

Measuring Workload of Clinical Trials: Transcultural Adaptation and Validation to Portuguese Language of Ontario Protocol Assessment Level (OPAL)

Roberta Monteiro Batista Sarmento[†] and Zenith Rosa Silvino

Aurora de Afonso Costa Nursing School, National Cancer Institute-Brazil (INCA), Fluminense Federal University, Niterói, RJ, Brazil

*Corresponding author: Roberta Monteiro Batista Sarmento, Aurora de Afonso Costa Nursing School, National Cancer Institute-Brazil (INCA), Fluminense Federal University, Niterói, RJ, Brazil, Tel: +55 (21) 3207-3954; E-mail: rsarmento@inca.gov.br

Received date: May 22, 2017; Accepted date: July 15, 2017; Published date: July 17, 2017

Copyright: © 2017 Sarmento RMB, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Introduction: The increase number of clinical protocols with different requirements and specificities, the demand for the quality of the data according to the good clinical practices evidence the need of an instrument capable of measuring the workload of clinical protocols, and to assist the management of research centers. The object of study is the instrument entitled the Ontario Protocol Assessment Level, created for measuring the workload of the research coordinator, focusing on the complexity of clinical protocols in oncology.

Aim: To perform a transcultural adaptation and validation of the instrument in terms of the Portuguese language.

Method: This is a methodological research, whose chosen scenario was the clinical research center of the Brazilian National Cancer Institute, located in Rio de Janeiro. The subjects were the clinical research coordinators. The research was approved by the ethics committee, under protocol 120.006.

Results: A significantly high degree of agreement between intra- and inter-observers was established; the agreement of the committee of specialists (the golden standard) was considered to be excellent ($ICC > 0.949$) in both research periods (1 and 2); this score demonstrates a high level of validation. The analytical process confirmed that the tool score did not overestimate nor underestimate the evaluation of the committee of specialists.

Conclusion: The instrument was considered valid and reliable based on the statistical tests performed. It provides the support required to calculate the workload generated by clinical protocols.

Keywords: Workload; Clinical protocols; Oncology

Introduction

Clinical research in Brazil has seen a significant amount of growth in the past few years. This is intended to produce information that could lead to an understanding of the mechanisms related to the promotion of health, to the prevention of illness and with regard to the therapy associated with health issues, as these are all essential to clinical decision-making process [1].

For this reason, it is extremely important to ensure the education of the professionals involved in the conduction of clinical studies, and the generation of information. Thus, the construction of indicators is essential as an instrument to verify the quality of the work, to support the evaluation of performances, and to determine changes in the processes used so far [2].

A fundamental part of the team of a research center is the clinical research coordinator (CRC). This person is responsible for taking care of all of the logistics, so that research undertaken in the center progresses as described by the research protocols. The CRC is responsible for the operational support of the tasks related to research projects involving human beings, fulfilling all the methodological and ethical mandatory requirements, ending when reliable results are obtained, and guaranteeing the wellbeing of the subjects of the research engaged in by center staff [3].

In a study investigating the role of the CRC, it was observed that in the Brazilian National Cancer Institute (INCA, in Portuguese), these professionals performed a higher number of specific tasks when compared to CRC internationally. In this institution, the role of the CRC ranges from data collection and administrative support, to the guidance of service quality based on data management, all of which results in quality research, leading to professional and institutional recognition [4].

With regard to the multiple roles the CRC have at INCA, and the large number of open clinical protocols, the development of indicators to analyze quality, and the creation of an instrument that enables the evaluation of coordinators' workload without compromising the quality of their data would have a great relevance for the institution.

Based on a review of the literature, the Ontario Institute for Cancer Research has found a lack of information regarding an understanding of the complexity of workload in clinical research. In order to solve this issue, the Institute developed the Ontario Protocol Assessment Level (OPAL), an instrument capable of measuring clinical protocols workload, not only based on the number of participants, but also through an evaluation of the degree of complexity of such protocols [5].

In view of the lack of any instrument in Brazil that can determine the workload of CRCs focusing on the complexity of oncology clinical protocols, this study aimed to create a transcultural adaptation of the

OPAL instrument, and to verify its validity and reliability at INCA's clinical research center.

With this study, the authors intend to contribute to a better distribution of oncology clinical protocols among the CRCs, helping to provide reliable data to promote the fulfillment of ethical and regulatory demands, and to enable the mapping and possible redirection of workload among the members of the team, which may include the need to better train the professionals involved.

Materials and Methods

This is a methodological study, aimed at the development, validation and evaluation of an instrument in terms of its transcultural adaptation. The study was authorized by the main author of the instrument, and performed at unit III of INCA (HCIII, in Portuguese), located in the city of Rio de Janeiro, Brazil. This center is recognized as a center of excellence in the elaboration and conduct of innovative oncologic clinical studies.

After reading all the material made available by the author, and the definition of the necessary items to be used with the instrument that would be part of the transcultural adaptation, the following stages were followed, based on the proposals of Guillemin, Bombardier and Beaton [6]:

(A) Translation – performed by two Portuguese-native translators. One was informed about the motivation behind the research, while the other was not informed about the goals. This step generated a synthesis of the two translated versions which was then used in a back translation.

(B) Evaluation of the back translation – once the previous step was complete, the material was translated back again by English-native translators, both unaware of the propositions of the research, thereby generating two different back translation versions.

(C) Review by a specialists' committee – composed of two nurses, two physicians, and one speech therapist, all bilingual and with professional experience in oncology and clinical research, the committee evaluated all the versions of the instrument – the original, the translation and the back translation, based on the analysis of semantic, idiomatic, cultural and conceptual equivalences. The semantic equivalence is related to the adequacy to the meaning of the words used in terms of vocabulary and grammar; the idiomatic equivalence is related to the adequacy of colloquial and idiomatic expressions; the experimental one is related to situations that are not coherent with the cultural context; and the conceptual view validates the concept used.

In order to perform this step, the translation reports were sent by the translators to the specialists' committee. The specialists were given a spreadsheet to describe the suggestions with regard to changing certain items that were under scrutiny. All questions that emerged during the meetings were answered through teleconferences with the author of the original instrument. The final product arose through the consensus of at least 80% of the participants, thus generating the final version of the instrument.

(D) Pre-test – to test the adapted instrument, 15 fictitious clinical protocols of different complexities and different numbers of patients, were created based on existing scenarios. In a meeting with the specialists' committee, the final version of the instrument was applied with the intention of finding workload scores, defined as the golden

standard, to be achieved in evaluation tests of the psychometric capacity of the instrument by the clinical team of research coordinators of the hospital.

Analysis of the instrument

The final version of the instrument was submitted for analysis in terms of comprehensibility using a sample of the target population, and submitted to tests aimed at validating the psychometric capacity of the instrument. In this stage, four nurses CRCs from the HCIII/INCA were included. These evaluated the reliability (test-retest and inter-observer) and validated the content of the instrument.

The reliability of the test-retest process is the capacity of the instrument to generate the same results when used twice with regard to the same subjects, with an interval of one or two weeks, and then comparing the paired results. In this study, it was observed that the reuse of the OPAL evaluation score instrument in terms of the same clinical protocol by the same group of professionals, with an interval of seven to fourteen days between the interviews led to similar results.

The reliability of the inter-observing process relates to the capacity of the instrument with regard to generating the same results when used on different subjects. It was evaluated by the simultaneous application of the instrument by different professionals, all using the OPAL score results for the same clinical protocol, independently of each other.

The degree of validity is defined as a level at which the data demonstrate the elements it was designed to measure. In other words, the results of the measurement correspond to the real state of the phenomenon being evaluated. It includes the validity of the content, the construct, and the criterion [7].

The validity of the content relates to the capacity to measure all aspects of the studied phenomenon. In this analysis, the ruling of the specialists' committee members was taken into consideration with regard to the final version of the instrument.

The validation of the construct is one of the most important aspects when it comes to evaluating the psychometric characteristics of an instrument. However, it is also the most complex and hard to identify. An instrument with a good level of validation in terms of its construct guarantees the evaluation of a theory or of a hypothesis being investigated or, in other words, it is related to the ability of the instrument to confirm the hypothesis under consideration [8]. To evaluate it, the results of scores generated by the research coordinators were compared to the values of the scores generated by the specialists' committee, considered as the golden standard for the process of validation.

The validity of the criterion is the level by which the measurement is related to already existing and well-accepted standards. In the case of this instrument, no previous instrument could be found that could be compared to it, therefore such an analysis was not attempted in the present study.

Data analysis

The results of the descriptive analysis were presented in the form of tables, and expressed by averages, standard deviations, median, minimum and maximum for all the numeric data created.

The inferential analysis consisted of methods of concordance: the intra-class correlation coefficient (ICC), attempted to verify if there was a significant concordance of scores in terms of the different

observed items, and for which the trust interval used was 95%; while a graphic analysis based on Bland Altman [9] for the OPAL score. This developed into the calculation of differences of intra- and inter-observer scores against the average result.

The association between the OPAL score and the corresponding qualitative classification was analyzed by Kruskal-Wallis' ANOVA and by the Dunn (non-parametric) multiple comparisons test. The criterion for determining significance was 5%. Statistical analysis was performed using statistical software SPSS version 20.0.

Ethical aspects

This research was approved by the Committee of Ethics in Research of INCA, under protocol number 120.006. An informed consent form was given to all participants in this study. With regard to studies regarding human beings, all ethical and legal precepts demanded in

Brazil were followed, as described under the Brazilian Ministry of Health Resolution 466/12 [10]. There are no interest conflict to be declared.

Results

Table 1 demonstrates the central values and the dispersion of OPAL score tendencies (in absolute points) of the workload found in the sample, composed of 15 clinical protocols. These numbers were analyzed with regard to two occasions (M1 and M2). It is possible to see that the average and median results are close (reflecting a similar magnitude), t around 5.5 to 6.0 points; the interquartile amplitude (IQA) was found to be 50% of the sample, which varied between 4.0 and 8.5 points; while 100% of the sample was found to be between 1.5 and 9.5 points (minimum and maximum).

Observer	Moment	Average	SD	Median	IQA			Minimum	Maximum
CRC 1	M1	6	2.2	5.5	4	-	8.5	1.5	9
	M2	6.2	1.8	5.5	5.5	-	8.5	3.5	9
CRC 2	M1	6	2.2	5.5	4	-	8.5	1.5	9
	M2	6.2	1.9	5.5	5	-	8.5	1.5	9
CRC 3	M1	6	2.1	5.5	4.5	-	8	3	9.5
	M2	6.1	2.1	5.5	5	-	8.5	1.5	9
CRC 4	M1	6	2.1	5.5	4.5	-	8.5	1.5	9
	M2	6.2	2.2	5.5	5	-	8.5	1.5	9
Golden Standard		6.1	2.1	5.5	5	-	8.5	1.5	9

Source: Study results, 2014

SD: standard deviation; IQA: interquartile amplitude (1st quarter – 3rd quarter).

Table 1: Descriptive analysis of the Ontario Protocol Assessment Level score for the workload in 15 protocols. Rio de Janeiro, Brazil, 2012.

The intra- and inter-observer reliability was evaluated through the ICC, which worked on the possibility of a significant agreement between the two OPAL scores during the two occasions, and also among the nurses of the specialists' committee (golden standard).

Table 2 demonstrates the ICC, its confidence interval of 95% (CI of 95%), and the descriptive level (p value) for the total OPAL score (in points) of the intra- and inter-observers analysis.

A significantly high intra and inter-observing concordance (p<0.0001) was observed among all pairs of nurses. In the intra-observer analysis, CRC 3 presented a higher level of concordance

(ICC=0.987) which is almost perfect, followed by CRC 2 (ICC=0.974) and CRC 4 (ICC=0.970), and finally, by CRC 1 (ICC=0.934), expressing, in a general sense, a high reproducibility (reliability) in terms of this score.

On the other hand, the inter-observer analysis found that the concordance compared to the specialists' committee (the golden standard) achieved an excellent level (ICC>0.949) during both occasions M1 and M2, demonstrating a high level of legitimacy (validity) in terms of the score.

Analysis	Observer	ICC	CI of 95%			p value
Intra-observer M1 x M2	CRC 1 x CRC 1	0.934	0.82	-	0.977	<0.0001
	CRC 2 x CRC 2	0.974	0.926	-	0.991	<0.0001
	CRC 3 x CRC 3	0.987	0.963	-	0.996	<0.0001
	CRC 4 x CRC 4	0.97	0.915	-	0.99	<0.0001

Intra-observer between M1 Nurse x Committee	CRC 1 x Committee	0.993	0.979	-	0.998	<0.0001
	CRC 2 x Committee	0.993	0.979	-	0.998	<0.0001
	CRC 3 x Committee	0.989	0.968	-	0.996	<0.0001
	CRC 4 x Committee	0.998	0.995	-	0.999	<0.0001
	CRC 1 x Committee	0.949	0.86	-	0.983	<0.0001
	CRC 2 x Committee	0.981	0.947	-	0.994	<0.0001
	CRC 3 x Committee	0.998	0.995	-	0.999	<0.0001
	CRC 4 x Committee	0.971	0.919	-	0.99	<0.0001

Source: Study results, 2014

ICC: intraclass correlation coefficient; **CI of 95%:** confidence interval of 95% for ICC

Table 2: Intra- and inter-observers analysis of the Ontario Protocol Assessment Level score. Rio de Janeiro, Brazil, 2012.

Among the CRC, CRC 1 was the one who presented the lowest level of concordance with the golden standard (ICC=0.949) with a wider confidence interval against ICC (CI 95%: 0.860-0.983) during occasion M2.

Table 3 demonstrates the average, standard deviation and the inferior and superior limits of the absolute differences in OPAL scores among nurses (CRC) and the specialists' committee (the golden standard), in terms of occasion M1. The limits (inferior and superior) are the "limits of concordance" according to Bland and Altman [9].

Observer	n	Average ^a	SD	IL 95%	-	SL 95%
CRC 1	15	-0.067	0.258	-0.573	-	0.439
CRC 2	15	-0.067	0.258	-0.573	-	0.439
CRC 3	15	-0.067	0.32	-0.694	-	0.56
CRC 4	15	0.033	0.129	-0.286	-	0.22

Source: Study results, 2014

Average^a: average of differences (golden standard – nurse); **SD:** standard deviation of the difference; **IL 95%:** Inferior Limit of 95% for the difference; **SL 95%:** Superior Limit of 95% for the difference

Table 3: Average, standard deviation and limits of concordance of Ontario Protocol Assessment Level score among nurses and the specialists' committee (the golden standard) during occasion M1. Rio de Janeiro, Brazil, 2012.

Table 4 showed that there is a significant association between the OPAL score and the classification (p=0.012). According to Dunn's multiple comparison test, at the 5% level, the level of high workload presented a score significantly higher than the low and medium load levels. There is no significant difference in the score between low and medium workload levels. Although there was a significant association, the score was not enough to discriminate the three classification levels (subjective evaluation). Probably, for a much larger sample this relationship can become more consistent.

Observer	Classification	n	median	min	max	Pvalue ^a	Significative differences ^b
Expert Committee	low	5	5	4	8	0,012	Low ≠ High
	medium	6	5,5	1,5	6,5		Medium ≠ High
	high	4	8,5	8,5	9		

Source: Study results, 2014

^a ANOVA de Kruskal-Wallis; ^b Dunn multiple comparison test, at level of 5%

Table 4: Analysis of the OPAL score according to the classification for the specialists' committee.

Table 5 showed that there is a significant association between the OPAL score and the classification for all four CRC, except for CRC 2 at time M2 (p=0.085). Similarly, although there was a significant association, the score (punctuation) was not enough to discriminate

the three classification levels (subjective assessment), except for CRC 4 at time M2. We have to consider the very small sample size at each grade level.

Observer	Classification	n	median	min	max	p value a	Significative differences b
CRC1 M1	low	5	4	4	5,5	0,049	
	medium	8	6	1,5	9		low ≠ high
	high	2	8,5	8,5	8,5		
CRC1 M2	low	5	4	3,5	6	0,022	
	medium	7	5,5	5,5	9		low ≠ high
	high	3	8,5	8,5	8,5		
CRC2 M1	low	5	4	4	5,5	0,049	
	medium	8	6	1,5	9		low ≠ high
	high	2	8,5	8,5	8,5		
CRC2 M2	low	5	5	4	5,5	0,085	
	medium	7	6,5	3	9		
	high	3	8,5	5,5	8,5		
CRC3 M1	low	4	4,3	4	5,5	0,009	low ≠ high
	medium	7	5,5	1,5	8		medium ≠ high
	high	4	8,5	7,5	9,5		
CRC3 M2	low	3	4	4	5	0,015	low ≠ high
	medium	7	5,5	1,5	8		medium ≠ high
	high	5	8,5	5,5	9		
CRC4 M1	low	5	4	1,5	5,5	0,008	
	medium	9	6,5	5,5	8,5		low ≠ medium
	high	1	9	9	9		
CRC4 M2	low	4	4	1,5	5	0,001	low ≠ high
	medium	5	5,5	5,5	5,5		medium ≠ high
	high	6	8,5	8	9		medium ≠ high

Source: Study results, 2014

^a ANOVA de Kruskal-Wallis; ^b Dunn multiple comparison test, at level of 5%

Table 5: Analysis of CRC scores' on different moments.

After the use of the graphic analysis developed by Bland Altman, during M1 (chosen due to the results of the analysis of ICC being similar to the ones found in M2), it was seen that the nurses presented straight limits of concordance (with a maximum of 1.25 points of amplitude), which demonstrates a high level of reliability if compared to the specialists' committee; no systematic behavior is seen in the differences in the measurements observed as described in average values or, in other words, a random distribution of the differences of the general average (lack of bias) and also, that the averages of the differences are near zero. This shows that the OPAL score on occasion M1 do not overestimate nor underestimate the evaluation of the specialists' committee.

It was also seen that CRC 4 presented a higher level of concordance, because his interval was the narrowest (-0.286 to 0.220) if compared to all the other observers.

Discussion

The importance of measuring the workload has emerged in clinical research centers. This load has increased both due to the need to achieve internationally recognized quality standards, and to the need to evaluate the costs generated in the carrying out of clinical studies and other spending on the part of the team involved in such studies [11-13].

The increasing complexity of treatments nowadays as found in clinical studies, their elevated cost, the emphasis on a more effective use of resources, the adherence to standards of good clinical practice, and the rise in demand for guaranteeing the quality of the data gathered, are all seen as the factors that increasingly generate workload in research centers [11,14-17].

When not focusing on these issues, the result is the use of methodologies that have not been approved, or even the use of simple estimates to calculate the workload, leading, in some cases, to non-realistic expectancies, excessive workload, and an inefficient use of available resources.

In this context, and due to the lack of any discussion in the scientific literature, and the absence of tools that could even consider the complexity related to the development of clinical studies, this paper proposes the transcultural adaptation and validation of OPAL, an instrument that is capable of determining the workload of clinical protocols, and which has already been validated in Canada.

After the development of all the stages of the transcultural adaptation of OPAL, and evaluating the methodological process used, it was seen that the discussions with the author of the original instrument and the specialists' committee were essential to achieve a coherent instrument for use in the Brazilian context that was suitable for the demands and specificities of the studied scenario.

With the transcultural adaptation, it was possible to identify points of divergence between Canadian and Brazilian realities. These were overcome with the insertion or removal of items that were seen as local characteristics in terms of the performance of clinical studies.

From the statistical tests performed, comparing the results given by different CRC, using the adapted OPAL instrument, and the specialists' committee, during both occasions, it was seen that there was a high level of concordance, thus leading us to declare confidence in the validation and the reliability of the adapted instrument.

Conclusion

In conclusion, the adapted OPAL does match the needs of users when it comes to calculating the workload generated by clinical protocols. However, it is important that the instrument is used consistently, and that the professionals responsible for distributing the protocols among the team, evaluate other factors that can influence the work of the clinical research coordinator.

There is an expectation to use the instrument at the HCIII in all studies happening in the research center, and periodic evaluations of these studies are already in place. For such an exercise, there will be created a commission to evaluate the studies, and a channel for open communication with the author of the original instrument will be developed. This will be done in order to assist professionals that use the instrument to resolve any doubts, and to permit them to be more coherent in terms of their assumptions regarding OPAL.

References

1. Ministério da Saúde (BR), Departamento de Ciência e Tecnologia (Decit) (2011) Fortalecendo a pesquisa clínica no Brasil: a importância de registrar os ensaios clínicos. *Rev Saúde Pública* 45: 436-439.
2. Schout D, Novaes HMD (2007) Do registro ao indicador: gestão da produção da informação assistencial nos hospitais. *Ciênc Saúde coletiva* 12.
3. Lousana G, Accetturi C, Castilho VC (2002) Guia prático para coordenadoras de estudos clínicos. *Interface* 1: 15-28.
4. Alves FVG, Dames KK, Lima R (2011) O Enfermeiro como Coordenador de Estudos Clínicos em Oncologia. *Rev bras Cancerol* 57: 75-84.
5. Smuck B, Bettello P, Berghout K, Hanna T, Kowaleski B, et al. (2011) Ontario Protocol Assessment Level: Clinical Trial Complexity Rating Tool for Workload Planning in Oncology Clinical Trials. *J Oncol Pract* 7: 80-84.
6. Guillemin F, Bombardier C, Beaton DE (1993) Cross-cultural adaptation of health related quality of life measures: literature review and proposed guidelines. *J clin epidemiol* 46: 1417-1432.
7. Beaton D, Bombardier C, Guillemin F (2002) Recommendations for the cross-cultural adaptation of health measures. American Academy of Orthopaedic Surgeons; Institute for Work & Health.
8. Schmidt DRC, Dantas RAS (2011) Análise da validade e confiabilidade da versão adaptada para o português do Questionário de Senso de Coerência de Antonovsky, entre profissionais de enfermagem. *Rev Latino Am Enfermagem* 19.
9. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods for clinical measurement. *Lancet* 1: 307-310.
10. Ministério da Saúde (BR) (2013) Conselho Nacional de Saúde. Resolução n. 466 de 12 dezembro de 2012. Aprova as diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos. *Diário Oficial da União, Brasília, DF* Seção 1: 59.
11. Roche K, Paul N, Smuck B, Whitehead M, Zee B, et al. (2002) Factors Affecting Workload of Cancer Clinical Trials: Results of a Multicenter Study of the National Cancer Institute of Canada Clinical Trials Group. *J Clin Oncol* 20: 545-556.
12. Raad J, Gool KV, Haas M, Haywood P, Faedo M, et al. (2010) Nursing takes time: workload associated with administering cancer protocols. *Clin J Oncol Nurs* 14: 735-741.
13. Swan BA, Griffin KF (2005) Measuring nurse workload in ambulatory care. *Nursing Economic* 23: 253-60.
14. Grazziano ES, Ferraz Bianchi ER (2010) Impacto del estrés ocupacional y burnout en enfermeros. *Enferm glob* 18.
15. Schmoeller R, Trindade LL, Neis MB (2011) Cargas de trabalho e condições de trabalho da enfermagem: revisão integrativa. *Rev Gaucha Enferm* 32.
16. Milani A, Mazzocco K, Stucchi S, Magon G, Pravettoni G, et al. (2017) How many research nurses for how many clinical trials in an oncology setting? Definition of the nursing time required by clinical trial – assessment tool (NTRCT – AT). *Int J Nurs Pract* 23.
17. Camargo TC (2002) A participação do enfermeiro em ensaios clínicos: uma revisão da literatura. *Revista Brasileira de Cancerologia* 48: 569-576.