phase that the organisation of the study has to be decided, data-collection procedures established, epidemiological tools designed, and data-collection and -computerisation personnel trained. It is therefore essential that sufficient time be devoted to the task, so as to ensure that no fieldwork begins until the procedures have been well defined, the epidemiological tools have been validated and distributed, and the researchers have received all the necessary training [6,9]. In line with the designated study objectives, this is the time to determine the precise nature of the information required and the manner of collecting same, without losing sight of the fact that the amount of data collected will inevitably exert a direct influence on the time employed and the end quality of the information [10].

The functions, both of the researchers and the various bodies involved, must be perfectly defined and delimited in the preliminary phases of the project, since it is upon these that the overall quality of the study will depend [9,11]. In the MPTR, three organisational levels with specific tasks and well-defined channels of communication were demarcated. We feel that herein lies one of the keys to the project's success, given that the execution of a uniform study in 96 widely dispersed health areas would be simply impossible unless all the parties involved have a clear idea as to what their responsibility is and to whom they are answerable when problems arise.

In the context of multicentre studies, special mention should be made of the CC, whose role in this type of project is crucial [12]. There is unanimity as regards entrusting the CC with the mission of ensuring the validity of the results, and it is this body that must thus take charge of organising and training researchers, implementing quality control and undertaking data handling and -analysis. In order to be able to perform these functions, mechanisms for co-ordination and feedback between the CC and the various organisational levels must be set up [9,13,14].

An important aspect is to ascertain whether data computerisation is to be delegated to the participating centres or carried out by the CC [15]. The decision must be taken on the basis of the amount of information, the time availa-

ble, the geographical spread of the centres and the resources available. Although this task tends to be centralised in the majority of studies, performing it locally is swifter, provides researchers with direct knowledge of their data without having to depend upon the information supplied by the CC and, by extension, enhances their involvement in the study. In contrast, the participation of a great number of individuals in this process calls for quality control to be tightened in respect of data entry [9].

When training researchers, it must be remembered that quality control can make no sense if those tasked with computerising the data fail to understand the importance of their work and so develop no commitment to it. It is at this point therefore that the objectives of the study must be described in detail, stress laid on the importance of having reproducible and high-quality information as a means of attaining said goals, and the implications of incomplete or low quality data discussed.

Opinions differ as to the real need for double data entry and the level at which this should be done. Some authors consider that the improvements in data quality do not justify the extra time and cost involved [16–19], given that in such cases all the study procedures must be doubled [16]. Others feel, however, that double data entry is justified because it has been used in numerous studies and serves to assure quality [4,6,20]. Finally, there are those that propose alternatives to this practice. In the MPTR, data entry control was deemed necessary in a sample of sufficient size to ensure that the results obtained were in line with what was judged acceptable [4–7].

The need to carry out regular audits of participant facilities in multicentre studies has been highlighted by bodies such as the National Cancer Institute (USA), which not only requires facilities to draw up a programmed audit schedule but has also published audit performance guidelines for the purpose [8]. Where researchers know that their work is going to be reviewed and assessed, they exercise greater care in the process of gathering the information, leading in turn to enhanced reliability of results. Periodically, project status reports should be issued and circulated to the researchers.

## Conclusions

In conclusion, it has to be said that the undertaking of a multicentre project implies the need to work with a heterogeneous and widely dispersed study population and research team, and yet at the same time attain a common goal by following a homogeneous methodology. This demands an additional effort in collecting the data: on the one hand, in order to unify methods and implement measures that minimise the variability injected by the high numbers of individuals participating in the process;

and, on the other hand, to establish mechanisms that monitor and measure the quality of the data collected. While both aspects are essential to ensure the validity of the results and therefore important to any study, there can be no doubt that they have to be that much more complete and comprehensive in multicentre studies. The MPTR is the largest TB study ever undertaken in Spain, and has yielded extremely valuable information on the disease [1–3]. We believe that this was possible due to the rigour with which the quality control mechanisms were implemented over the course of the study and served to enable highly reproducible and valid results to be obtained.

### Members of the Study Group

listed in [ <http://www.isciii.es/unidad/sgecnsp/centros/cne/infgralcne.html> ]

### Acknowledgements

This work was funded by a FIS grant, Exp. 99/0016. Dolores Guerra was funded with 2 grants by the Instituto de Salud Carlos III (96/4186 y 97/4040).

### References

- Díez Ruiz-Navarro M, Huerta Álvarez C, Moreno Casbas T, Guerra Pérez MD, Caloto González MT, et al: **La tuberculosis en España: resultados del Proyecto Multicéntrico de Investigación sobre Tuberculosis (PMIT)**. Editada por el Instituto de Salud Carlos III. Madrid, 1999
- Grupo de Trabajo del PMIT: **Incidencia de la tuberculosis en España: resultados del Proyecto Multicéntrico de Investigación en Tuberculosis (PMIT)**. *Med Clin (Barc)* 2000, **114**:530-537
- Grupo de Trabajo del PMIT: **El diagnóstico y tratamiento de la tuberculosis en España: Resultados del Proyecto Multicéntrico de Investigación sobre Tuberculosis (PMIT)**. *Med Clin (Bar)* 2001, **116**:167-173
- Reynolds-Haertle RA, McBride R: **Single vs. Double data entry in CAST**. *Controlled Clin Trials* 1992, **13**:487-494
- Zalokar M, Hallstrom A, Gillespie MJ: **Distributed data entry and quality control**. *Controlled Clin Trials* 1985, **6**:243
- Neaton JD, Duchene AG, Svendsen KH, Wentworth D: **An examination of the efficiency of some quality assurance methods commonly employed in clinical trials**. *Stat Med* 1990, **9**:115-124
- Newhouse MM for the MPS Coordinating Center: **Data entry design and data quality**. *Controlled Clin Trials* 1985, **6**:229
- Weiss RB: **Systems of protocol review, quality assurance, and data audit**. *Cancer Chemother Pharmacol* 1988, **42(Suppl)**:S88-S92
- Gassman JJ, Owen WW, Kuntz TE, Martin JP, Amoroso WP: **Data quality assurance, monitoring and reporting**. *Controlled Clin Trials* 1995, **16**:104S-136S
- Hosking JD, Newhouse MM, Bagniewska A, Hawkins B: **Data collection and transcription**. *Controlled Clin Trials* 1995, **16**:66S-103S
- Díez M por el Grupo de Trabajo del PMIT: **Consideraciones respecto a los estudios multicéntricos: a propósito del Proyecto Multicéntrico de Investigación sobre Tuberculosis**. *Gac Sanit* 2000, **14**:247-249
- Blumenstein BA, James KE, Lind BK, Mitchell HE: **Functions and organization of coordinating centres for multicenter studies**. *Controlled Clin Trials* 1995, **16**:4S-29S
- Volter JM: **Quality assurance in a cooperative group**. *Cancer Treat Rep* 1985, **69**:1189-1193
- Cassel GH, Ferris FL: **Site visits in a multicenter ophthalmic clinical trial**. *Controlled Clin Trials* 1984, **5**:251-262
- McFadden ET, LoPresti F, Bailey LR, Clarke E, Wilkins PC: **Approaches to data management**. *Controlled Clin Trials* 1995, **16**:30S-65S
- Day S, Fayers P, Harvey D: **Double data entry: what value, what price?**. *Controlled Clin Trials* 1998, **19**:15-24
- King DW, Lashey R: **A quantifiable alternative to double data entry**. *Controlled Clin Trials* 2000, **21**:94-102
- Zhang J, Hu W: **Single or double data entry: considerations based on a simple binomial model**. *Controlled Clin Trials* 1998, **19**:56-58
- Gibson D, Harvey AJ, Everett V, Parmar MKB, on behalf of the CHART Steering Committee: **Is double data entry necessary?**. *The CHART trials. Controlled Clin Trials* 1994, **15**:482-488
- Gagnon J, Province MA, Bouchard C, Leon AS, Skinner JS, Wilmore JH, et al: **The HERITAGE Family Study: quality assurance and quality control**. *AEP* 1996, **6**:520-529

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMedcentral will be the most significant development for disseminating the results of biomedical research in our lifetime."

Paul Nurse, Director-General, Imperial Cancer Research Fund

Publish with **BMC** and your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours - you keep the copyright



Submit your manuscript here:  
<http://www.biomedcentral.com/manuscript/>

[editorial@biomedcentral.com](mailto:editorial@biomedcentral.com)