



Field work I: selecting the instrument for data collection^{*}

Joao Luiz Bastos¹
Jeovany Martínez Mesa³

Rodrigo Pereira Duquia²
Renan Rangel Bonamigo²

David Alejandro González-Chica¹

DOI: <http://dx.doi.org/10.1590/abd1806-4841.20143884>

Abstract: The selection of instruments that will be used to collect data is a crucial step in the research process. Validity and reliability of the collected data and, above all, their potential comparability with data from previous investigations must be prioritized during this phase. We present a decision tree, which is intended to guide the selection of the instruments employed in research projects. Studies conducted along these lines have greater potential to broaden the knowledge on the studied subject and contribute to addressing truly socially relevant needs.

Keywords: Data collection; Health surveys; Methods; Questionnaires; Review

INTRODUCTION

This article discusses one of the most trivial aspects of a researcher's daily tasks, which is to select among various available options the instruments to perform data collection that meet the intended objectives and, at the same time, respect budgetary and temporal restrictions as well as other equally relevant issues when conducting a research. The instrument for data collection is a key element of the traditional questionnaires, which are used to investigate various topics of interest among participants of scientific studies. It is through questionnaires / instruments aimed to assess, for example, sun exposure, family history of skin diseases and mental disorders, that it is possible to measure these phenomena and analyze their associations in health surveys. In this paper, we discuss only questionnaires and their elementary components - the instruments; the reader should refer to specialized literature for knowledge and proper management of other resources available for data collection, including, for example, equipment to measure blood pressure, exams on cutaneous surface lesions and collection of biological material in studies focused on biochemical markers, such as blood parameters etc. Even so, it is argued that the guiding principles presented in this text widely apply, with minor adaptations, to all data collection processes.

As discussed earlier, all scientific investigations, including those in the field of Dermatology,

must start with a clear and predefined question.¹ Only after formulating a pertinent research question may the researcher and his/her team plan and implement a series of procedures, which will be able to answer such a question with acceptable levels of *validity* and *reliability*. This means that the scientific activity is organized by framing questions and executing a series of procedures to address them, including, for example, the use of questionnaires and their constituent instruments. Such procedures should be recognized as processes that respect ethical research guidelines and whose results are accepted by the scientific community, i.e. they are valid and reliable. However, it should be clarified before moving forward, albeit briefly and partially, what is commonly meant by validity and reliability in science.

In general, validity is considered to be present in an instrument, procedure or research as a whole, when they produce results that reflect what they initially aimed to evaluate or measure.² A research can be judged both in terms of internal validity when its conclusions are correct for that sample of studied individuals, as well as external validity, when its results can be generalized to other contexts and population domains.³ For example, in a survey that estimates the frequency of pediatric atopic dermatitis in Southeast Brazil, the closer the results are to the examined subjects' reality, the greater their internal validity. In other

Received on 28.07.2014.

Approved by the Advisory Board and accepted for publication on 29.07.2014.

^{*} Work performed at the Universidade Federal de Santa Catarina, Porto Alegre Health Sciences Federal University and Latin American Cooperative Oncology Group (LACOG) - Porto Alegre (RS), Brazil.

Financial Support: None

Conflict of Interest: None

¹ Universidade Federal de Santa Catarina (UFSC) - Florianópolis (SC), Brazil.

² Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSIPA) - Porto Alegre (RS), Brazil.

³ Latin American Cooperative Oncology Group (LACOG) - Porto Alegre (RS), Brazil.

words, if the actual frequency of atopic dermatitis were 12.5% for this region and population, a research that achieved a similar result would be considered internally valid.⁴ The ability to generalize or extrapolate those results to other regions in the country would be reflected in the study's external validity. Furthermore, to be valid in any dimension this research should have used an established instrument, able to distinguish individuals who actually have this dermatological condition from those who do not have it. So, the study's validity research depends on the validity of the very instruments that are used.

A research instrument is deemed reliable when it is able to consistently generate the same results after being applied repeatedly to the same group of subjects. This concept is often used in multiple stages of the research process including, for example, when a data collection supervisor performs a quality control check, reapplying some questions to the same subjects already interviewed or even during the construction of a new instrument in the test-retest phase in which the reliability and consistency of the given answers are examined. The *Acne-Specific Quality of Life Questionnaire* was considered reliable after recording consistent data on the same individuals in an interval of seven days between the first and second administrations.⁵ Moreover, a study will be more reliable as more precise instruments are used in data collection and as more subjects are recruited - studies with a significant number of participants present results with a smaller margin of error. It is noteworthy that, although the concept of reliability extrapolates the question of temporal consistency (test-retest), we will address this aspect in a more limited fashion in this article. The interested reader should consult specific publications for further discussion of this topic.⁶

Resuming our original question, it must be noted that the need for careful selection of instruments to be used in scientific investigations must have a solid theoretical basis and should not be considered as a mere fad. Ultimately, the wrong choice of an instrument can compromise the internal validity of the study, producing misleading results, which are therefore unable to answer the research question originally formulated. Besides, the choice of an instrument also has implications in the ability to generalize the research results (external validity), and to compare them with those of other studies conducted nationally or internationally on the same subject - researchers using *equivalent* instruments can establish an effective dialogue, which enables a more comprehensive analysis of the phenomenon in question, including its antecedents and consequences.⁷

In order to justify the need to carefully select the instruments to be used in scientific research and

also provide basic guidelines so that these decisions are based on solid grounds, we will divide this article into the following sections: (1) On the comparative nature of scientific research; (2) How to select the most appropriate instrument for my research when there are prototypes available in the scientific literature; and (3) What to do when there are no available instruments to assess the phenomenon of interest to the researcher.

ON THE COMPARATIVE NATURE OF SCIENTIFIC RESEARCH

The inherently comparative nature of scientific research represents an aspect that may sometimes pass unnoticed even to the more experienced researcher. However, the careful examination of a project's theoretical framework, the discussion of a scientific article and also the study results are sufficient to easily demonstrate this comparative nature.

Investigations in the field of Social Anthropology, for example, are based on comparisons of complex cultural systems; the identification of idiosyncrasies in a particular cultural system is only possible after its confrontation with the characteristics of another system.⁸ So, the conclusion that a specific South American indigenous population exhibits distinct kinship relations from those observed in the hegemonic Western family composition only occurs when these two forms of cultural systems are compared.

The same occurs in the healthcare field - comparisons are crucial to arrive at conclusions, including the evaluation of consistency of certain scientific findings among a set of previously conducted studies. Likewise, if a researcher is interested in examining the quality of life of patients affected by the pain caused by lower-limb ulcers, he and his team should necessarily make comparisons.

In this case, the comparison is between two distinct groups of subjects with lower-limb ulcers, one of them with pain and the other without it, to ascertain whether the levels of quality of life found in both groups are similar or not. If the researcher observes, by comparison, that the group with pain has a diminished quality of life compared to the group without pain, he may conclude that there is a negative correlation between quality of life and pain related to lower-limb ulcers.

However, the comparative principle goes beyond contrasting internal groups in a study, as illustrated above. Researchers of a particular subject, for example, the development of melanocytic lesions, can only confirm that the use of sunscreen prevents their occurrence, when multiple scientific studies evaluating this question have previously shown it. In other words, by comparing the results generated by several

investigations on the same topic, the scientific community can judge the consistency of the findings and thus make a solid conclusion about the subject matter.

Considering that the comparison of results from different studies is a key aspect of the production and consolidation of scientific knowledge, the following question arises: How should one conduct scientific studies so that their results are comparable to each other? Invariably, the answer to this question includes the use of scientific research instruments that are valid, reliable, and equivalent in different studies. So, what are the basic elements of the selection and use of these tools that enables this scientific dialogue? This is exactly what the subsequent section aims to answer.

HOW TO SELECT THE MOST APPROPRIATE INSTRUMENTS FOR MY RESEARCH, WHEN THERE ARE PROTOTYPES AVAILABLE IN THE SCIENTIFIC LITERATURE

We assume that the researcher has already formulated a clear and pertinent research question, which he or she wants to answer by conducting a scientific research. To illustrate the situation, imagine that a researcher is interested in estimating the frequency of depression and anxiety in a population of caregivers of pediatric patients with chronic dermatoses. The research question could be worded specifically in this way: *What is the frequency of anxiety and depression in caregivers of children under five years of age, with chronic dermatoses (atopic dermatitis, vitiligo and psoriasis) residing in the city of Porto Alegre in 2014?*

Considering that the phenomenon to be evaluated is restricted to anxiety and depression, how should the investigator proceed in this regard? There are at least two possible alternatives: the researcher can develop a set of entirely original items (instrument) to measure both mental disorders cited or select valid and reliable instruments already available in the scientific literature to assess such disorders.

Both alternatives have their own implications. Developing a new instrument means conducting an additional research project that will require considerable effort and time to be carried out. The scientific literature on the development and adaptation of instruments emphatically condemns this decision.⁹ Often, researchers who choose to develop new instruments overestimate the deficiencies of the existing ones and disregard the time and effort needed to construct a new and appropriate prototype. In most cases, the optimistic and to some extent naive expectations of these researchers are frustrated by the development of a new instrument whose flaws are potentially similar to or even greater than the ones found in existing instruments, but with an additional aggravating factor: the possibility of comparing the results of a study

performed with the newly developed instrument to those of previous studies employing other measuring tools is, at least initially, nonexistent. In general, we recommend developing new instruments only when there are no other options for measuring the phenomenon in question or when the existing ones have huge and confirmed limitations.

If the researcher has taken the (right) decision to use an existing instrument to assess anxiety and depression, we suggest that he or she should cover the following steps:⁹

- 1 - Conduct a very broad and thorough literature search to retrieve the instruments that assess the phenomenon in question. The bibliographic search can start in the traditional bibliographical resources in healthcare, such as PubMed, but it must also take in consideration those available in other scientific fields, such as psychology and education whenever necessary;
- 2 - Identify all the available instruments to measure the phenomenon of interest. Eventually, some may not have been published in books, book chapters or as scientific articles. In these cases, it is essential to make contact with the researchers working in the area to ask them about the existence of unpublished measuring instruments (gray literature);
- 3 - Based on the elements presented in Table 1, reassess the course of development of each identified instrument, seeking to distinguish those with established results, good indicators of validity and reliability and in particular, those extensively used by the scientific community.¹⁰ Ideally, the instruments of choice are those that were also evaluated by independent research groups, i.e., groups which were not involved with their initial design; and
- 4 - Select an instrument that meets the goals of your study, considering ethical, budgetary and time constraints, among others. Whenever the chosen instrument has been created in a research context significantly distinct from that of your investigation, search the literature for studies of cross-cultural adaptation that aimed to produce an equivalent version of the instrument for the language and cultural specificities of your research context.¹¹ Thus, as argued by Reichenheim & Moares, "the process of cross-cultural adaptation should be a combination between a component of literal translation of words and phrases from one language to another and a meticulous tuning process, that addresses the cultural context and lifestyle of the target-population to which the version will be applied."¹⁰

Proceeding as described above, the privileged scenario will be the one in which studies addressing

TABLE 1: Aspects regarding validity and reliability (quality) of measurement instruments*

Aspect	Characteristic	Conceptual definition and strategies	What to observe
Validity	Dimensional validity	This refers to the correspondence that should exist between the instrument's internal structure and the one that was theorized regarding the phenomenon to be evaluated. For example, if the instrument aims to measure mental disorders and includes depression and anxiety as its two dimensions of interest, a statistical analysis of it should reveal such dimensions.	Results of exploratory and confirmatory factor analyzes, demonstrating the correspondence between the postulated structure for the phenomenon and the loading of the instrument items on their respective dimensions. Returning to the example, a factor analysis of the instrument for common mental disorders should demonstrate that the questions regarding anxiety are grouped in the dimension that concerns them (anxiety) and the questions about depression are associated with their underlying factor (depression).
	Construct validity	The instrument's ability to measure what it intends to assess when there is not another tool considered the "gold standard" for measuring the phenomenon of interest. Construct validity can be determined by several methods, including: <ul style="list-style-type: none"> • Extreme groups: the instrument is applied to two groups, one supposedly with the presence of the characteristic of interest and the other without it. • Convergent validity: comparison between the assessments obtained with the instrument of interest versus those resulting from another scale used for measuring the same phenomenon. • Discriminant or divergent validity: it can be obtained by testing the correlation between the results of an instrument and those of another one used for measuring a different construct. 	Finding that the instrument confirms the hypothesis that one group has the feature of interest and the other does not, is an indication of the instrument's validity through the comparison of extreme groups. In the convergent validity example, it is expected that the results from both instruments point in the same direction (that they are positively correlated with each other). The correlation between the results of different instruments must be zero when evaluating the discriminant validity.
	Criterion-related validity	Ability of the instrument to measure what it proposes, whenever there are instruments considered as the "gold standard". The verification of this validity involves the application of two instruments, the one intended to be used and another considered as reference, and also by the observation of their correlation. Criterion validity is typically divided into two subtypes: <ul style="list-style-type: none"> • Concurrent or simultaneous validity: tests the correlation of the instrument of interest with a "gold standard" after applying both simultaneously. • Predictive validity: determined by the ability of the instrument to predict a future event, which will be based on the subsequent application of the reference instrument. 	In both cases the correlation between the instrument of interest and the "gold standard" one support the validity argument for the former.
Reliability	Internal consistency	As an illustration, if we wish to measure the functional capacity of individuals and we have several items (questions) to measure it, they should have a high correlation among themselves. The measures used to assess internal consistency are the Cronbach's alpha coefficient and the Kuder-Richardson coefficient, among others. In all cases, it is possible to estimate the internal consistency with a single application of the instrument to the sample under evaluation.	The minimum acceptable value for these coefficients is 0.8.
	Temporal stability	Stability may be assessed in different ways, including: <ul style="list-style-type: none"> • The degree of agreement between different observers, using the same instrument (inter-observer reliability). • The consistency of the observations made by the same examiner at different moments in time (intra-observer reliability or test-retest). 	The minimum acceptable value for these coefficients is 0.5.

*This Table was designed based on data published by Reichenheim & Moraes¹⁰ and Streiner & Norman⁹ that must be consulted if the reader wishes to advance further in these topics.

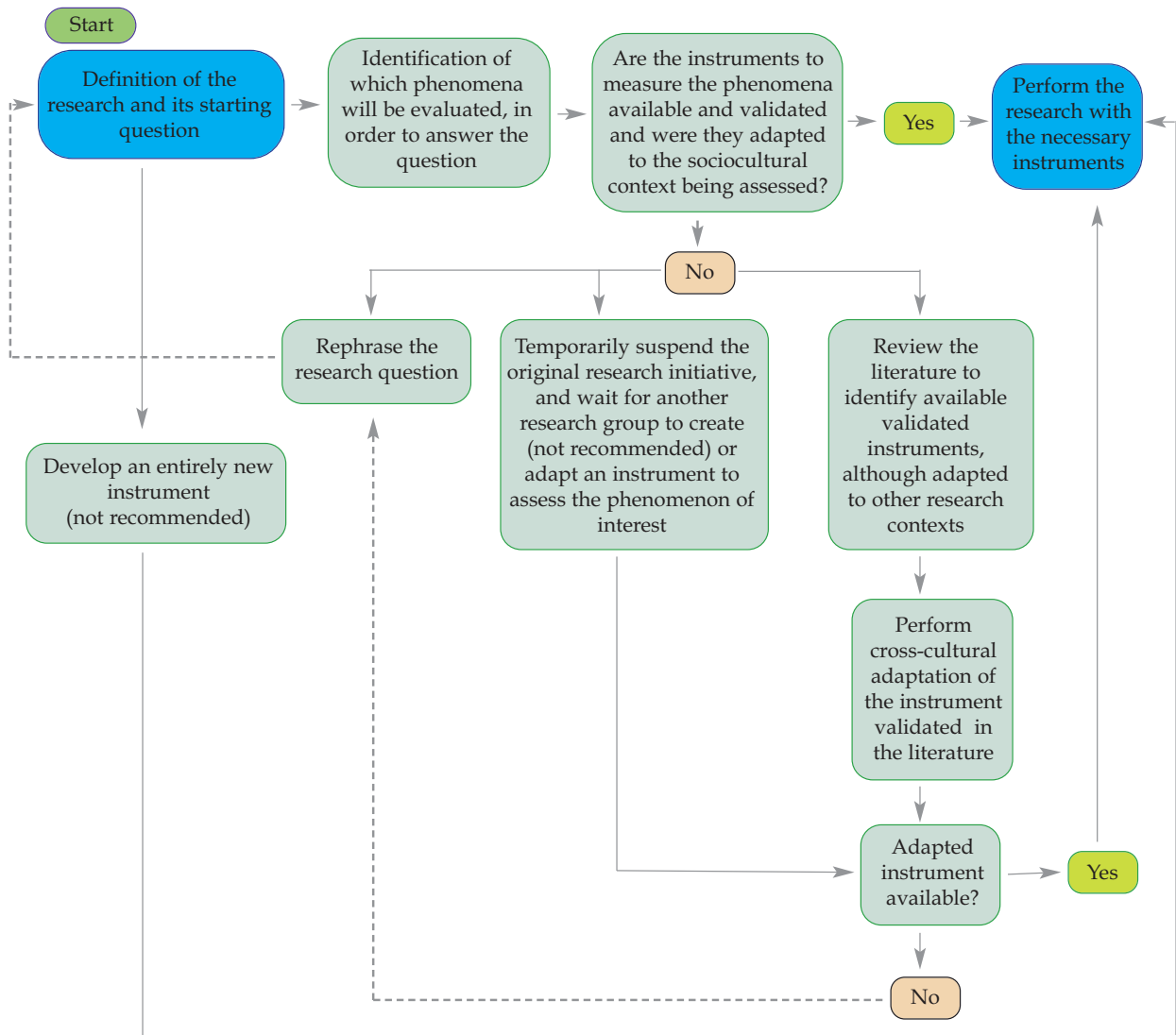


FIGURE 1: Decision tree to guide the process of choosing an instrument to collect scientific research data

the same phenomena shall be conducted with equivalent instruments to assess them and therefore their results will be readily comparable. This would be the same as having a study conducted in different countries on the topics of depression and anxiety in caregivers of pediatric patients with chronic skin diseases and each one would use a version of the instrument adapted for the respective research contexts. So, while in Brazil the *Hospital Anxiety and Depression Scale* would be used in a version adapted to Brazilian Portuguese, the equivalent version of this same instrument in Japanese would be used in Japan. Therefore, the rates of these common mental disorders, estimated by both studies would be directly comparable at the end of each survey.

WHAT TO DO WHEN THERE ARE NO AVAILABLE INSTRUMENTS FOR MEASURING THE PHENOMENON OF INTEREST TO THE RESEARCHER

Whenever the researcher is confronted with the lack of instruments for measuring the phenomenon of interest, it is possible to follow at least one of these leads:

- 1 - Ultimately, review the research question and replace it with one that does not involve the assessment of the phenomenon for which there are no measurement tools available;
- 2 - Develop an ancillary research program, whose main objective is to perform a cross-cultural adaptation of a measurement instrument to the context in which the investigation will be conducted. In

this case, one must consider the need to postpone the original study until the adapted version of the instrument is available - something that takes in the most optimistic prediction, two to three years; or

3 - Temporarily suspend the research initiative, waiting until other researchers have provided an adapted version of the selected instrument, making it possible to execute the study in a similar socio-cultural context.

The synthesis of the entire process suggested in this article is illustrated in the decision tree, depicted in Figure 1. We believe that the conduct of studies along these lines has an even greater potential to increase the knowledge on the particularities of any topic of interest and, ultimately, contribute to approaching socially relevant demands. □

REFERENCES

1. Martínez-Mesa J, González-Chica DA, Bastos JL, Bonamigo RR, Duquia RP. Sample size: how many participants do I need in my research? *An Bras Dermatol.* 2014;89:609-15.
2. Gil AC. *Métodos e técnicas de pesquisa social.* São Paulo: Atlas; 1999.
3. Pereira MG. *Epidemiologia: teoria e prática.* Rio de Janeiro: Guanabara Koogan; 2002.
4. Solé D, Camelo-Nunes IC, Wandalsen GF, Mallozi MC, Naspitz CK; Brazilian ISAAC Group. Prevalence of atopic eczema and related symptoms in Brazilian schoolchildren: results from the International Study of Asthma and Allergies in Childhood (ISAAC) phase 3. *J Investig Allergol Clin Immunol.* 2006;16:367-76.
5. Kamamoto Cde S, Hassun KM, Bagatin E, Tomimori J. Acne-specific quality of life questionnaire (Acne-QoL): translation, cultural adaptation and validation into Brazilian-Portuguese language. *An Bras Dermatol.* 2014;89:83-90.
6. Carmines EG, Zeller RA. Reliability and validity assessment. Lewis-Beck MS, editor. Thousand Oaks: SAGE; 1979.
7. Berry JW, Poortinga YH, Segall MH, Dasen PR. *Cross-cultural psychology: research and applications.* Cambridge: Cambridge University Press; 2007.
8. DaMatta R. *Relativizando: uma introdução à antropologia social.* Rio de Janeiro: Rocco; 1987.
9. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use.* 2nd ed. Oxford: Oxford University Press; 1998.
10. Reichenheim ME, Moraes CL. Qualidade dos instrumentos epidemiológicos. In: Almeida-Filho N, Barreto ML, editores. *Epidemiologia & saúde: fundamentos, métodos e aplicações.* Rio de Janeiro: Guanabara Koogan; 2012. p. 150-64.
11. Reichenheim ME, Moraes CL. Operationalizing the cross-cultural adaptation of epidemiological measurement instruments. *Rev Saúde Pública.* 2007;41:665-73.

MAILING ADDRESS:

João Luiz Bastos
 Universidade Federal de Santa Catarina
 Trindade
 88040-970 - Florianópolis - SC
 Brazil
 E-mail: joao.luiz.epi@gmail.com

How to cite this article: Bastos JL, Duquia RP, González-Chica DA, Mesa JM, Bonamigo RR. Field work I: selecting the instrument for data collection. *An Bras Dermatol.* 2014;89(6):918-23.