

# Review of measurement instruments in research ethics in the biomedical sciences, 2008–2012

Research Ethics

2014, Vol. 10(3) 141–150

© The Author(s) 2014

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1747016114538963

rea.sagepub.com

**Barbara K Redman**

NYU School of Medicine, USA

## Abstract

There is an urgent need in biomedical science to understand whether regulations are being met, prerequisite to goals of subject protection and integrity in research practice. This article presents an update of a 2006 summary of measurement instruments in research ethics with psychometric information in the years 2008–2012. A review of 25 instruments identified seven used in the time period 2008–2012 and which had accumulated at least one study of its psychometric qualities beyond its developmental phase. Many of these instruments had been accumulating psychometric information over more than a decade. Two additional but still underdeveloped instruments addressing important bioethical issues – coercion and therapeutic misconception – are included because they address important issues in research ethics. Bioethicists use a wide range of methods for knowledge development and verification; each method should meet stringent standards of quality. Measurement instruments that meet these standards have the potential to greatly ease the work of institutional review boards and other regulatory bodies as well as to enhance empirical work on human research ethics.

## Keywords

institutional review board, measurement, research ethics

## Introduction

In order to understand treatments and outcomes and to accumulate knowledge over multiple studies, valid and reliable measurement instruments are important

---

### Corresponding author:

Barbara K Redman, NYU School of Medicine, 227 East 30th Street, 7th floor, Room 722, New York, NY 10016, USA.

Email: [jessica.wico@nyumc.org](mailto:jessica.wico@nyumc.org)

for the biomedical sciences. But they are also essential for the same reasons, to support ethical and regulatory interventions in the conduct of research, including assuring informed consent, detecting harms, especially among the vulnerable, recruitment of subjects and trust in researchers. Much empirical research on these matters has used surveys and questionnaires developed for a particular study with little understanding of the quality of the measurement and that are rarely replicated in future studies. This approach makes it difficult to interpret the accumulated body of knowledge and to constantly improve the practice of research.

Societal expectations for protection of human subjects and of research integrity require a stock of well-validated measures in all areas of the research process. They are to be used as evidence in judgments about whether ethical and regulatory standards have been met and to stimulate discussion about whether standards should be changed. The purposes of this article are to: (i) report a systematic review of recent instruments for measuring constructs in research ethics; (ii) evaluate them according to psychometric standards; and (iii) comment on their adequacy as a group to support the practice of research.

## **Methods**

In addition to a broad reading of the research ethics and regulatory literature, instruments were retrieved from PubMed, by a search under the term ‘ethics empirical research’, yielding 1181 entries for the years 2008–2012 and sorted by those whose titles indicated they were relevant to the research process (accessed 15 September 2013). Instruments need not have been originated during this 5 year period but must have been used during it and must have accumulated at least one study of its psychometric qualities beyond its developmental phase to begin to get a sense of its validity and reliability.

Citations were screened for measurement instruments for which psychometric information was available, and through citation indexes (Google Scholar and Scopus) for other studies that had used the instrument and added psychometric information about measurement characteristics and usefulness in addressing questions of research practice. In some instances, such as informed consent for research, a published review of instruments was available and consulted.

Well-established sources suggest the kinds of psychometric data that should be available to evaluate a measurement instrument and the standards they should meet. Validity has to do with meaning or interpretation of scores on a measure. Content validity is evaluated logically by consensus of opinion that appropriate content is covered. Criterion validity is the correlation between a measure and a gold standard of the same attribute, predictive if one is interested in predicting future states. Construct validity assesses the extent to which the instrument measures the attribute it purports to measure. Factors are statistically identified clusters

of items that measure one or more constructs. Convergent and discriminant validity describe the ability to detect differences in groups known to be similar or different in the attribute being measured. Many consider criterion, convergent and discriminant validity as forms of evidence that fall under the umbrella of construct validity, and that is how they are reported in this study. Internal consistency reliability is a measure of the homogeneity of items in a scale, usually measured by Cronbach's alpha, which for research purposes must show a value between .70 and .90. Test-retest reliability measures the stability of scores over time. Responsiveness (also called sensitivity) is the ability to detect subtle but significant change, often as an outcome of an intervention (McDowell, 2006).

## Results

Twenty-five instruments were retrieved: one very well summarized elsewhere and not considered in detail here, seven included in this review plus two developed in the years of the review with a single rigorous test of their psychometric properties but not yet tested a second time (please see Table 1); fifteen instruments were identified as missing sufficient psychometric information in the time frame designated above.

Clusters of instruments measuring important areas of ethical research practice follow. Elements of informed consent or capacity to give it has received most development, perhaps reflecting a requirement by some institutional review boards that adequacy of consent must be documented by some 'objective' means, particularly in subjects with diminished capacity or lack of formal education.

## Informed consent

Assessing understanding, appreciation, reasoning and choice related to capacity, the MacArthur competence assessment tool for clinical research (MacCAT-CR) has been widely used and its psychometric characteristics well studied. It uses a semi-structured interview with standard hypothetical protocols, easing comparison across studies (Appelbaum and Grisso, 2001). Because it is so well-documented and requires specific training, it will not be further considered here.

## *Quality of informed consent measure*

The quality of informed consent measure (QuiC) was developed to measure subjects' actual (part A) and perceived (part B) understanding of cancer clinical trials (see Appendix in Joffe et al., 2001, for instrument and scoring details). Some items are specific to trial phase I (toxicity and dose finding), II (preliminary efficacy) or III (randomized controlled) trials. QuiC is based on elements outlined in federal

**Table 1.** Measurement characteristics of reviewed instruments.

Instrument	Content validity	Construct validity	Internal consistency <sup>a</sup>	Test–retest reliability <sup>b</sup>
Quality of informed consent	Yes	Some evidence	–	.66–.77
Decision making control instrument	Yes	Other instrument	.83	–
Measuring trust on medical researchers	Yes	FA	.87, .72	–
Trust in medical researchers scale	Yes	Known groups Other instruments & known groups	.84	–
Reactions to research participation questionnaire	Yes	Adequate	Adequate	–
Attitudes to randomized trials questionnaire	Yes	Predictive of trial enrollment	–	–
Survey of organizational research climate	Yes	Discriminant validity	.8–.87	.72–.83
Promising but tested only once				
Coercion assessment scale	Yes	FA	Adequate	–
Therapeutic misconception scale	Yes	Predictive	–	–

<sup>a</sup>Standard is 0.7–0.9 for judgments about individuals.

<sup>b</sup>Standard is 0.7–0.8.

FA = factor analysis.

regulations and by expert panels, supporting content validity. Test–retest reliability was .66 for the objective scale and .77 for subjective understanding. This instrument should be useful to monitor the informed consent process as a screen for both disclosure and capacity and to study interventions intended to improve its quality (Joffe et al., 2001). Participants with poorer objective understanding tended to report higher expectations of benefit from the research (Weinfurt et al., 2012), supportive of validity. Additionally, an adaptation of QuiC has been developed for storage and future use of biospecimens for research studies, showing significant discrepancy between what subjects thought they understood as measured by the subjective scale and the objective scale measuring what they really understood after they had given consent (Klima et al., 2014), signaling a need for further work on informed consent in this area of research.

### *Decision making control instrument*

Some work has focused on elements of informed consent believed to be under-developed both conceptually and in measurement. The Decision Making Control

Instrument (DMCI) centers on the construct of voluntariness, defined as perception of control over a specific decision, developed in the context of parents making research and treatment decisions for their seriously ill children. To parents, ‘a sense of control meant they perceived the decision was up to them, were aware of the option to reject or withdraw from the protocol, and perceived that the decision was made without feeling pressured, intimidated or coerced’; this kind of careful definition of the construct is essential to assuring that the item pool addresses all of its relevant domains (Miller et al., 2009: 24) (content validity). The nine-item DCMI (see Miller et al., 2009: Appendix) measures self-control, absence of control, and others’ control with internal consistency reliabilities from .68 to .87, overall scale .83. Scores were associated with measures of trust and decision self-efficacy, supporting construct validity. DCMI must be much more widely tested but could be used to explore potential causes of decreased perceived voluntariness including their physicians’ request to participate in research (Miller et al., 2011).

## **Trust in medical researchers**

Trust in its medical researchers is thought to affect willingness to enroll in trials and subsequent compliance with protocols.

### *Measuring trust in medical researchers*

Two scales with beginning development are available. Measuring trust in medical researchers is focused on physicians who do medical research and was developed based on a conceptual model of researcher trust to include safety, fidelity, honesty, and global trust. A 12-item and a 4-item scale were developed from previous survey items and focus groups (see Hall et al., 2006: Table 1) and tested with diabetes and asthma patients. Both scales measure a single factor and have acceptable internal reliability (.87 and .72). Scoring involves summing ‘strongly disagree’ (1) to ‘strongly agree’ (5), reverse scoring when necessary. Being black and levels of education were negatively related to trust (Hall et al., 2006). Two studies in persons with breast cancer found that regret about trial participation was negatively correlated with trust in medical researchers (Mancini et al., 2012a) and that dissemination of results to trial participants via the internet did not increase trust in medical researchers (Mancini et al., 2012b).

### *Trust in medical researchers scale*

The trust in medical researchers scale is built on a conceptual model of general fear of participation in medical research, mistrust of research personnel, and feelings that researchers act differently toward disadvantaged groups. To incorporate

content validity, items underwent a cognitive pretest of think-aloud interviews with African American volunteers. The 12-item questionnaire has an internal consistency of .84, a subscale on participation deception of .78, and a subscale on researcher honesty of .75 (for items, see Mainous et al., 2006: Table 1). Some evidence toward construct validity was obtained with other trust scales. African American respondents showed significantly less trust than did white counterparts. It is thought that investigators could tailor their recruitment efforts based on information from the scale (Mainous et al., 2006). Jefferson et al. tested an intervention to increase clinical research participation on Alzheimer's disease. African American groups have higher rates of Alzheimer's disease and participate in research at lower rates than do whites, but group discussion to increase participation did not alter trust scores (Jefferson et al., 2013).

## **Vulnerable populations**

### *Reactions to research participation questionnaire*

There has long been concern among institutional review boards (IRBs) that the process of research participation for those who had experienced trauma would itself create distress through the process of reporting on the trauma and its effects. A cluster of related instruments has been developed to answer the general question and to monitor negative reactions among those research participants. The reactions to research participation questionnaire (RRPQ) was first documented in 2000 with a 10-item scale asking degree of agreement, reflecting concerns from accident and assault survivors (content validity) (Ruzek and Zatzick, 2000).

A 24-item revised RRPQR consists of five subscales: participation, personal benefits, emotional reactions, perceived drawbacks, and global evaluation. RRPQ-C and RRPQ-P are adaptations to research participation among traumatically injured children and their parents (Kassam-Adams and Newman, 2002).

Although most participants report a positive benefit–risk ratio from study participation, a sub-set express distress and regret about study participation. Because it is not clear how to identify this group a priori, administering these instruments by interview and using them to monitor all trauma studies is suggested. Studies show adequate internal consistency, construct and concurrent validity (Newman et al., 2001).

## **Attitudes about trials**

### *Attitudes to randomized trials questionnaire*

Recruitment requires that potential subjects understand the nature of trials and randomization. The attitudes to randomized trials questionnaire (ARTQ) was

developed with face validity to understand why accrual to clinical trials of cancer therapy is very low and differentiated among those who seemed comfortable with randomization, those with concerns who with fuller explanation would consider randomization, and those firmly against randomization and participation in trials. ARTQ (which may be found in Fallowfield et al., 1998, table 5) is used to identify a potential subject's set of attitudes and has been found to be predictive (with 80% accuracy) of trial enrollment (Fleissig et al., 2001). Among a largely African American population, which is under-represented in cancer trials, ARTQ items changed after an educational intervention, although there was no control group in this study (Ford et al., 2012).

## **Organizational research climate**

### *Survey of organizational research climate*

The survey of organizational research climate (SORC) appears to be the first standardized tool available to measure organizational research environments that foster or undermine research integrity. Content validity was supported by research integrity experts, focusing on regulatory quality, responsible conduct of research resources, integrity norms, and inhibitors. SORC was tested with faculty and post-doctoral fellows at 40 academic health centers in top-tier research universities in the US, asking their perceptions about the climate in their university and department (the instrument and its scoring may be found in Martinson et al., 2013). Its seven subscales showed internal consistency reliability of .8–.87 and a test–retest reliability of .72–.83. The patterns of relationships between SORC subscales and measures of organizational justice discriminated among perceptions of research environment. Results can identify underperforming organizational units (Martinson et al., 2013). Further supporting validity, more positive perceptions of the research climate were associated with higher likelihood of desirable and lower likelihood of undesirable research practices (Crain et al., 2013). SORC has not yet been tested in other kinds of academic settings.

## **Promising but still developing**

### *Coercion assessment scale*

Two instruments representing ethically significant concerns are still being developed. The coercion assessment scale (CAS) measures perceived coercion for study participation among substance-abusing offenders, a group clearly vulnerable to economic hardship, comorbid psychiatric disorders, and social stigmatization. Fifteen percent of the study group felt they could not say no to study participation. Although the seven-item instrument has adequate internal consistency and a

single-factor structure, further validation of the scale and its interpretation is required (Dugosh et al., 2010).

### *Therapeutic misconception scale*

Therapeutic misconception (when research subjects do not distinguish between research and treatment) was identified in the 1980s and found to be widespread. It is worrisome because it may undercut informed consent. Initial development of the 10-item therapeutic misconception scale (TMS) reflects beliefs about degree of individualization, likeliness of benefit, and misunderstanding of study purpose to benefit future patients. TMS showed modest predictive value when validated against the current gold standard of a therapeutic misconception interview, but may be useful for identification of number of mistaken beliefs as a target for additional subject education (Appelbaum et al., 2012).

## **Stock of measurement instruments in research**

Psychometrically sound measurement instruments are an important tool for research ethics, and interest in their development is reflected in the rise of several journals focused on empirical research in bioethics. Clear operationalization of constructs and their measurement offer opportunities for learning, yet many areas of research ethics still lack such instruments, including privacy, confidentiality, and researcher professionalism. Also, for very few instruments has there been investment in repeated studies of validity and reliability that allow interpretation with some generalizability to populations of interest. Table 1 documents the limited psychometric data available for included instruments; they should undergo much more testing with the varied populations with which they will be used.

A recent summary of measurement in informed consent found most instruments developed specifically for each study, with lack of common definition and measurements, with most not providing any description concerning how the tool was developed, hindering comparison of findings and improvement of informed consent (Sand et al., 2010). This situation appears to be common in research ethics. Nevertheless, some of the instruments reviewed above show excellent attention to definition, formal tests of validity and reliability, and are poised to invite continued development in important areas of assessing research ethics.

Evidence of meeting psychometric standards is just as essential to interpreting the results of instruments as is meeting statistical standards for analysis of data. Just as all studies use the skills of statisticians, so scholars developing instruments should engage the skills of psychometricians.



## Limitations

Because classification of the medical literature does not allow thorough identification of measurement instruments, some may have been missed.

## Discussion

Significant progress in developing and testing measurement instruments relevant to research ethics has occurred since they were last reviewed as a group in 2006 (Redman, 2006). However, sustained effort will be needed to further define and measure constructs, to test instruments to understand their generalizability to relevant populations, and to accumulate bodies of research important to discussions in research ethics. When this work is further advanced, investigators could submit results from standardized measures as evidence that, for example, consent was informed, that concerns of vulnerable populations had been addressed, that subject recruitment of underrepresented groups had been tailored to increase their participation, or that therapeutic misconception had been minimized.

## Declaration of conflicting interests

The author declares that there are no conflicts of interest.

## Funding

No funding was involved.

## References

- Appelbaum PS and Grisso T (2001) *MacCAT-CR: MacArthur Competence Assessment Tool for Clinical Research*. Sarasota, FL: Professional Resource Press.
- Appelbaum PS, Anatchkova M, Albert K, Dunn LB and Lidz CW (2012) Therapeutic misconception in research subjects: Development and validation of a measure. *Clinical Trials* 9: 748–761.
- Crain AL, Martinson BC and Thrush CR (2013) Relationships between the Survey of Organizational Research Climate (SORC) and self-reported research practices. *Science and Engineering Ethics* 19: 835–850.
- Dugosh KL, Festinger DS, Croft JR and Marlowe DB (2010) Measuring coercion to participate in research within a doubly vulnerable population: Initial development of the Coercion Assessment Scale. *Journal of Empirical Research on Human Research Ethics* 5: 93–102.
- Fallowfield LJ, Jenkins V, Brennan C, Sawtell M, Moynihan C and Souhami RL (1998) Attitudes of patients to randomised clinical trials of cancer therapy. *European Journal of Cancer* 34: 1554–1559.
- Fleissig A, Jenkins V and Fallowfield L (2001) Results of an intervention study to improve communication about randomised clinical trials of cancer therapy. *European Journal of Cancer* 37: 322–331.

- Ford M, Wahlquist A, Blake R, et al. (2012) Assessing and intervention to improve clinical trial perceptions among predominately African-American communities in South Carolina. *Progress in Community Health Partnerships: Research, Education and Action* 6: 249–263.
- Hall MA, Camacho F, Lawlor JS, Depuy V, Sugarman J and Weinfurt K (2006) Measuring trust in medical researchers. *Medical Care* 44: 1048–1053.
- Jefferson AL, Lambe S, Romano RR, Liu D, Islam F and Kowall N (2013) An intervention to enhance Alzheimer's disease clinical research participation among older African-Americans. *Journal Alzheimer's Disease* 36: 597–606.
- Joffe S, Cook EF, Cleary PD, Clark JW and Weeks JC (2001) Quality of Informed Consent: A new measure of understanding among research subjects. *Journal of the National Cancer Institute* 93: 139–147.
- Kassam-Adams N and Newman E (2002) The reactions to research participation questionnaires for children and for parents (RRPQ-C and RRPQ-P). *General Hospital Psychiatry* 24: 336–342.
- Klima J, Fitzgerald-Butt SM, Kelleher KJ, et al. (2014) Understanding of informed consent by parents of children enrolled in a genetic biobank. *Genetics in Medicine* 16: 141–148.
- McDowell I (2006) *Measuring Health* (3rd edn). New York: Oxford University Press, 30–45.
- Mainous AG, Smith DW, Geesey ME and Tilley BC (2006) Development of a measure to assess patient trust in medical researchers. *Annals of Family Medicine* 4: 247–252.
- Mancini J, Genre D, Dalenc F, et al. (2012a) Patients' regrets after participating in a randomized controlled trial depended on their involvement in the decision making. *Journal of Clinical Epidemiology* 65: 635–642.
- Mancini J, Genre D, Dalenc F, et al. (2012b) Transparency in the presentation of trial results may not increase patients' trust in medical researchers. *Clinical Trials* 9: 90–93.
- Martinson BC, Thrush CR and Crain AL (2013) Development and validation of the Survey of Organizational Research Climate (SORC). *Science and Engineering Ethics* 19: 813–834.
- Miller VA, Reynolds WW, Ittenbach RF, Luce MF, Beauchamp TL and Nelson RM (2009) Challenges in measuring a new construct: Perception of voluntariness for research and treatment decision making. *Journal of Empirical Research on Human Research Ethics* 4: 21–31.
- Miller VA, Ittenbach RF, Harris D, et al. (2011) The decision making control instrument to assess voluntary consent. *Medical Decision Making* 31: 730–741.
- Newman E, Williams T, Sinclair R and Kaloupek D (2001) Empirically supported ethical research practice: The costs and benefits of research from the participants' view. *Accountability in Research* 8: 309–329.
- Redman BK (2006) Review of measurement instruments in clinical and research ethics, 1999–2003. *Journal of Medical Ethics* 32: 153–156.
- Ruzek JI and Zatzick DF (2000) Ethical considerations in research participation among acutely injured trauma survivors: An empirical investigation. *General Hospital Psychiatry* 22: 27–36.
- Sand K, Kaasa S and Loge JH (2010) The understanding of informed consent information—definitions and measurements in empirical studies. *American Journal of Bioethics Primary Research* 1(2): 4–24.
- Weinfurt KP, Seils DM, Lin L, et al. (2012) Research participants' high expectations of benefit in early-phase oncology trials: Are we asking the right question? *Journal of Clinical Oncology* 35: 4396–4400.