CLINICAL ETHICS

Review of measurement instruments in clinical and research ethics, 1999–2003

B K Redman

J Med Ethics 2006;32:153–156. doi: 10.1136/jme.2005.012567

Every field of practice has the responsibility to evaluate its outcomes and to test its theories. Evidence of the underdevelopment of measurement instruments in bioethics suggests that attending to strengthening existing instruments and developing new ones will facilitate the interpretation of accumulating bodies of research as well as the making of clinical judgements. A review of 65 instruments reported in the published literature showed 10 with even a minimal level of psychometric data. Two newly developed instruments provide examples of the full use of psychometric and ethical theory. Bioethicists use a wide range of methods for knowledge development and verification; each method should meet stringent standards of quality.

found 14 reporting some form of reliability and only one reporting validity for the method used. The lack of validity of measures used to assess children's competence is a significant limitation of prior research on consent, and the lack of a standard and operational definition of competence in research makes it very difficult to compare findings across studies. The majority of studies looked at samples of healthy children and not those with psychiatric or medical problems.²

A review of 35 instruments measuring ethical constructs and published during 1998 found many in the early stages of development.³ Only nine included at least four of the five pieces of information on validity and reliability (the same as those identified in Table 1) necessary to make a judgement about their adequacy as measurement instruments.

These few reviews show problems with: the use of interviews/impressions with no standard for the judgement that follows the measurement; groups of studies in which large numbers of instruments of unknown validity and reliability, and disparate definitions of the constructs being investigated are used; testing of instruments on samples unlike those for whom the problem exists; and the lack of instruments in particular areas. These measurement deficiencies result in an inability to compare findings across studies and unknown confidence in and difficulty with interpreting the accumulated body of research. This article: (1) reports a systematic review of instruments for measuring ethical constructs, defined as ideas commonly used in bioethics theory, research, or practice; (2) evaluates them according to psychometric standards; and (3) suggests the implications of the current level of development of measurement instruments in bioethics.

METHODS

In addition to a broad reading of the ethics and clinical literature, instruments were retrieved from PubMed (which has incorporated Bioethicsline) by searches under the term "ethics empirical research" alone and with "informed consent", "patient preference", "health care rationing", "patient selection", "study", "trust", "autonomy", "beneficence", "nonmaleficence", and "justice". Citations were screened for measurement instruments for which psychometric information was available and through citation indexes for other studies that had used the instruments. Instruments to measure moral development in individuals (including health professionals), decision making competence,

The prominent role played by bioethics in influencing societal values about health, illness, and medicine, and in everyday transactions in treatment and research, requires definition and measurement of its constructs. This is congruent with the concept of empirical ethics and is especially important for testing theories, for understanding the frequency with which particular ethically problematic situations occur, and for performance improvement in clinical and research settings.

Reviews of some areas of scholarship and practice in bioethics provide a glimpse of the level of development of measurement instruments. Kim and colleagues reviewed 32 studies (28 study samples) of decision making competence of cognitively impaired elderly persons.¹ Four used general clinical interviews/impressions as methods of assessing capacity. The remaining 24 used 18 different instruments to measure various decisional abilities. With few exceptions every research group developed its own instrument. The result of the variety of ways of assessing decisional abilities was that the constructs did not have a consistent meaning across studies. The overwhelming majority of studies used hypothetical decision making scenarios for testing the decisional abilities of cognitively impaired elderly individuals, with unknown generalisability to real world situations. With the absence of a clear criterion or reference standard of incompetence, some used a psychometric standard (two standard deviations below the mean), which may not result in ethically relevant categorisations.

A review of 29 studies of measures of children's competence for assent or consent

Correspondence to:
B K Redman, Wayne State
University, 5557 Cass
Avenue, Suite 100, Detroit,
MI 48202, USA; ae9080@
wayne.edu

Received 20 April 2005
In revised form
28 June 2005
Accepted for publication
29 June 2005

Table 1 Measurement characteristics of reviewed instruments

Instrument	Content validity	Construct validity	Internal consistency*	Test-retest reliability†	Sensitivity to intervention
Moral Distress Scale	Yes	Known groups	–	0.86	–
Hospital Ethical Climate Survey	Yes	Known groups	0.68–0.89	–	–
Ethical Issues Scale	Yes	FA, another instrument	0.74–0.91	–	–
Schedule of Attitudes Toward Hastened Death	Yes	FA	0.89	–	–
Preference of Life- Sustaining Treatment Questionnaire	Yes	Theory correlates FA	–	0.81	–
Domino Physician Assisted Suicide Scale	–	–	0.91	0.87	–
Multi-Dimensional Measure of Informed Choice	Yes	Predictive	0.82–0.83	–	–
		Discriminative	0.68–0.78		
Quality of Informed Consent	Yes	–	–	0.66–0.77	–
Reactions to Research Participation Questionnaire	–	FA	0.62–0.68	–	–
			0.78–0.80		
Trust in Physician Scale	–	Predictive directions, FA	0.85–0.91	0.77	–

*Standard is 0.7–0.9; middle 0.9's for judgements about individuals.

†Standard is 0.7–0.8.

FA, factor analysis.

and methods of economic appraisal and valuation of health states were not included as these have been adequately reviewed elsewhere.^{1 4 5}

Citations for the instruments reviewed by Redman in 2002 were updated to April 2004 and new instruments added.³ The criteria for inclusion in this review were: availability of information on reliability and validity; use of the instrument in two or more studies (necessary for a minimal level of accumulation of psychometric data), at least one of which was published between 1999 and 2003; and use in publications on bioethical theory, practice, or research.

Well-established sources suggest the kinds of psychometric data that should be available to evaluate a measurement instrument and numerous standards they should meet.⁶ Internal consistency reliability is a measure of the homogeneity of items in a scale; test-retest reliability measures the stability of scores over time. Validity describes the range of interpretations that can appropriately be placed on a measurement score. 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. 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. Responsiveness (also called sensitivity) is the ability to detect subtle but significant change, often as an outcome of an intervention.

RESULTS

Of the 35 instruments retrieved from the previous review³ and 30 from subsequent published literature, 10 met the criteria for inclusion. Of the 55 not meeting inclusion criteria, three provided no information on reliability or validity, 25 had been used only once, and 23 had been used more than once but not between 1999 and 2003, and therefore did not have recent psychometric information, one was in an area of bioethics not included in this review, and three were judged as used in psychosocial, not bioethical, research theory or practice. The 10 instruments that met the criteria are described below in five clusters: professional climate/issues, end-of-life, informed consent, research, and trust. Two additional instruments showed promise. Table 1 describes the types of psychometric data available for each instrument, including numerical standards where they exist.

Professional climate/issues

Moral Distress Scale

Moral distress is the painful psychological disequilibrium that results from recognising the ethically appropriate action yet not taking it because of obstacles such as lack of time, supervisory reluctance, an inhibiting medical power structure, institutional policy, or legal considerations. Moral distress is an element of job stress in nurses. The Moral Distress Scale is used with nurses caring for adults in hospitals. Occupational health nurses did not show moral distress on this scale; critical care nurses did (known groups). Fifteen per cent of the nurses studied had left a position because of moral distress. The most distressing items dealt with working where the number of staff was so low that care was inadequate, and carrying out physicians' orders for unnecessary tests and treatments for terminally ill patients.^{7 8}

Factors predicting moral distress, interventions to decrease it, and its impact on patients have not been identified. Longitudinal studies are necessary to discover what part of moral distress precipitates nurse resignation.⁹

Hospital Ethical Climate Survey

The ethical climate of an organisation, as perceived by a group of its workers, is believed to affect ethical practice, job satisfaction, and quality of care. Climate can be viewed as a set of institutional practices and assessed by measuring how decisions having ethical content are solved, or the presence of organisational conditions that allow employees to engage in ethical reflection, or both. These include the option to disagree with one another, the inclusion of those with a stake in decisions, access to information to make informed decisions, and encouragement of questioning and debate.¹⁰

Content validity was supported by expert judgement and construct validity by findings of a relationship between the Hospital Ethical Climate Survey and the Integrity Audit, and by the known group technique. Five factors were found, but a different five factors were identified in the Turkish version of this instrument.^{10 11}

Ethical Issues Scale

The Ethical Issues Scale is a survey instrument used to measure the frequency of nurses' encounters with ethical issues in practice. Items were developed from the literature on ethical conflict in nursing practice, from focus group interviews of practising nurses, and from a panel of nurses with expertise in bioethics. Factor analysis showed three scales: end of life treatment decisions, patient care issues, and human rights issues. Comparison with other methods of

measurement has not yet been accomplished. This scale has been used in statewide samples of practising nurses in Maryland and in the New England states.^{12 13}

End of life

Schedule of Attitudes Toward Hastened Death

Understanding why terminally ill patients may desire a more rapid death or request physician assisted suicide has become an important element in both palliative care and public policy. The Schedule of Attitudes Toward Hastened Death has been used with persons with AIDS and amyotrophic lateral sclerosis, and with those who are terminally ill with cancer.^{14–16} Content validity was established by experts in palliative care and consultation-liaison psychiatry. The total score was significantly correlated with clinician ratings on the Desire for Death Rating Scale, with ratings of depression and psychological distress, with pain intensity and physical symptom distress, and, in another study, with hopelessness, suffering, and anxiety.^{15 17} Factor analysis showed a single factor.¹⁸ Depression and hopelessness were the strongest predictors of desired death, supporting validity of the Schedule.¹⁶

Preference of Life-Sustaining Treatment Questionnaire

This questionnaire was developed to study whether elderly hospitalised patients formed preferences regarding the use of specific medical treatments and the level of certainty with which these preferences were made. It assesses patients' preferences for or against nine treatment options: temporary or permanent tube feeding, resuscitation, temporary or permanent use of a respirator, amputation, kidney dialysis, radiation, chemotherapy, antibiotics, and blood transfusion, under hypothetical conditions of intact cognitive ability, permanently confused, and permanently unconscious. Content validity was estimated from geriatric physicians and nurses.¹⁹ Recently, this scale has been shortened for use in busy clinical settings, with little loss of accuracy.²⁰

Domino Physician Assisted Suicide Scale

This is apparently the only available instrument measuring attitudes toward physician assisted suicide for which data on psychometric characteristics are available. It has been used with college students who will be informed citizens of the future, adult caregivers of elderly parents, and elderly individuals close to the end of life. One major factor was found.^{21 22}

Informed consent

Multidimensional Measure of Informed Choice

An informed choice is based on relevant knowledge, consistent with the decision maker's values, and is behaviourally implemented. This measure consists of an eight-item scale of knowledge of Down's syndrome, a four-item scale assessing attitudes towards undergoing a serum screening test for the syndrome, and a record of test uptake. Content validity was assessed by comparing women's scores on the knowledge scale with their responses to open-ended questions designed to elicit their understanding of the test.²³ Those who made a choice rated on the Multidimensional Measure of Informed Choice as informed 6 weeks later rated their decision as being more informed, better supported, and of higher quality than did women whose choice was categorised as uninformed (predictive validity). A lack of association between this measure and anxiety shows discriminant validity. By itself, knowledge appears not to be a good proxy for informed choice.²⁴

Quality of Informed Consent

Standardised methods for assessing the adequacy of informed consent to research have been lacking but could

be used by institutional review boards for monitoring purposes and as an outcome measure for interventions intended to improve informed consent. This instrument incorporates the basic elements of informed consent specified in federal regulations, assesses therapeutic misconceptions, and measures actual (objective) and perceived (subjective) understanding of cancer clinical trials. Content validation came from two independent panels. There is an adult form and one for parents, written at eighth grade reading level, taking 7 minutes to complete. It is meant to be used only with cancer trials and with trials that conform to the conventional phases I, II, and III.^{25 26}

Research

Reactions to Research Participation Questionnaire

Researchers and institutional review boards typically identify potential costs and benefits of research by relying on common sense, clinical judgement, prior experience, and imagined/personal substitution with the participant; these are subjective appraisals and may be based on biased opinions and untested assumptions. This questionnaire identifies human participants' perceptions of the costs and benefits of research participation and their responses to research procedures such as informed consent and recruitment. Factor analysis showed five factors, two of which aligned with the ethical constructs of benefit and cost-risk ratio.²⁷ Kassam-Adams and Newman developed a Reactions to Research Participation Questionnaire for children and one for parents.²⁸

Trust

Trust in Physician Scale

Interpersonal trust in a clinical context is a person's belief that the physician's words and actions are credible and can be relied upon.²⁹ Trust was correlated with patient satisfaction with the physician and the length of the relationship, and was higher among patients who actively chose and predicted continuity with their physician.³⁰ Construct validity was supported by an inverse correlation between higher trust scores and scepticism. The scale loaded on a single factor.³¹ Kraetschmer and colleagues found that autonomous patients had relatively low levels of trust, with passive respondents (more frequently female, those with less education, and aged over 65 years) more likely to display blind trust.³²

This scale can help to identify physician and patient behaviours that promote or block trust and relationships characterised by low trust, and test methods to increase it.³⁰

Instruments showing special promise

Two among the many promising instruments that have been used only once are nevertheless worthy of mention, one because the approach to its development is psychometrically sophisticated, and the other because it is closely linked with ethical theory.

Preferences for Care Near the End of Life

Although 70% of deaths in US hospitals have been preceded by decisions to limit medical treatment, authors do not agree on ways to name the multiple dimensions of preferences for care near the end of life. Content validity was estimated by the usual literature review and advice of experts, but then placed in a subject matter grid to assure that various dimensions were covered. Factor analysis shows five factors. Concurrent validity could be tested by comparing scores to expressed preferences on an advance directive. Predictive validity could be studied by noting whether this measure predicts decision making when it is time to express wishes. The instrument was shown to be stable in a group of healthy adults but must be tested with chronically ill and terminally ill participants.³³

Ideal Patient Autonomy Scale

Evidence based patient choice is founded on a strong liberal, individualist interpretation of patient autonomy. Not all patients are in favour of such an interpretation and should not be held to a standard they are unlikely to satisfy. Two scales reflect the paternalistic and consumerist poles of the liberalist individualist model, a third reflects Socratic autonomy and procedural independence, and the fourth scale reflects ideals of risk disclosure. The Ideal Patient Autonomy Scale is clearly distinct from the generally used psychological preference instruments. In an elderly population with low levels of education, the factor structure was: (1) the doctor knows best; (2) the patient should decide; (3) the patient is entitled to a wish not to participate; and (4) the provision of obligatory risk information. A correlation between the doctor knows best and the right to non-participation was in the expected direction, supporting construct validity. This scale is constructed with the goal of providing data for improving existing moral theories.³⁴

Implications of the current level of development of measurement instruments

From the Table 1 summary of measurement characteristics of the 10 instruments meeting the study criteria, it is possible to see that most had estimates of reliability and evidence of content validity, and some, but not all, forms of construct validity. None had established sensitivity to interventions even though most were designed in part to measure the effects of an intervention. Content validity was frequently assumed by judgement of professional experts and not by patients/participants. The 10 instruments represent a narrow range of the entire field of important ethical constructs. From the pattern of instrument development, one may assume that only nurses suffer from moral distress and only physicians are concerned with trust.

These instruments serve several purposes: (1) describing compliance with existing moral norms (Quality of Informed Consent, Trust in Physician Scale); (2) empirical testing of practices prescribed by normative theories (Preference of Life-Sustaining Treatment Questionnaire, Multidimensional Measure of Informed Choice, Reaction to Research Participation Questionnaire); and (3) generating new material for normative study (Moral Distress Scale, Hospital Ethical Climate Survey, Ethical Issues Scale).³⁵

Two limitations of this study must be noted. In the process of screening the literature, the first or second use of an instrument might have been missed, other search terms might have yielded additional instruments, additional instruments might have qualified for inclusion beyond the 5-year time span of 1999–2003, and citation indices do not include all journals. Secondly, judgement about what constitutes an ethical construct may vary. Despite these limitations, this review reaches a conclusion similar to the reviews cited in the opening paragraph of this article, that the development of instruments designed to measure bioethics constructs in the areas reviewed is in an early stage of development.

Bioethics uses a wide range of methods for knowledge development and verification, although the field of psychology, from whence the theory and practice of psychometrics comes, has historically not been as well integrated into bioethics as have the social sciences.³⁵ The considerable underdevelopment of measurement instruments in bioethics may well be impeding the development of important bodies of research. This is especially true in an era of evidence based medicine that proclaims that what happens to patients should be founded, to the greatest extent possible, on evidence.³⁶

Competing interests: none declared

REFERENCES

- Kim SYH, Karlawish JHT, Caine, eds. Current state of research on decision-making competence of cognitively impaired elderly persons. *Am J Geriatr Psychiatry* 2002;10:151–65.
- Miller VA, Drotar D, Kodish E. Children's competence for assent and consent: a review of empirical findings. *Ethics Behav* 2004;14:255–95.
- Redman BK. *Measurement instruments in clinical ethics*. Thousand Oaks, CA: SAGE, 2002:283–5.
- Rest JR, Narvaez D. *Moral development in the professions: psychology and applied ethics*. Hillsdale, NJ: Lawrence Erlbaum, 1994:10–12.
- Newbold D. A brief description of the methods of economic appraisal and the valuation of health states. *J Adv Nurs* 1995;21:325–33.
- McDowell I, Newell C. *Measuring health*, 2nd ed. New York: Oxford University Press, 1996:29–42.
- Corley MC, Elswich RK, Gorman M, et al. Development and evaluation of a moral distress scale. *J Adv Nurs* 2001;33:250–6.
- Meltzer LS, Huckabay LM. Critical care nurses' perceptions of futile care and its effect on burnout. *Am J Crit Care* 2004;13:202–8.
- Corley MC. Nurse moral distress: a proposed theory and research agenda. *Nurs Ethics* 2002;9:636–49.
- Olson LL. Hospital nurses' perceptions of the ethical climate of their work setting. *Image J Nurs Sch* 1998;30:345–9.
- Bahcecik N, Ozturk H. The Hospital Ethical Climate Survey in Turkey. *JONAS Healthc Law Ethics Regul* 2003;5(4):94–9.
- Fry ST, Damrosch S. Ethics and human rights issues in nursing practice. A survey of Maryland nurses. *Mid Nurse* 1994;13(7):11–12.
- Fry ST, Duffy ME. The development and psychometric evaluation of the Ethical Issues Scale. *J Nurs Scholarsh* 2001;33:273–7.
- Pessin H, Rosenfeld B, Burton L, et al. The role of cognitive impairment in desire for hastened death: a study of patients with advanced AIDS. *Gen Hosp Psychiatry* 2003;25:194–9.
- Rabkin JG, Wagner GJ, Del Bene M. Resilience and distress among amyotrophic lateral sclerosis patients and caregivers. *Psychosom Med* 2000;62:271–9.
- Breitbart W. Depression, hopelessness, and desire for hastened death in terminally ill patients with cancer. *JAMA* 2000;284:2907–11.
- Rosenfeld B, Breitbart W, Stein K, et al. Measuring desire for death among patients with HIV/AIDS: the Schedule of Attitudes Toward Hastened Death. *Am J Psychiatry* 1999;156(1):94–100.
- Rosenfeld B, Breitbart W, Galiotta M, et al. The Schedule of Attitudes Toward Hastened Death. *Cancer* 2000;88:2868–75.
- Cohen-Mansfield J, Rabinovich BA, Lipson S, et al. The decision to execute a durable power of attorney for health care and preferences regarding the utilization of life-sustaining treatments in nursing home residents. *Arch Intern Med* 1991;151:289–94.
- Cohen-Mansfield J. Development of a framework to encourage addressing advance directives when resources are limited. *J Aging Health* 2002;14:24–41.
- Domino G, Kempton S, Cavender J. Physician assisted suicide: a scale and some empirical findings. *Omega* 1996–1997;34:247–57.
- Domino G. Community attitudes toward physician assisted suicide. *Omega* 2002;46:199–210.
- Marteau TM, Dormandy E, Michie S. A measure of informed choice. *Health Expectations* 2001;4:99–108.
- Michie S, Dormandy E, Marteau TM. The multi-dimensional measure of informed choice: a validation study. *Patient Educ Couns* 2002;48:87–91.
- Joffe S, Cook EF, Cleary PD, et al. Quality of informed consent: a new measure of understanding among research subjects. *J Natl Cancer Inst* 2001;93:139–47.
- Joffe S, Cook EF, Cleary PD, et al. Quality of informed consent in cancer clinical trials: a cross-sectional survey. *Lancet* 2001;358:1772–7.
- Newman E, Willard T, Sinclair R, Kaloupek D. Empirically supported ethical research practice: the costs and benefits of research from the participants' view. *Accountability in Research* 2001;8:309–29.
- Kassam-Adams N, Newman E. The reactions to research participation questionnaires for children and parents (RRPQ-C and RRPQ-P). *Gen Hosp Psychiatry* 2002;24:336–42.
- Anderson LA, Dedrick RF. Development of the Trust in Physician Scale: a measure to assess interpersonal trust in patient-physician relationships. *Psychol Rep* 1990;67:1091–100.
- Thom DH, Ribisl KM, Stewart AL, et al. Further validation and reliability testing of the Trust in Physician Scale. *Med Care* 1999;37:510–17.
- Freburger JK, Callahan LF, Currey SS, et al. Use of the Trust in Physician Scale in patients with rheumatic disease: psychometric properties and correlates of trust in the rheumatologist. *Arthritis Care Res* 2003;49:51–8.
- Kraetschmer N, Sharpe N, Urowitz S, et al. How does trust affect patient preferences for participation in decision-making? *Health Expectations* 2004;7:317–26.
- Gauthier DM, Froman RD. Preferences for care near the end of life: scale development and validation. *Res Nurs Health* 2001;24:298–306.
- Stiggelbout AM, Molewijk AC, Otten W, et al. Ideals of patient autonomy in clinical decision making: a study on the development of a scale to assess patients' and physicians' views. *J Med Ethics* 2004;30:268–74.
- Sugarman J, Sulmasy DP. *Methods in medical ethics*. Washington, DC: Georgetown University Press, 2001:11–13.
- Eddy DM. Evidence-based medicine: a unified approach. *Health Aff* 2005;24(1):9–17.