



CIRUGÍA ESPAÑOLA

www.elsevier.es/cirugia


Original article

Protocolization of multicenter clinical studies in the digital era. Is useful data centralization by a data-manager?



Ríos Antonio^{a,b,*} Puñal-Rodríguez José Antonio^c Moreno Pablo^d Mercader-Cidoncha Enrique^e
 Ferrero-Herrero Eduardo^f Durán Manuel^{g,h} Ruiz-Merino Guadalupeⁱ Ruiz-Pardo José^j
 Rodríguez José Manuel^{a,b} Pedro Ramón Gutiérrez^{k,l}

^a Unidad de Cirugía Endocrina, Servicio de Cirugía General y de Aparato Digestivo, Instituto Murciano de Investigación Bio-Sanitaria (IMIB-Arrixaca), Hospital Clínico Universitario Virgen de la Arrixaca, Servicio Murciano de Salud, Murcia, Spain

^b Departamento de Cirugía, Pediatría y Obstetricia, y Ginecología, Universidad de Murcia, Murcia, Spain

^c Servicio de Cirugía General y Aparato Digestivo, Complejo Universitario de Santiago de Compostela, Spain

^d Cirugía Endocrina, Hospital Universitario de Bellvitge, L'Hospitalet de Llobregat, Barcelona, Spain

^e Sección de Cirugía Endocrino-Metabólica, Hospital Universitario Gregorio Marañón, Instituto de Investigación Biosanitaria Gregorio Marañón, Madrid, Spain

^f Servicio de Cirugía General, Aparato Digestivo y Trasplante de Órganos Abdominales, Hospital Universitario 12 de Octubre, Madrid, Spain

^g Servicio de Cirugía General y del Aparato Digestivo, Hospital Universitario Rey Juan Carlos, Móstoles, Madrid, Spain

^h Facultad de Ciencias de la Salud, Universidad Rey Juan Carlos, Alcorcón, Madrid, Spain

ⁱ FFIS, Fundación para la Formación e Investigación Sanitarias de la Región de Murcia, Murcia, Spain

^j Servicio de Cirugía General y del Aparato Digestivo, Hospital Torrecardenas, Almería, Spain

^k Servicio de Urología, Complejo Hospitalario Universitario de Canarias (CHUC), Santa Cruz de Tenerife, Spain

^l Departamento de Cirugía, Universidad de La Laguna (ULL), San Cristóbal de La Laguna, Santa Cruz de Tenerife, Spain

ARTICLE INFO

Article history:

Received 8 November 2022

Accepted 21 February 2023

Available online 20 October 2023

Keywords:

Multicenter clinical study

Infrequent carcinoma

Data-manager

Protocolization

Concordance

Database

A B S T R A C T

Introduction: In multicenter studies, the protocolization of data is a critical phase that can generate biases. The objective is to analyze the concordance and reliability of the data obtained in a clinical multicenter study between the protocolization in the center of origin and the centralized protocolization of the data by a data -manager.

Methods: National multicenter clinical study about an infrequent carcinoma. A double protocolization of the data is carried out: (a) center of origin; and (b) centralized by a data manager. The concordance between the data is analyzed for the global data and for the two groups of the project: (a) study group (Familiar carcinoma, 30 researchers protocolize); (b) control group (Sporadic carcinoma, 4 people protocolize). Interobserver variability is evaluated using Cohen's kappa coefficient.

Results: The study includes a total of 689 patients with carcinoma, 252 in the study group and 437 in the control group. Regarding the concordance analysis of the tumor stage, 2.5% of disagreements were observed and the concordance between people who protocolize was

* Corresponding author.

E-mail address: arzrios@um.es (A. Ríos).

<http://dx.doi.org/10.1016/j.cireng.2023.02.010>

2173-5077/© 2023 AEC. Published by Elsevier España, S.L.U. All rights reserved.

near perfect (Kappa = 0.931). Regarding the evaluation of the recurrence risk, disagreements occurred in 7% of the cases and the concordance was near perfect (Kappa = 0.819). Regarding the sonography evaluation (TIRADS), the disagreements were 6.9% and the concordance was near perfect (Kappa = 0.922). Also, 4.6% of transcription errors were detected.

Conclusions: In multicenter clinical studies, the centralized data protocolization by a data-manager seems to present similar results to the direct protocolization in the database in the center of origin.

© 2023 AEC. Published by Elsevier España, S.L.U. All rights reserved.

Protocolización de estudios clínicos multicéntricos en la era digital. ¿Tiene utilidad la centralización de datos por un data-manager?

RESUMEN

Palabras clave:

Estudio multicéntrico clínico
Carcinoma infrecuente
Data-manager
Protocolización
Concordancia
Base de datos

Introducción: En los estudios multicéntricos la protocolización de los datos es una fase crítica que puede generar sesgos, sobre todo en estudios clínicos con presupuesto limitado. El objetivo es analizar la concordancia y confiabilidad de los datos obtenidos en un estudio multicéntrico clínico entre la protocolización en el centro de origen y la protocolización centralizada mediante un data-manager.

Método: Estudio clínico multicéntrico de prevalencia nacional sobre un carcinoma familiar infrecuente, realizándose una doble protocolización de los datos: (a) en el centro de origen; y, (b) centralizada con un data-manager. La concordancia se analiza para el global de los datos y para los dos subgrupos del proyecto: (a) grupo a estudio (Carcinoma familiar. Protocolizan 30 investigadores); y (b) grupo control (Carcinoma esporádico. Protocolizan 4). Las diferencias interobservador se evalúan mediante el Índice de Kappa de Cohen.

Resultados: Se incluyen 689 pacientes, 252 del grupo a estudio y 437 del grupo control. Respecto al análisis de concordancia del estadio tumoral se han objetivado un 2,5% de discordancias, siendo alta la concordancia entre protocolizadores (Kappa = 0,931). Respecto a la valoración del riesgo de recidiva las discordancias fueron del 7% de los casos, siendo alta la concordancia (Kappa = 0,819). Respecto a la clasificación ecográfica TIRADS las discordancias son del 6,9% y la concordancia es alta (Kappa = 0,922). Se han detectado un 4,6% de errores de transcripción.

Conclusiones: En los estudios multicéntricos clínicos la protocolización centralizada de los datos por un data-manager parece presentar resultados similares a la protocolización directa en la base de datos en el centro de origen.

© 2023 AEC. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

Introduction

In rare pathologies, multicenter clinical studies are being promoted to achieve sufficiently large series over a relatively short period of time that would make it possible to draw representative conclusions.^{1,2} However, these studies also have drawbacks, particularly since patients are managed at different hospitals and with protocols that are not always the same, which generates a heterogeneity that may affect the results. Furthermore, remote work conducted by different researchers can lead to different biases.³

One step where biases can occur is data management.³ Currently, the trend is to generate an online cloud database

that allows all study centers to manage data simultaneously. However, several researchers are involved in the data entry process, and this may lead to increased heterogeneity when compiling the database. In contrast, centralized data entry, generally by a data manager, entails greater expense and has not been shown to be beneficial.³

In most clinical trials, especially drug trials, electronic data collection is done directly, using a database and data capture systems to provide greater data integrity and shorten the study time. However, these systems are not easy to implement⁴ since they require prior installation by the research center, user training and patient participation — all of which come at a cost. Therefore, these systems are not very cost-effective for multicenter clinical projects,^{1,4} as their cost and

the infrastructure required are not affordable. Furthermore, the collection of electronic data through a capture system is not free of risks, and its advantages can be lost when not used properly.⁵

What has become widespread in multicenter clinical projects (most without specific funding) is the use of online cloud databases for data processing. However, whatever potential biases or errors this online data collection system may entail has not been tested against a system of data entry by a data manager,^{6,7} especially when you consider that a reliable data storage system does not guarantee the reliability of data entry. Errors in data transcription have been reported, and fields near the end of the data entry page are associated with a higher number of errors. However, even in these instances, there are studies that indicate that control by a data manager can be considered an avoidable cost, while some authors suggest using an electronic control to monitor the reliability of data entry.⁶ Cost studies on the electronic data collection process in clinical trials have shown a reduction in collection costs of more than 50%, with savings ranging from 49 %-62 %, depending on the project.⁸

The aim of this study is to conduct a multicenter clinical study to analyze the agreement between the completion of the data collection protocol at the hospital of origin versus centralized data entry by a data manager in order to assess the reliability of the data.

Methods

Study design

We have conducted a nation-wide multicenter study in Spain, which has been endorsed by the Spanish Association of Surgeons (*Asociación Española de Cirujanos*, or AEC). This study was the first main objective of our project. The study included patients with familial cancer of low incidence among the population.

We designed a specific data collection protocol for this pathology to collect socio-personal, clinical, therapeutic and follow-up variables. This protocol included an attached report specifying the data to be collected for each of the variables. In addition, contact information was provided to resolve any questions.

Data collection

The data were collected at the center of origin by each of the researchers collaborating with the project, who complied with the project protocol. Said protocol may have been on paper or digitalized, but the data did not flow directly into the project database.

A double protocol was carried out.

- 1) *Protocol 1–Data entry at the center of origin:* Data are registered as they are obtained, following the protocol at the center of origin.

- 2) *Protocol 2–Centralized data entry by a data manager:* Data are registered in a centralized manner, by a single data manager, who is able to resolve any concerns about the analyzed variables with the study center of origin.

The agreement between the protocol conducted with the data from the center of origin and the centralized protocol using a data manager has been analyzed in 3 different groups:

- 1) *Group 1 - Overall study.* Agreement was analyzed for the total sample under study.
- 2) *Group 2–Familial cancer group.* Agreement in the familial cancer group was analyzed. In this group, 30 researchers participated in the collection of data at the study centers of origin, which were the different hospital units participating in the project.
- 3) *Group 3–Sporadic carcinoma control group.* Agreement was analyzed in the control group of patients with sporadic carcinoma. In this group, 4 researchers participated in the data entry at the center of origin, as these data came from only 4 hospital units.

Variables analyzed

Most socio-personal, clinical and therapeutic variables were excluded from the agreement analysis, where the presence of discordance is exceptional, except for transcription error. Three groups of variables were analyzed:

- 1) *Prognostic classification according to the TNM system* of the American Joint Committee on Cancer. Within this system, the following are analyzed:
 - 2) Size (T)
 - 3) Node involvement (N)
 - 4) Systemic involvement (M)
 - 5) TNM stage
- 6) *Assessment of the risk of recurrence according to the ATA* (American Thyroid Association), which proposes 3 risk groups:
 - 7) Low risk
 - 8) Moderate risk
 - 9) High risk
- 10) *Ultrasound assessment and TI-RADS* (Thyroid Imaging Reporting and Data System), according to the Korean TI-RADS classification.⁹ There are seven risk groups:
 - 11) TI-RADS 1
 - 12) TI-RADS 2
 - 13) TI-RADS 3
 - 14) TI-RADS 4a
 - 15) TI-RADS 4b
 - 16) TI-RADS 4c
 - 17) TI-RADS 5

Lastly, transcription errors detected during the data entry protocol process were assessed.

Statistical analysis

The statistical analysis was performed with a software package (SPSS, version 19.0 for Windows, SPSS, Chicago, Illinois). Interobserver differences were evaluated using Cohen's Kappa Index. The agreement coefficient has a value that ranges between 0 and 1, and its interpretation corresponds with the following ranges:

- 0 No agreement.
- 0.01–0.19 Very low agreement.
- 0.20 – 0.39 Low agreement.
- 0.40 – 0.59 Moderate agreement.
- 0.60 – 0.79 Good agreement.
- 0.80 – 0.99 High agreement.
- Perfect agreement.

P-values <.05 were considered statistically significant. There was agreement among the protocol data processors only if the correlation coefficient was greater than 0.00, with or without statistical significance.

For the comparison of groups 2 and 3, the Student's t test was applied. Fisher's exact test was applied when the contingency tables had cells with an expected frequency <5.

Results

General project characteristics

The study included a total of 689 patients with cancer, divided between the familial cancer study group (n = 252) and the sporadic cancer control group (n = 437).

Agreement analysis regarding the prognostic classification of the TNM system

1) Assessment of the T component of the TNM system

The data management protocol for the T component of the TNM system presented a disagreement rate of 4.2% (n = 29). In the familial carcinoma group 2, with 30 data entry researchers, the disagreement rate reached 6% (n = 15). Meanwhile, in the sporadic carcinoma control group 3, with 4 protocol data entry researchers, discrepancies reached 3.2% (n = 14) (Table 1). These differences were not statistically significant (P = 0.125).

Both from a global level (Kappa = 0.939) as well as in the familial cancer group 2 (Kappa = 0.920) and in control group 3 (Kappa = 0.950), the agreement among the data entry researchers was high (Table 2).

2) Assessment of the N component of the TNM system

Data entries for the N component (lymph node involvement) of the TNM system presented a disagreement rate of 0.7% (n = 5), all of which were from the familial cancer group 2 (2%; n = 5) (P = .003) (Table 1).

Both globally (Kappa = 0.979) as well as in the familial cancer group 2 (Kappa = 0.949), agreement among the protocol data entry operators was high, and agreement was perfect in control group 3 (Kappa = 1.000) (Table 3).

3) Assessment of the M component of the TNM system

Data entries for the M component (systemic involvement) of the TNM system presented perfect agreement among the 3 groups (Tables 1 and 4).

4) Assessment of the TNM stage

Data entries for the TNM stage presented a disagreement rate of 2.5% (n = 17). This percentage was greater in the familial

Table 1 – Percentage of disagreement between the initial assessment conducted at the hospital of origin and the final assessment centralized by a data manager.

Variables	Agreement	Disagreement	Percentage
Group 1 – Global series			
T	660	29	4.2%
N	684	5	0.7%
M	689	0	0%
Stage	672	17	2.5%
ATA Risk Assessment	641	48	7%
TI-RADS	549	41	6.9%
Transcription errors	657	32	4.6%
Group 2 – Familial cancer			
T	237	15	6%
N	247	5	2%
M	252	0	0%
Stage	242	10	4%
ATA Risk Assessment	234	18	7.1%
TI-RADS	201	14	6.5%
Transcription errors	240	12	4.8%
Group 3 – Control group			
T	423	14	3.2%
N	437	0	0%
M	437	0	0%
Stage	430	7	1.6%
ATA Risk Assessment	407	30	6.9%
TI-RADS	348	27	7.2%
Transcription errors	417	20	4.6%

Table 2 – Agreement between the initial assessment made at the hospital of origin and the final assessment centralized by a data manager on the assessment of the T component (size) of the TNM system.

Initial assessment at the Center of Origin	Centralized assessment by a Data Manager					
	T1a	T1b	T2	T3	T4	Total
Group 1 – Overall series						
Kappa coefficient for inter-rater reliability 0,939						
P < .0001						
T1a	306	3	0	5	0	314
T1b	7	144	6	1	0	158
T2	0	2	99	1	0	102
T3	0	0	2	105	0	107
T4	0	0	2	0	6	8
Total	313	149	109	112	6	689
Group 2 – Familial cancer						
Kappa coefficient for inter-rater reliability 0,920						
P < .0001						
T1a	74	3	0	3	0	80
T1b	4	65	1	0	0	70
T2	0	2	48	0	0	50
T3	0	0	2	48	0	50
T4	0	0	0	0	2	2
Total	78	70	51	51	2	252
Group 3 – Control group						
Kappa coefficient for inter-rater reliability 0,950						
P < .0001						
T1a	232	0	0	2	0	234
T1b	3	79	5	1	0	88
T2	0	0	51	1	0	52
T3	0	0	0	57	0	57
T4	0	0	2	0	4	6
Total	235	79	58	61	4	437

cancer group 2 (4%; n = 10) than in the sporadic carcinoma control group 3 (1.6%; n = 7), although said differences were not significant (P = .094) (Table 1).

Both globally (Kappa = 0.931) as well as in the familial cancer group 2 (Kappa = 0.904) and in control group 3 (Kappa = 0.950), agreement among the data entry managers was high (Table 5).

Agreement analysis for the ATA risk of recurrence staging system

Data entries for the ATA risk of recurrence classification presented a disagreement rate of 7% (n = 48). In the familial cancer group 2, the discrepancies represented 7.1% (n = 18),

Table 3 – Agreement between the initial assessment at the study center of origin and the centralized final assessment by a data manager regarding the N component (lymph node involvement) of the TNM system.

Initial assessment at the Center of Origin	Assessment by a Data Manager			
	N0	N1a	N1b	Total
Group 1 – Overall series				
Kappa coefficient for inter-rater reliability 0,979				
P <.0001				
N0	541	0	0	541
N1a	5	95	0	100
N1b	0	0	48	48
Total	546	95	48	689
Group 2 – Familial cancer				
Kappa coefficient for inter-rater reliability 0,949				
P <.0001				
N0	189	0	0	189
N1a	5	39	0	44
N1b	0	0	19	19
Total	194	39	19	252
Group 3 – Control group				
Kappa coefficient for inter-rater reliability 1,000				
P <.0001				
N0	352	0	0	352
N1a	0	56	0	56
N1b	0	0	29	29
Total	352	56	29	437

Table 4 – Agreement between the initial assessment at the study center of origin and the centralized final assessment by a data manager regarding the M component M (systemic involvement) of the TNM system.

Initial assessment at the Center of Origin	Assessment by a Data Manager		
	M0	M1	Total
Group 1 – Overall series			
Kappa coefficient for inter-rater reliability 1,000			
P < .0001			
M0	687	0	687
M1	0	2	2
Total	687	2	689
Group 2 – Familial cancer			
Kappa coefficient for inter-rater reliability 1,000			
P < .0001			
M0	251	0	251
M1	0	1	1
Total	251	1	252
Group 3 – Control group			
Kappa coefficient for inter-rater reliability 1,000			
P < .0001			
M0	436	0	436
M1	0	1	1
Total	436	1	437

while in the sporadic carcinoma control group 3, the discrepancies represented 6.9% (n = 30) of the cases (Table 1). These differences were not significant (P = .890).

Both at a global level (Kappa = 0.819) as well as in the familial carcinoma group 2 (Kappa = 0.829) and in control group 3 (Kappa = 0.813), agreement among the data entry operators was high (Table 6).

Table 5 – Agreement between the initial assessment at the study center of origin and the centralized final assessment by a data manager regarding staging according to the TNM system.

Initial assessment at the Center of Origin	Assessment by a Data Manager				
	I	II	III	IV	Total
Group 1 – Overall series					
Kappa coefficient for inter-rater reliability 0,931					
P < .0001					
I	540	4	3	0	547
II	1	44	1	0	46
III	3	0	59	5	67
IV	0	0	0	29	29
Total	544	48	63	34	689
Group 2 – Familial cancer					
Kappa coefficient for inter-rater reliability 0,904					
P < .0001					
I	186	2	1	0	189
II	1	24	0	0	25
III	2	0	24	4	30
IV	0	0	0	8	8
Total	189	26	25	12	252
Group 3 – Control group					
Kappa coefficient for inter-rater reliability 0,950					
P < .0001					
I	354	2	2	0	358
II	0	20	1	0	21
III	1	0	35	1	37
IV	0	0	0	21	21
Total	355	22	38	22	437

Agreement analysis for the TI-RADS ultrasound classification

The assessment of the TI-RADS ultrasound classification presented a disagreement rate of 6.9% (n = 41). In the familial

Table 6 – Agreement between the initial assessment at the study center of origin and the centralized final assessment by a data manager regarding the ATA risk of recurrence.

Initial assessment at the Center of Origin	Assessment by a Data Manager			
	Low risk	Moderate risk	High risk	Total
Group 1 – Overall series				
Kappa coefficient for inter-rater reliability 0,819				
P < .0001				
Low risk	498	40	1	539
Moderate risk	1	127	6	134
High risk	0	0	16	16
Total	499	167	23	689
Group 2 – Familial cancer				
Kappa coefficient for inter-rater reliability 0,829				
P < .0001				
Low risk	177	13	0	190
Moderate risk	1	46	4	51
High risk	0	0	11	11
Total	178	59	15	252
Group 3 – Control group				
Kappa coefficient for inter-rater reliability 0,813				
P < .0001				
Low risk	321	27	1	349
Intermediate risk	0	81	2	83
High risk	0	0	5	5
Total	321	108	8	437

Both overall (Kappa = 0.922) as well as in the familial cancer group 2 (Kappa = 0.921) and in control group 3 (Kappa = 0.922), agreement among the data processors was high (Table 7).

Transcription errors have been detected in 4.6% (n = 32) of cases, with similar percentages found in group 2 (4.8%; n = 12) and group 3 (4.6%; n = 20) ($P = .939$) (Table 1).

In clinical trials and multicenter studies, data collection notebooks or registers are basic tools used by researchers to

Regardless of the type of data management, it is important to ensure that information is entered into the database in a consistent and accurate manner,¹³⁻¹⁵ as even a small rate of error can have a big impact on the results of a study.¹⁶ Random or systematic errors that go unnoticed are sources of bias in the database. Furthermore, the availability of reliable electronic systems is not enough to guarantee the validity of

Initial assessment at the Center of Origin	Assessment by a Data Manager							
	TI-RADS 1	TI-RADS 2	TI-RADS 3	TI-RADS 4a	TI-RADS 4b	TI-RADS 4c	TI-RADS 5	Total
Group 1 – Overall series								
Kappa coefficient for inter-rater reliability 0,922								
P <.0001								
TI-RADS 1	0	1	0	0	0	0	0	1
TI-RADS 2	0	34	3	0	0	0	0	37
TI-RADS 3	0	2	27	0	0	0	0	29
TI-RADS 4a	0	0	0	104	2	0	0	106
TI-RADS 4b	0	0	0	1	155	19	0	175
TI-RADS 4c	0	0	1	0	8	211	4	224
TI-RADS 5	0	0	0	0	0	0	18	18
Total	0	37	31	105	165	230	22	590
Missing Data								99
Group 2 – Familial cancer								
Kappa coefficient for inter-rater reliability 0,921								
P <.0001								
TI-RADS 1	0	0	0	0	0	0	0	0
TI-RADS 2	0	12	0	0	0	0	0	12
TI-RADS 3	0	0	1	0	0	0	0	1
TI-RADS 4a	0	0	0	43	0	0	0	43
TI-RADS 4b	0	0	0	1	45	9	0	56
TI-RADS 4c	0	0	0	0	3	91	1	95
TI-RADS 5	0	0	0	0	0	0	9	9
Total	0	12	1	44	48	100	10	215
Missing Data								37
Group 3 – Control group								
Kappa coefficient for inter-rater reliability 0,922								
P <.0001								
TI-RADS 1	0	1	0	0	0	0	0	1
TI-RADS 2	0	22	3	0	0	0	0	25
TI-RADS 3	0	2	26	0	0	0	0	28
TI-RADS 4a	0	0	0	61	2	0	0	63
TI-RADS 4b	0	0	0	0	110	10	0	120
TI-RADS 4c	0	0	1	0	5	120	3	129
TI-RADS 5	0	0	0	0	0	0	9	9
Total	0	25	30	61	117	130	12	375
Missing Data								62

population-based cross-sectional studies. In fact, the relevance of a study largely depends on 2 main factors: quality data collection from the medical records, and the reliability of the data transferred from the medical records to the electronic system. Any weakness in these 2 stages will invalidate the study.¹⁶⁻¹⁹ Thus, the importance of training clinical researchers must be emphasized, particularly when it comes to training in good clinical practices.²⁰⁻²²

For all these reasons, although multicenter studies offer the advantage of obtaining information on rare pathologies in a relatively short period of time, they are more heterogeneous studies and have a greater risk of biases.^{1,3,22-24} Therefore, a balance must be sought between minimizing the potential biases of a study while not increasing the costs, which would make them unviable. In this context, the concept of centralized data entry by a data manager implies a more homogeneous protocol. However, it also entails a greater cost for the research project, which is sometimes unaffordable,^{1,3,22-24} and usually implies a longer data entry period versus when the protocol is followed independently by the different research groups.

Our multicenter project, which was conducted within the context of a national scientific society in Spain, shows how an adequate study design, detailed protocol, and complete explanation of the implications of each variable resulted in an acceptably high rate of agreement. Nevertheless, even though agreement was high (and in some cases perfect) for almost all the variables analyzed, the detailed analysis of the data revealed a percentage of disagreement that reached 7% for certain variables. Furthermore, these discrepancies were accentuated when the number of data processors was higher. Thus, we observed more disagreement in the familial cancer group 2, where there were 30 researchers entering data, versus the sporadic carcinoma control group 3, with only 4 researchers entering data.

In addition, the discrepancies mainly correlated with variables whose determination involves some type of nuance. For instance, agreement rates were almost perfect for concrete variables like lymph node involvement (N) or the presence of metastasis (M), with non-agreement rates of 0.7% and 0%, respectively. For the "T" variable, however, which is not only determined by size but also by factors such as the impact on neighboring structures, the non-agreement rate was greater than 4%. This situation became more obvious for variables where the classification was even more complex. Thus, we observed a disagreement rate of 7% for the ATA recurrence risk assessment and 6.9% in the TI-RADS ultrasound classification.

Therefore, although the agreement between direct data entry at the hospital of origin and centralized data processing by a data manager is high, there is a percentage of disagreement for important prognostic variables that ranges from 2.5% for TNM tumor stage to 7% for ATA the risk assessment. This implies a percentage of error or bias only in the process of data entry in multicenter studies. Thus, for these variables and although agreement is high, it is at the lower limit of said agreement — close to what is considered merely "good agreement". To this, we must add the transcription error rate of 4.6%, as detected by the data manager.

Complying with data processing protocols is a slow and sometimes tedious process, and it is not always done by the main researcher at the hospital of origin, as the work is frequently delegated to others. All of this leads to possible errors in transcription. We should emphasize that the objective of this study is not to detect transcription errors, which suggests that this percentage may possibly be higher, since only evident data processing errors were detected. In order to detect all errors, the protocol variables would have to be reviewed in their entirety.

Hence, data collection must be well supervised to improve study quality. Although the use of a data manager makes no difference, the transcription error rate was nearly 5%. Therefore, it is necessary to create clear data entry protocols that are easy to complete, and we recommend that an experienced member of the data collection team review the main variables of the study. Otherwise, in the design phase of the study, the database should compile fewer variables but of greater quality and interest for the analysis.

We agree with other authors in that, if the staff members who log the data are well trained and supervised, this is not the weakest link in the data management chain,^{6,25,26} even though we have shown that this process does entail a margin of error.

One factor that our project has not corroborated (possibly because it is a short protocol) is that the position of the data entry field in the protocol may play an important role in the proportion of errors. The last few positions are associated with more errors than the initial fields, especially when the fields are numerical. This has been attributed to data manager fatigue when questionnaires have too many fields to complete.⁶ These results therefore suggest that, to create more effective questionnaires, the most important information should be collected in the first few fields, the number of fields should be reduced, and "combo boxes" or text boxes should be used instead of fields with direct numerical insertion (especially in the last part of the questionnaire).

A limitation of this study is that it was based on a single data manager, making it difficult to generalize our conclusions. However, we should reiterate that the decision to use a single data manager was made to improve the reliability of data entry by reducing bias among data entry operators. To this point, it would be interesting to be able to compare our outcomes with other multicenter studies in other contexts for external validation of the results. Lastly, the quality of the data in multicenter studies with no data manager should be assessed, specifically regarding who collects the data (administrative staff, residents, etc) and under whose supervision, since these factors can influence the results, in which case the presence of a data manager could present differences.

Based on the data from our study, the existence of a data manager is not necessary given the high agreement between the data entry types, which represents savings in both time and costs.⁸ The drawbacks that online data processing may present can be resolved (as in our study) with standard databases that do not require direct costs or maintaining an online system, although it does not present the advantages offered by online products in terms of quality control. Nonetheless, they may be sufficient for multicenter projects like the one evaluated in our study. All these factors are

important in multicenter studies of rare pathologies, where the contribution of cases is important, but at the same time funding is limited or even non-existent. In a Phase II Clinical Trial, data management costs can be almost 30% of the costs of Phase III clinical trials,⁸ and this added expense will subsequently impact the costs of marketed drugs. In multicenter clinical studies, these costs will determine the viability of the project itself.²⁷

Lastly, it should be noted that, although there are no differences, and a data manager may not be necessary to log the data, it should be mentioned that data managers usually carry out other functions over the course of a project, such as coordinating patient testing, resolving problems that may arise in the participating study centers, serving as a liaison among researchers, etc. Therefore, if the budget of a project is able to cover the cost, hiring a data manager usually speeds up the completion of a project.

In conclusion, centralized data processing by a data manager in multicenter clinical studies seems to present similar results compared to direct data entry in the database at the hospital of origin.

Ethics approval

The study protocol was approved by institute's committee.

Ethics Committee Code: 2021-2-13-HCUVA.

This human study has been reviewed by the appropriate ethics committee and has therefore been performed in accordance with the ethical standards laid down in the 2000 Declaration of Helsinki as well as the Declaration of Istanbul 2008.

Funding

No specific funding sources have been used for this study.

Précis

In multicenter clinical studies, the existence of a data manager is not required and the protocolization can be performed directly in the database in the center of origin due to the near perfect concordance between the two protocolization systems.

Authorship: Contribution to study by the Authors

- 1 Conception and design: Ríos A.
- 2 Acquisition of a substantial portion of data: Ríos A, Ferrero-Herrero E, Puñal-Rodríguez JA, Duran M, Mercader-Cidoncha E, Ruiz-Pardo J, Rodríguez JM.
- 3 Analysis and interpretation of data: Ríos A, Gutiérrez PR.
- 4 Drafting of the manuscript: Ríos A.
- 5 Critical revision of the manuscript for important intellectual content: Ríos A, Gutiérrez PR, Mercader-Cidoncha E.
- 6 Statistical expertise: Ríos A.
- 7 Obtaining funding for this project or study: Ríos A.

- 8 Supervision: Ríos A, Ferrero-Herrero E, Puñal-Rodríguez JA, Gutiérrez PR, Durán M. Mercader-Cidoncha E, Ruiz-Pardo J, Rodríguez JM.
- 9 Final approval of the version to be published: Ríos A, Ferrero-Herrero E, Puñal-Rodríguez JA, Gutiérrez PR, Mercader-Cidoncha E, Ruiz-Pardo J, Rodríguez JM.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

Thank the Spanish Association of Surgeons and the Endocrine Surgery Section for their support in carrying out this work. Thank the 24 Spanish Endocrine Surgery Units for having collaborated in the realization of this project and for having provided the data.

REFERENCES

1. Serra-Aracil X, Pascua-Sol M, Badia-Closa J, Navarro-Soto S. Cómo poner en marcha y desarrollar un estudio multicéntrico prospectivo, controlado y aleatorizado. *Cir Esp.* 2020;98:119-26. <http://dx.doi.org/10.1016/j.ciresp.2019.11.012>.
2. Hemminki K, Eng C, Chen B. Familial risks for nonmedullary thyroid cancer. *J Clin Endocrinol Metab.* 2005;90:5747-53. <http://dx.doi.org/10.1210/jc.2005-0935>.
3. Pannucci CJ, Wilkins EG. Identifying and avoiding bias in research. *Plast Reconstr Surg.* 2010;126:619-25. <http://dx.doi.org/10.1097/PRS.0b013e3181de24bc>.
4. Welker JA. Implementation of electronic data capture systems: barriers and solutions. *Contemp Clin Trials.* 2007;28:329-36. <http://dx.doi.org/10.1016/j.cct.2007.01.001>.
5. Schmier JK, Kane DW, Halpern MT. Practical applications of usability theory to electronic data collection of clinical trials. *Contemp Clin Trials.* 2005;26:375-85. <http://dx.doi.org/10.1016/j.cct.2007.01.001>.
6. Mauri D, Karampoiki V, Mauri J, Kamposioras K, Alexiou G, Ferentinos G, Peponi C (SESy Europe Task Force). et al. Double-blind control of the data manager doesn't have any impact on data entry reliability and should be considered as an avoidable cost. *BMC Med Res Methodol.* 2008;8:66. <http://dx.doi.org/10.1186/1471-2288-8-66>.
7. Howells K. Living without a data management system. *IDrugs.* 2008;11:733-7.
8. Pavlović I, Kern T, Miklavcic D. Comparison of paper-based and electronic data collection process in clinical trials: costs simulation study. *Contemp Clin Trials.* 2009;30:300-16. <http://dx.doi.org/10.1016/j.cct.2009.03.008>.
9. Kwak JY, Han KH, Yoon JH, Moon HJ, Son EJ, Park SH, et al. Thyroid imaging reporting and data system for US features of nodules: a step in establishing better stratification of cancer risk. *Radiology.* 2011;260:892-9. <http://dx.doi.org/10.1148/radiol.11110206>.
10. Lee N, Millman A, Osborne M, Cox J. ABC of medical computing. Storing and managing data on a computer. *BMJ.* 1995;311:562-5. <http://dx.doi.org/10.1136/bmj.311.7004.562>.

11. Millman A, Lee N, Brooke A. ABC of medical computing. Computers in general practice – I. BMJ. 1995;311:800–2. <http://dx.doi.org/10.1136/bmj.311.7008.800>.
12. Brandt CA, Argraves S, Money R, Ananth G, Trocky NM, Nadkarni PM. Informatics tools to improve clinical research study implementation. Contemp Clin Trials. 2006;27:112–22. <http://dx.doi.org/10.1016/j.cct.2005.11.013>.
13. Los RK, van Ginneken AM, Roukema J, Moll HA, Lei van der J. Why are structured data different? Relating differences in data representation to the rationale of OpenSDE. Med Inform Internet Med. 2005;30:267–76. <http://dx.doi.org/10.1080/14639230500367563>.
14. de Lusignan S. The barriers to clinical coding in general practice: a literature review. Med Inform Internet Med. 2005;30:89–97. <http://dx.doi.org/10.1080/14639230500298651>.
15. Marks RG. Validating electronic source data in clinical trials. Control Clin Trials. 2004;25:437–46. <http://dx.doi.org/10.1016/j.cct.2004.07.001>.
16. Mullooly JP. The effects of data entry error: an analysis of partial verification. Comput Biomed Res. 1990;23:259–67. [http://dx.doi.org/10.1016/0010-4809\(90\)90020-D](http://dx.doi.org/10.1016/0010-4809(90)90020-D).
17. Levitt SH, Aeppli DM, Potish RA, Lee CK, Nierengarten ME. Influences on inferences: effect of errors in data on statistical evaluation. Cancer. 1993;72:2075–82. [http://dx.doi.org/10.1002/1097-0142\(19931001\)72:7<2075](http://dx.doi.org/10.1002/1097-0142(19931001)72:7<2075).
18. Arndt S, Tyrrell G, Woolson RF, Flaum M, Andreasen NC. Effects of errors in a multicenter medical study: preventing misinterpreted data. J Psychiatr Res. 1994;28:447–59. [http://dx.doi.org/10.1016/0022-3956\(94\)90003-5](http://dx.doi.org/10.1016/0022-3956(94)90003-5).
19. Crombie IK, Irving JM. An investigation of data entry methods with a personal computer. Comput Biomed Res. 1986;19:543–50. [http://dx.doi.org/10.1016/0010-4809\(86\)90028-5](http://dx.doi.org/10.1016/0010-4809(86)90028-5).
20. Saleh M, Naik G, Jester P, Joiner C, Westfall E, Kimberlin DW, et al. Clinical Investigator Training Program (CITP) - A practical and pragmatic approach to conveying clinical investigator competencies and training to busy clinicians. Contemp Clin Trials Commun. 2020;19100589. <http://dx.doi.org/10.1016/j.conctc.2020.100589>.
21. Swezey T, McGuire FH, Hurley P, Panhuis J, Goldstein K, Chuck T, et al. More than a box to check: research sponsor and clinical investigator perspectives on making GCP training relevant. Contemp Clin Trials Commun. 2020;19100606. <http://dx.doi.org/10.1016/j.conctc.2020.100606>.
22. Bechtel J, Chuck T, Forrest A, Hildebrand C, Panhuis J, Pattee SR, et al. Improving the quality conduct and efficiency of clinical trials with training: recommendations for preparedness and qualification of investigators and delegates. Contemp Clin Trials. 2020;89105918. <http://dx.doi.org/10.1016/j.cct.2019.105918>.
23. Huang GD, Bull J, McKee KJ, Mahon E, Harper B, Roberts JN. Clinical trials recruitment planning: a proposed framework from the Clinical Trials Transformation Initiative. Contemp Clin Trials. 2018;66:74–9. <http://dx.doi.org/10.1016/j.cct.2018.01.003>.
24. Chung KC, Song JW. A Guide on Organizing a Multicenter Clinical Trial: the WRIST study group. Plast Reconstr Surg. 2010;126:515–23. <http://dx.doi.org/10.1097/PRS.0b013e3181df64fa>.
25. Day S, Fayers P, Harvey D. Double data entry: what value, what price? Control Clin Trials. 1998;19:15–24. [http://dx.doi.org/10.1016/S0197-2456\(97\)00096-2](http://dx.doi.org/10.1016/S0197-2456(97)00096-2).
26. Reynolds-Haertle RA, McBride R. Single vs. double data entry in CAST. Control Clin Trials. 1992;13:487–94. [http://dx.doi.org/10.1016/0197-2456\(92\)90205-E](http://dx.doi.org/10.1016/0197-2456(92)90205-E).
27. Ríos A, Rodríguez MA, Puñal JA, Moreno P, Mercader E, Ferrero E, et al. Biological behavior of familial papillary thyroid microcarcinoma: Spanish multicenter study. Langenbecks Arch Sur. 2022. <http://dx.doi.org/10.1007/s00423-022-02704-4>. Epub ahead of print. PMID: 36251077.