How will the sustainable development goals deliver changes in well-being? A systematic review and meta-analysis to investigate whether WHOQOL-BREF scores respond to change

Suzanne M Skevington, Tracy Epton

International Hub for Quality of Life Research, Manchester Centre for Health Psychology, Division of Psychology and Mental Health, University of Manchester, Manchester, UK

Correspondence to
Dr Suzanne M Skevington; suzanne.skevington@manchester.ac.uk

ABSTRACT

Introduction The Sustainable Development Goals (SDGs) 2015 aim to '…promote well-being for all', but this has raised questions about how its targets will be evaluated. A cross-cultural measure of subjective perspectives is needed to complement objective indicators in showing whether SDGs improve well-being. The WHOQOL-BREF offers a short, generic, subjective quality of life (QoL) measure, developed with lay people in 15 cultures worldwide; 25 important dimensions are scored in environmental, social, physical and psychological domains. Although validity and reliability are demonstrated, clarity is needed on whether scores respond sensitively to changes induced by treatments, interventions and major life events. We address this aim.

Methods The WHOQOL-BREF responsiveness literature was systematically searched (Web of Science, PubMed, EMBASE and Medline). From 117 papers, 15 (24 studies) (n=2084) were included in a meta-analysis. Effect sizes (Cohen's d) assessed whether domain scores changed significantly during interventions/events, and whether such changes are relevant and meaningful to managing clinical and social change.

Results Scores changed significantly over time on all domains: small to moderate for physical (d=0.37; CI 0.25 to 0.49) and psychological QoL (d=0.22; CI 0.14 to 0.30), and small for social (d=0.10; CI 0.05 to 0.15) and environmental QoL (d=0.12; CI 0.06 to 0.18). More importantly, effect size was significant for every domain (p<0.001), indicating clinically relevant change, even when differences are small. Domains remained equally responsive regardless of sample age, gender and evaluation interval.

Conclusion International evidence from 11 cultures shows that all WHOQOL-BREF domains detect relevant, meaningful change, indicating its suitability to assess SDG well-being targets.

INTRODUCTION

When the Sustainable Development Goals (SDGs) superseded the Millennium Development Goals (MDGs)[1] in 2015, the United Nations (UN) built on MDG success in meeting the needs of low/middle income countries, and renewed plans for action and change.[2] While SDG 3 aims to 'ensure healthy lives, and promote well-being for all, at all ages', this forms only part of a broader health and well-being (WB) brief within the SDGs. During the MDG programme, environmental sustainability became a growing concern, as health and WB costs from hazardous events (eg, droughts, flooding) escalated, and these are increasingly linked to climate change.[3] In this paper, we inquire how we will know whether the SDGs have significantly improved WB in 2030. Evaluation is a central theme of this work.

Learning from the MDGs

Despite considerable international agreement about MDG aims and interventions to implement them, decisions were inconclusive about which outcome measures should evaluate MDG goals. Effective MDG interventions included training village health workers, free insecticide treatment of bed nets, eliminating basic health service fees in low/middle-income countries, improving access to sexual and reproductive care and offering combination drug therapies for HIV infection.[4] One line of thinking is that outcome measures should be intervention-specific, and hence different assessments are needed to address heterogeneous interventions designed to improve quality of life (QoL) and survival.[4] Although specific outcome measures are suited to evaluating single, and sometimes similar interventions, their data limit conclusions when outcomes from diverse targets need to be compared. Despite broad agreement that quality, timely and reliable disaggregated data are essential to measuring progress, and recognition that gross domestic product

Key questions

What is already known about this topic?
► Sustainable Development Goal 3 (SG3) aims to '… promote well-being for all, at all ages' through targeting new interventions, but subjective assessments are needed to complete and complement existing objective indices (eg, gross domestic product), and enable accurate conclusions to be drawn in 2030.
► A cross-cultural assessment—the WHOQOL-BREF— offers a unique subjective, generic quality of life (QoL) measure designed with users, to assess global health and well-being; data exist on 60 000 adults living in 100 countries.
► Although validity and reliability is established, it is unclear whether WHOQOL-BREF scores respond to changes induced by treatment, interventions and major life events; a systematic review of the literature with meta-analysis examines whether these changes are clinically and socially meaningful.

What are the new findings?
► When changes in scores for the four WHOQOL-BREF domains were tested (eg, before and after treatment), both physical and psychological QoL domains showed small-to-moderate changes; small changes were found for the social and environmental QoL domains.
► Nevertheless, the effect size for every domain was significant, indicating that each domain delivers relevant and meaningful change, even when the differences are small.
► WHOQOL-BREF scores remained responsive, regardless of the sample age, gender and evaluation interval.

Recommendations for policy
► The unique features and high performance of the WHOQOL BREF make it suitable for monitoring subjective QoL in SDG projects where well-being changes are expected (eg, SDG 3; SDG 9; SDG 17) and also to compare achievements of global targets, nationally and internationally.
► As WHOQOL-BREF domains detect relevant changes in QoL following treatments, interventions and major life events (eg, earthquake, abuse), practitioners, policymakers and researchers in clinical and social fields can be assured of measurement responsiveness.

(GDP) alone would be insufficient to evaluate the SDGs, in 2015 the UN[2] reported low agreement about which alternative and additional measures to select. Although good generic measures are available to compare many different populations, conditions and settings, rather than endorsing one of these, the UN-SDG panel decided to go 'beyond GDP' by developing new measures of progress; consequently, they proposed subjective well-being (SWB) indicators, for example, positive mood[5] (Sustainable Development Solutions Network, p69)[5]. Without generic subjective data from global cultures, it will not be possible to draw conclusions about progress in 2030 when SDG outcomes are finally and formally evaluated. Is it possible to find out whether 'anyone has been left behind'[3] (ICUN, UNEP, WWF, p3–35)[3], without asking about perceptions?

Within the context of this debate, we consider an international, well-developed, multilingual, generic tool, with an established track record. Acceptable and feasible to assess subjective QoL in many cultures worldwide, the WHOQOL-BREF was developed through an international collaboration convened by the WHO, Geneva. The WHOQOL group aimed to advance standards in cross-cultural measurement, and produce a global measure that would complement 'objective' indices on standard of living (eg, GDP) and health (eg, mortality rate).[6] Multiple language versions of the WHOQOL were simultaneously developed in 15 cultures worldwide, from an internationally agreed protocol which has since been used by a network of developers in around 100 cultures. Endorsement of a suitable subjective measure to evaluate SDG goals will enable ongoing projects to take immediate advantage of evaluating key outcomes from interventions, treatments and major life events, using a common metric.

Requirements for evaluating the SDGs
A good subjective measure for SDG purposes would need to show (a) measurement properties of reliability and validity, at global, regional and national levels, and moreover that its scores respond sensitively to changes in people's lives; (b) a capacity to detect positive and negative changes, reflecting experience; (c) acceptability to different cultures, and feasibility to use; (d) easy administration and scoring, with scores that are readily interpretable by decision makers and practitioners; (e) a capacity to assess well people, including those receiving health promotion or disease prevention interventions, and unhealthy populations of all types (eg, chronic mental health, neglected tropical diseases, intractable infections, non-communicable diseases, road traffic injuries); (f) that the results can indicate where inequalities between groups exist, so that 'nobody is left behind'; and (g) standard procedures to enable cultural adaption, translation and standardisation when new language versions are needed, potentially for 7300+ cultures worldwide. Such properties will enable sound comparisons of diverse cultures, for example community happiness in Bhutan; berry picking in Sweden; 'feeling fed-up' in Britain. The availability of this type of tool opens up opportunities for 'invisible', indigenous groups (eg, 'first nations'), to give 'voice' to their QoL views when engaging with SDG projects, and perhaps for the first time, to communicate them to policymakers.

Although generic economic indices like GDP and population health estimates (eg, disability-adjusted life years) were used in MDG evaluations, no single measure can fulfil this challenging brief. When it comes to SWB and health, there is no substitute for obtaining meaningful information directly from individuals about their personal experiences, and then testing whether the intervention significantly changes it. Although the UN acknowledged the importance of assessing experience in 2015, and designated WB (defined as positive mood) as an indicator for SDG 17 among 100 indicators in total,[5] by March 2016, WB had disappeared from

the new final list of 230 UN approved SDG indicators.[7] Furthermore, all 230 are exclusively objective. Without subjective empirical data, comparing SDG achievements remains partial and therefore contentious. Incomplete data for MDG targets frustrated global policymaking during the planning period leading to publication of the SDG plan in 2015: they reported 'progress was limited and hard won'.[2] A global evaluation strategy of using a common metric is urgently needed to record the personal experience of all communities participating in SDG projects in the next 12 years. A standardised subjective measure added to existing population-based indicators would provide a more comprehensive and conclusive evaluation. Without sound data it will be impossible to conclude whether WB has been significantly improved by the SDG programme, wasting more years, and considerable human and financial resources.

WB or QoL?

Although promoting WB is a specified aim of SDG 3, the UN[2] did not say whether WB or QoL should be measured alongside health. Conceptual distinctions between subjective QoL and SWB are opaque, raising questions about whether these concepts are identical, 'nested' within each other or different.[8] Lack of clarity confounds policymakers and practitioners who seek to make informed choices about the best measure to use. QoL assessment is well suited to measuring SDG 3 outcomes, with its reputation for rigorously developed multidimensional instruments that are relevant to mental and physical health. Furthermore, the five dimensions of mood (positive and negative), cognition and evaluation that typically compose SWB[9] can be readily mapped onto some of the 25 QoL dimensions assessed by the WHOQOL. New international empirical findings[10] strongly suggest that SWB is 'nested' within QoL. QoL as measured by the WHOQOL offers a scale with exceptional multidimensionality and conceptual integration, so is highly appropriate to address the very wide range of life qualities addressed by the 17 SDGs.

WB has also been viewed as purely 'objective', and assessed by counting material household goods (eg, TVs, cars, house size),[11] but some indices show a loose relationship with subjective measures. Over a decade of analysis examining the relationship between GDP and SWB, Easterlin and Sawangfa[12] showed that this association is neither strong nor linear, as generally assumed. While SWB in lower income countries does increase as GDP rises, among higher income countries rising GDP has little or no association with SWB,[12] illustrating that objective GDP indicators are no substitute for subjective measures of WB or QoL.

Although the history of health-related QoL assessment reveals several high-quality generic measures (eg, European Quality of Life Scale-five dimensions (EQ-5D), Short Form-36 Health Survey (SF-36)), the unique conceptual breadth of the WHOQOL with its 25 facets scored in four domains enables WB to be evaluated for many

SDGs, for health goals where WB is intrinsic (SDG 3), and for other SDGs (SDG 11) eg, where QoL evidence could show that people feel physically safer and secure in sustainable cities. In SDG 17, better QoL could strengthen ways to implement and revitalise the global partnership for sustainable development. Multiple dimensions of a profile are especially valuable for unpacking the impact the of QoL dimensions where trade-offs are unknown; for example, action to conserve biodiversity may enhance psychological QoL, and may also conflict with more immediate human needs for adequate housing, better nutrition and being employed (relating to SDG 15, SDG 2 and SDG 16),[2] so potentially abrading environmental QoL. Including a quality generic, subjective profile could complement objective economic measures (eg, GDP), and balance the evaluation portfolio.

Advances in QoL assessment

The WHOQOL suite of instruments offers a series of patient and person-reported outcome measures (PROMs) that provide insights into thoughts, feelings and personal experiences relating to QoL and health. The WHOQOL was the first generic, cross-cultural PROM designed explicitly for global use. Focus groups of patients, health professionals and community members were convened in 15 cultures, pooling their proposed concepts and wording for the questionnaire. Due to the way it was developed, it is highly acceptable to users. Acceptability promotes lower survey refusal rates and reduces missing data, which in turn improve the accuracy that decision makers need when drawing conclusions and taking action. Improvements to validity that arise from using PROMs methods have accelerated routine use in randomised controlled trials (see https://commonfund.nih.gov/promis/overview).

A strength of the WHOQOL development was the innovatory 'spoke-wheel' methodology that was devised by a multidisciplinary, multinational collaboration to simultaneously create many highly equivalent language versions from a commonly agreed concept.[13] Issues of national importance to QoL were identified locally, and then appended to WHOQOL-BREF core items, to complete the QoL concept for that country,[14] for example, eating and appetite in Hong Kong. This process tailored the evaluation to the culture, and hence resonates with the UN Inter Agency Expert Group policy[7] to allow countries to devise culturally appropriate indicators that are not on the approved international SDG list. National progress in SDG targets will be tracked then later aggregated up to regional and international levels.

Over 20 years, a WHOQOL-BREF manual has facilitated around 100 culturally adapted translations globally, completed by over 60 000 sick and well adults. This process potentially enables universal QoL comparisons between cultures as diverse as Assam tea pickers, Wall Street bankers and Peruvian mineworkers, by producing highly equivalent cross-cultural data. Similarly, the WHOQOL-BREF's globally agreed importance items

are clustered and scored in physical, psychological, social and environmental QoL domains. This breadth of dimensions uniquely extends the WHOQOL concept well beyond the physical and psychological, commonly found in health-related QoL measures, as it includes QoL related to social relationships, the environment (eg, physical safety, financial resources, home environment), and an unusual spiritual, religious and personal beliefs component within the psychological domain (see detail in references [6 15 16]). These features extend its application to innovative settings beyond health. A profile of scores offers a fine-grained multidimensional analysis, and a priori predictions can be made about which domains and items will likely respond most to the impact of an intervention. Furthermore, the WHOQOL-BREF together with its importance ratings has been recently used at a one-to-one level to assess QoL, and prioritise individual patient healthcare.[17 18] This research is underpinned by the WHO definition of QoL: 'An individual's perception of their position in life, in the context of the culture and value systems in which they live, and in relation to their goals, expectations, standards and concerns' (The WHOQOL Group, p. 43)[19]. Briefly, QoL is subjective and in the eye of the beholder, as it is invisible to observers; hence, it is best judged by the person themselves.[20]

The WHOQOL-BREF results are an aid to policymaking and allocating budgetary resources in the US State of Connecticut Department of Mental Health and Addiction Services (DMHAS).[21] Since 2007, 14 000+ adult mental health service users have been invited to complete the WHOQOL-BREF annually, which takes about 6 min in English. This evidence is disaggregated by service, demographics and health, enabling policymakers to prioritise service delivery to disadvantaged groups with the aim of redressing QoL deficits. Following a 3-year pilot, DMHAS annually allocates budgetary resources to promote effective interventions, and make service improvements,[21] illustrating how disaggregated WHOQOL-BREF data (eg, by age, education, gender, culture) can be used to reduce inequalities. It would therefore be valuable for delivering SDG 3 targets to relevant groups, being eminently suited to solving global policy issues.

The importance of responsiveness

Global psychometric properties of the WHOQOL-BREF show that scores are reliable and valid across diverse cultures,[15] and international reference data are available for gender, age and health status groups.[16] For some countries, national norms are available (eg, Australia).[22] Other studies show that the scores distinguish sick people from well, and across a wide range of physical and psychological disorders, and social and environmental conditions;[23] this is evidence of good discriminant validity.

Among key psychometric properties, responsiveness (also sometimes referred to as sensitivity to change) is perhaps the most important requirement. In clinical healthcare and interventions, responsiveness shows the degree to which score changes are clinically relevant.

Responsiveness is the ability of a measure to detect clinical change, particularly where differences in outcomes occur, including small ones. Responsive instruments have scores with the capacity to change sensitively and appropriately, in accordance with the person's condition or situation, and this information needs to be interpreted. However, there are many different ways of measuring responsiveness. In a longitudinal study of depressed patients in primary care in six cultures who received anti-depressant medication over 9 months, changes in WHOQOL-BREF domain scores were examined in relation to changing depressive symptoms to see how much QoL scores would increase if depression improved, and decrease if deterioration occurred.[24] Without such measurement information about change and its relevance, the success of an intervention or trial remains unknown.

The systematic review (SR) technique used in this study can assess responsiveness to change by combining and assessing known evidence in meta-analysis. As methods of calculating responsiveness are controversial,[25] we followed Norman *et al.*'s recommendation[26] to limit calculations to Cohen's effect size.[27] Through evaluating collective knowledge about WHOQOL-BREF responsiveness, we shall draw conclusions about the clinical and social relevance of this measure. For instance, this information has clinical value when managing long-term treatment for chronic conditions, such as Parkinson's disease.[20]

Responsive instruments can then be used to monitor new services, modes of service delivery, the impact of policy changes, reactions to political crises and environmental disasters, such as those outlined as SDGs. For SDG targets, it will be essential to determine whether QoL has improved significantly following planned interventions for instance, providing adequate and reliable nutrition to improve QoL. This question could be investigated in Addis Abba today using the Amharic version of the WHOQOL-BREF, and support action in this low income country. As no other generic global health-related QoL measure has the same capacity to embrace the plethora of situations and conditions outlined in the SDGs, the WHOQOL-BREF should be the subjective measure of choice.

Establishing the responsiveness of social and environmental domains in the WHOQOL-BREF is important to understanding the breadth of QoL impacts from adverse conditions (eg, earthquakes) and afterwards. As yet, the WHOQOL-BREF's responsiveness has not been comprehensively examined with regard to a variety of situations and populations, despite several studies which assessed outcomes after diverse interventions and major life events. Examples include radiation exposure,[28] earthquake survival,[29] housing elderly Sichuan Chinese in an earthquake zone,[30] Palestinian conflict in Gaza,[31] Nigerian refugees,[32] women torture survivors,[33] slum upgrades[34] and poverty in low/middle-income countries.[35] By using the WHOQOL-BREF in these challenging fields, this

indicates that social and development practitioners view it as a suitable tool.

The ability of instrument scores to respond to changing conditions is perhaps the most important psychometric property of health and WB measures. This property is no more applicable than when monitoring change induced by environmental and social interventions, including many addressing SDG targets. Gathering multidimensional subjective information about how QoL is affected provides much needed information when trying to understand why some interventions succeed, and if they fail, how that intervention might be adjusted to become more effective.

The main aim of the present research was to conduct a systematic review and meta-analysis of global evidence on the responsiveness of the WHOQOL-BREF to change. Furthermore, to assess the relevance or meaning of WHOQOL-BREF domain scores in clinical and social settings.

METHODS
Search strategy
Keyword searches were conducted through Web of Science/Knowledge, PubMed, Medline and EMBASE, up to May 2017. There were two search filters: the first identified use of the WHOQOL-BREF through the search terms 'WHOQOL-BREF' and 'WHOQOL'. The second filter was for responsiveness, searching the terms 'responsive', 'responsiveness', 'sensitivity' and 'sensitivity to change'. A grey literature search accessed documents from international governmental organisations (eg, UN; WHO); a hand search of government organisations was conducted.

Inclusion/exclusion criteria
For inclusion in this review, studies had to (a) assess responsiveness of the WHOQOL-BREF by reporting measures for all four domains, (b) include adult participants (adult was defined by the study culture), (c) include data (or provide data on request) that could be used to calculate Cohen's d, (d) score the WHOQOL-BREF according to manual instructions, (e) provide a report in English or French (or translation) and (f) apply a longitudinal or repeated measures design (recommended by the Scientific and Advisory Committee of the Medical Outcomes (SACMOS) Trust).[36] Studies with the following designs were included: (a) pre-intervention and postintervention/treatment compared with control, (b) postevent compared with a control group who had not experienced this event, (c) pre-intervention and postintervention/treatment with no control, (d) pre-event and postevent with no control, and (e) change over time as an event/illness progressed.

Studies were excluded if (a) they had not assessed WHOQOL-BREF responsiveness for all four domains; (b) participants were not adult; (c) they did not report (or provide) sufficient information to calculate

responsiveness; (d) WHOQOL-BREF scoring and transformation did not follow official procedures; (e) translations other than English or French were unavailable; and (f) a longitudinal or repeated measures design was not used.

Moderator variables
Information about potential moderator variables was extracted for each study. These were: characteristics of the sample (ie, mean participant age, percentage of females), and study characteristics (ie, the interval of time between baseline and follow-up assessments).

Coding reliability
All data extraction was completed independently by both authors. Where disagreements arose, these were resolved by discussion between the two, until 100% agreement was reached.

Meta-analytic strategy
The effect size used in the analysis was Cohen's d.[27] Values of effect size were interpreted according to Cohen's threshold criteria: <0.20 is interpreted as trivial, 0.20 to 0.50 as small, 0.50 to 0.80 moderate and >0.80 large. For independent samples, means, SDs and Ns at pre-intervention and postintervention were used to calculate effect size. If only postintervention means, SDs and Ns were reported, these were used. For repeated measures samples, means, SDs, N and correlation between two time points were used. If these data were unavailable, mean change, N and t value were used, or mean change, N and P value.

A random effects model weighted by sample size was used to calculate the overall effect size (d), the CI, the significance of heterogeneity (Q) and the extent of heterogeneity (I^2). The *metan* command was used in STATA V.11, for overall effect size,[37] and the *metareg* command was used to explore moderators.

RESULTS
Study selection
Figure 1 shows the flow of articles in the study. The literature search identified 117 articles. After eliminating duplicates, this left 83 articles for screening. After assessing the full text of 59 articles for eligibility, a total of 15 papers containing 24 studies were included in the final review. Key reasons for exclusions were they (a) presented responsiveness for WHOQOL measures other than the WHOQOL-BREF (eg, WHOQOL-100) (n=4); (b) collected WHOQOL-BREF data as the 'gold standard', but only reported the responsiveness of other measures (n=12); (c) presented reports in languages without a translation (eg, Chinese, Korean) (n=3); (d) did not report results for all four domains (n=5); (e) did not provide sufficient data to calculate effect size (n=5); (f) showed scoring problems which could not be rectified (n=6); and (g) provided literature reviews without useful data (n=9).
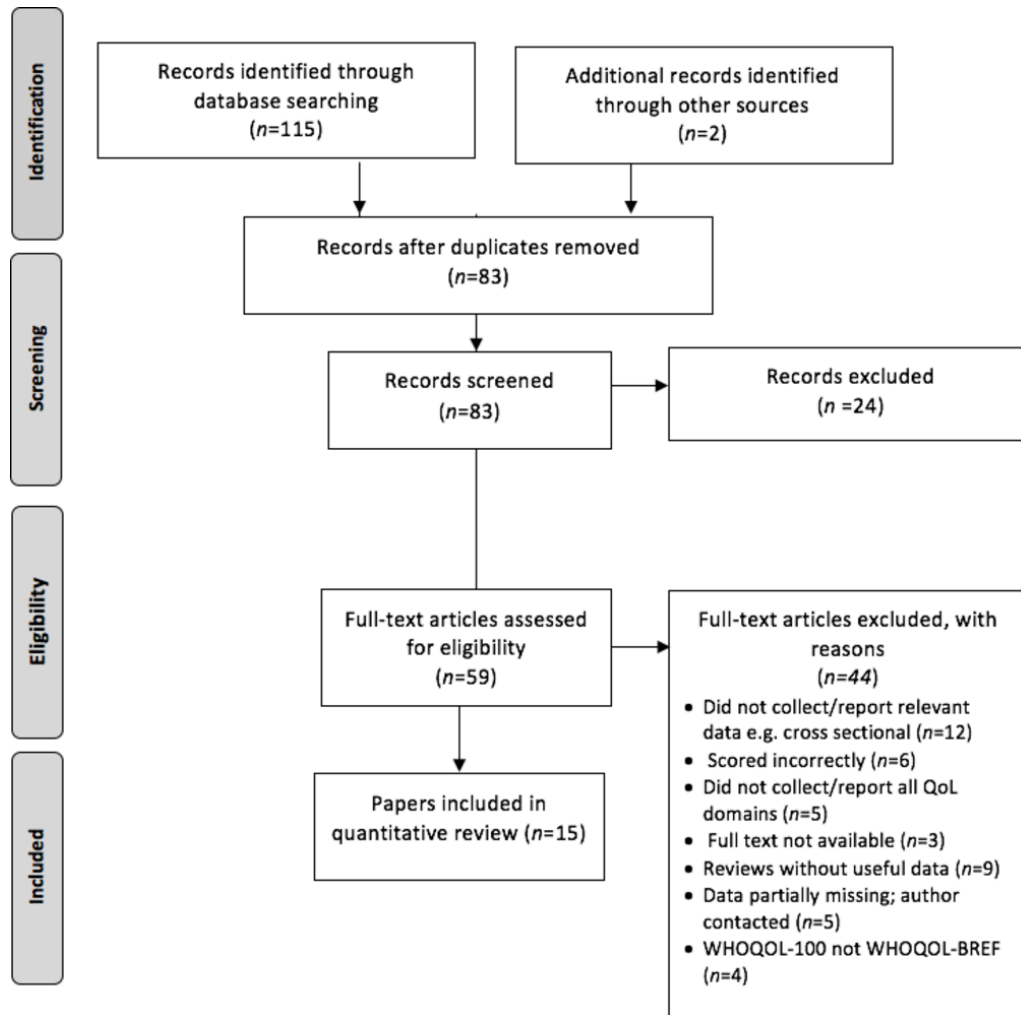
Figure 1 Flow of papers through the study. QoL, quality of life.

Characteristics of included studies
Details of the included studies are shown in table 1.

Sample characteristics
Reports on the WHOQOL-BREF responsiveness are presented historically in table 1, starting from the year 2000. Studies were conducted in a total of 11 cultures worldwide: Europe (n=15), East and South Asia (n=6), Australasia (n=2) and South America (n=1). Samples included a range of illnesses and disabilities categorised as physical health (n=17), mental health (n=3) and well (n=4). They included inpatients, outpatients, primary care, preventative care, social care and the community. The mean sample age was 50.26 years (SD=13.55), and included 60.46% women (SD=23.37). Study sampling was quasi-experimental with allocation to naturalistic groups, randomised and controlled groups, 'convenience' samples and structured survey populations.

Study design
Studies varied in design: pre-intervention and postintervention data with no control (n=18); pre-event and postevent with no control (n=3); change after an event, compared with a control group without experience of the event (n=2); and pre-intervention and postintervention compared with a control (n=1). SACMOS[36] information categorised the type of responsiveness in each study, and is recorded in table 1 (right-hand column).

Study characteristics
The mean interval between collecting the baseline measures/intervention and follow-up, was 33.19 weeks (SD=51.15).

Responsiveness
Across 24 tests (n=2084) the WHOQOL-BREF showed small-to-moderate responsiveness in detecting differences over time or health status in all four domains: physical (d=0.37; CI 0.25 to 0.49), psychological (d=0.22; CI 0.14 to 0.30), social (d=0.10; CI 0.05 to 0.15) and environmental QoL (d=0.12; CI 0.06 to 0.18) (see table 2 and online supplementary figures S2-S5 for forest plots: Physical (figure 2); Psychological (figure 3); Social (figure 4); Environmental (figure 5) QoL. These results suggest that the WHOQOL-BREF is responsive across all the domains, and across a wide variety of interventions and events.

As there was significant heterogeneity for each domain (see table 2), moderation analyses were conducted. No

Table 1 Sample characteristics, study characteristics, study design and effect sizes

Study, sub-sample and location	Measurement instruments	Sample, sample size, mean age, gender and time between measurements	Responsiveness/ sensitivity (effect sizes)		Study design/ calculation of effect sizes
O'Carroll et al[41] Scotland	WHOQOL-BREF WHOQOL-100	Liver transplant patients (n=50; age 49.6 years; 62% female) vs waiting list controls (n=21; age 53.9 years; 66.7% females) Presurgery and 3 m follow-up	Phys: Psych: Soc: Env:	1.29 0.82 0.55 0.79	Responsiveness calculated from presurgery and postintervention means and SDs (3 m follow-up) for surgery and waiting list controls
Hwang et al[42] Taiwan	WHOQOL-BREF	Older community adults who had falls (n=24) vs no falls (n=190) Mean age and gender not reported Follow-up 3 m after initial assessment	Phys: Psych: Soc: Env:	0.06 0.37 0.38 0.11	Responsiveness calculated from mean change and SDs for fallers vs non-fallers at 3 m follow-up
Taylor et al[43] New Zealand	WHOQOL-BREF HAQ	Inpatients with rheumatoid arthritis (n=72; age 60.6 years; 71.6% female) Outpatients (n=142; age 60.7 years; 71.8% female) Follow-up 2 weeks after admission	Phys: Psych: Soc: Env:	1.06 0.61 0.32 0.39	Responsiveness effect size reported as means, SD and correlations from inpatient admission to 2-week follow-up
van de Willige et al[44] The Netherlands	WHOQOL-BREF EQ-5D TTO	Patients with chronic schizophrenia (age 36.0 years; 45% female): RCT treatment (n=37) vs treatment as usual (n=39) Tested after treatment, 9 and 18 m	Phys: Psych: Soc: Env:	0.29 0.24 0.00 0.00	Responsiveness calculated from baseline and 18 m means (after receiving treatment) and t value
Chiu et al[40] Taiwan	WHOQOL-BREF Glasgow Coma Scale	Patients with traumatic brain injury (n=199; age 45.4 years; 35.7% female) Unemployed who remained unemployed (n=42) vs became employed (n=10) Follow-up 6 m	Phys: Psych: Soc: Env:	0.22 0.25 0.09 0.19	Responsiveness calculated from change means and SDs for those employed vs unemployed at 6 m follow-up
Ackerman et al[45] Australia	WHOQOL-BREF AQOL WOMAC K10 MHAQ	Hip and knee replacement surgery candidates: waiting list (n=279; median age 68.0 years; 57% women) After surgery (n=74; median age 68 years; 44% female) Control: population norms (n=396) Follow-up 3 m	Phys: Psych: Soc: Env:	0.98 0.42 0.13 0.43	Responsiveness calculated from mean change (from baseline to 3 m) and P values Discriminatory validity calculated using hip and knee replacement patients 3 m follow-up vs population norms
Alsaker et al[38] Norway	WHOQOL-BREF SF-36	Abused women (n=87) living in a violent partnership, then living in shelters after leaving partner (n=22; age 40 years; 100% female) Follow-up 12 m after leaving	Phys: Psych: Soc: Env:	0.06 0.24 0.06 0.00	Responsiveness calculated from mean change from baseline to 12 m (after leaving violent partner) and P value

Continued

Table 1 Continued

Study, sub-sample and location	Measurement instruments	Sample, sample size, mean age, gender and time between measurements	Responsiveness/ sensitivity (effect sizes)		Study design/ calculation of effect sizes
Zhao et al[46] China (Hong Kong)	WHOQOL-BREF SF-36 MLHF ChQOL	Inpatients with congestive heart failure (n=32; age 66.6 years; 48.72% female). Follow-up 4 weeks after treatment	Phys: Psych: Soc: Env:	0.93 0.18 0.08 0.06	Responsiveness calculated from mean change (after 4 weeks of treatment) and t values
Thakar et al[47] India	WHOQOL-BREF SF-36 Nurick grade	Patients with cervical spondylotic myelopathy (n=70; age 51.9 years; 5.71% female) Comparison with normative scores Follow-up 12 to 18 m after surgery	Phys: Psych: Soc: Env:	0.41 0.41 0.04 0.41	Responsiveness calculated from mean change (after 18 m after surgery) and P value
Skevington and McCrate[23] overall UK	WHOQOL-BREF SF-36	Total sample (n=4452; age 44.5 years; 67.52% women) Received one of 13 interventions for diabetes, arthritis, irritable bowel, fatigue, pain, schizophrenia (n=1864) Well controls (n=2761) Follow-up between 4 weeks and 12 m			Discriminant validity calculated from means and SDs of sick and well people, at one time point only
Skevington and McCrate[23] depression UK	As above	Depressed patients receiving treatment as usual (n=38; age 77.84 years; 52.6% female) Follow-up 3 m	Phys: Psych: Soc: Env:	0.66 0.77 0.03 0.31	Responsiveness calculated from means, SDs and correlations between baseline and follow-up at 3 m
Skevington and McCrate[23] arthritis UK	As above	Patients with rheumatoid arthritis receiving treatment as usual (n=18; age 55.28 years; 50.0% female) Follow-up 12 m	Phys: Psych: Soc: Env:	0.82 0.36 0.24 0.12	Responsiveness calculated from means, SDs and correlations between baseline and follow-up at 12 m
Skevington and McCrate[23] bowel disorders UK	As above	Outpatients with irritable bowel syndrome/disorder and patients with chronic fatigue syndrome in primary care receiving treatment as usual (n=84; age 47.75 years; 77.4% female) Follow-up 6 m	Phys: Psych: Soc: Env:	0.03 0.03 0.02 0.05	Responsiveness calculated from means, SDs and correlations between baseline and follow-up at 6 m
Skevington and McCrate[23] diabetes UK	As above	Patients with diabetes receiving treatment as usual in primary care (n=162; age not reported; 36.6% female) Various follow-up times	Phys: Psych: Soc: Env:	0.09 0.06 0.16 0.00	Responsiveness calculated from means, SDs and correlations between baseline and follow
Skevington and McCrate[23] diabetes UK	As above	Patients with diabetes receiving treatment as usual (n=50; age not reported; 26.0% female) Various follow-up times	Phys: Psych: Soc: Env:	0.28 0.04 0.26 0.18	Responsiveness calculated from means, SDs and correlations between baseline and follow-up

Table 1 Continued

Study, sub-sample and location	Measurement instruments	Sample, sample size, mean age, gender and time between measurements	Responsiveness/ sensitivity (effect sizes)		Study design/ calculation of effect sizes
Skevington and McCrate[23] students UK	As above	Students (university) receiving negative emotional disclosure intervention (n=42; age 19.57 years; 83.3% female) Follow-up 4 weeks	Phys: Psych: Soc: Env:	0.06 0.18 0.12 0.12	Responsiveness calculated from means, SDs and correlations between baseline and follow-up at 4 weeks
Skevington and McCrate[23] bowel disorders UK	As above	Patients with irritable bowel syndrome joining self-help group intervention in community (n=72; age 47.99 years; 69.4% female) Follow-up 8 m	Phys: Psych: Soc: Env:	0.21 0.16 0.09 0.04	Responsiveness calculated from means, SDs and correlations between baseline and follow-up at 8 m
Skevington and McCrate[23] schizophrenia UK	As above	Patients with schizophrenia moving from psychiatric ward to community residential home (n=7; age 75.67 years; 57.1% female) Various follow-ups	Phys: Psych: Soc: Env:	0.11 0.06 0.05 0.25	Responsiveness calculated from means, SDs and correlations between baseline and follow-up
Skevington and McCrate[23] chronic pain UK	As above	Patients with chronic pain receiving aromatherapy and massage intervention (n=29; age 45.74 years; 86.2% female) Follow-up 4 weeks	Phys: Psych: Soc: Env:	0.25 0.10 0.12 0.14	Responsiveness calculated from means, SDs and correlations between baseline and follow-up at 4 weeks
Skevington and McCrate[23] orthopaedic surgery UK	As above	Hip and knee surgery (arthroplasty) patients (n=58; age 69.72 years; 69.0% female) Follow-up 3 m	Phys: Psych: Soc: Env:	0.55 0.54 0.07 0.07	Responsiveness calculated from means, SDs and correlations between baseline and follow-up at 3 m
Valenti et al[39] Italy	WHOQOL-BREF	Earthquake survivors (n=397; age 52.2 years; 38.64% female) Follow-up at 6, 12 and 18 m after event	Phys: Psych: Soc: Env:	0.002 0.0001 0.002 0.002	Responsiveness calculated from mean change (over 18 m) and P value
Sartorio et al[48] Italy	WHOQOL-BREF ORWELL97 TSD-OC EuroQoL	Inpatients with obese (n=249; age 47.0 years; 69.0% female) on multidisciplinary weight programme Follow-up at end of 3-week intervention	Phys: Psych: Soc: Env:	0.21 0.12 0.07 0.11	Responsiveness calculated from mean change (after 3-week weight reduction programme) and P value)
Oliveira et al[49] Brazil	WHOQOL-BREF SF-36 FACT-B+4	Patients with breast cancer (n=32; age 49.2 years; 100% female) Follow-up 1 m after surgery	Phys: Psych: Soc: Env:	0.53 0.02 0.12 0.00	Responsiveness from change 1 m after surgery, effect size reported and SE calculated from CI
Yeh et al[50] Taiwan	WHOQOL-BREF OHRQoL	Dental caries restoration treatment (n=126; age not reported; 20.7% female) Pretreatment and post-treatment: 2 weeks	Phys: Psych: Soc: Env:	0.17 0.10 0.16 0.06	Responsiveness from pre and post means (2 weeks after treatment) and P value

Continued

Table 1 Continued

Study, sub-sample and location	Measurement instruments	Sample, sample size, mean age, gender and time between measurements	Responsiveness/ sensitivity (effect sizes)		Study design/ calculation of effect sizes
Lin et al[51] Taiwan	WHOQOL-BREF PCOSQ	Outpatients with polycystic ovarian syndrome (n=50; age not reported; 100% female) Tested at diagnosis, then 1–2 m after drug treatment initiated	Phys: Psych: Soc: Env:	0.37 0.18 0.00 0.15	Responsiveness calculated from mean change (baseline to 1 m) after drug treatment and t value

.AQoL, Assessment of Quality of Life; ChQOL , Chinese Quality of Life Instrument; EQ-5D (EuroQoL), European Quality of Life Scale-five dimensions; Env WHOQOL-BREF, Environmental domain of quality of life; FACT-B+4, Functional Assessment of Cancer Therapy – Breast plus Arm Mobility; HAQ, Health Assessment Questionnaire; K10, Kessler Psychological Distress Scale; MHAQ ,Shortened version of HAQ; MLHF, Minnesota Living with Heart Failure; Nurick grade of cervical spondylotic myelopathy; OHRQoL ,Oral Health Related Quality of Life rating; ORWELL 97, Obesity-Related Wellbeing; PCOSQ, Polycystic Ovary Syndrome Health-Related Quality of Life Questionnaire (Chinese); Phys WHOQOL-BREF, Physical quality of life; SF-36, Psych WHOQOL-BREF, Psychological domain of quality of life; QOL, quality of life ; Short Form-36 Health Survey; Soc WHOQOL-BREF, Social relationships domain of quality of life; TSD-OC, Obesity-Related Disability test; TTO, Time Trade Off; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

differences in responsiveness were found due to the sample characteristics of age and gender, or for the study characteristic of the interval between baseline and follow-up assessments (see table 3). This suggests that the WHOQOL-BREF is equally responsive regardless of the age and gender of the population studied, or the time interval between the intervention/event and measurement.

DISCUSSION

A systematic review of the global literature and meta-analysis of results confirmed that all four WHOQOL-BREF domain scores change significantly over time in response to an intervention, treatment or major life event. Across 24 studies containing 2084 adults living in 11 diverse cultures we found small to moderate effects for the physical and psychological domains, and small effects for the social relationships and environmental QoL domains. Furthermore, WHOQOL-BREF domains were found to be responsive to change, regardless of the sample's mean age, its gender composition, and the elapsed time between baseline and follow-up assessments. These findings demonstrate that WHOQOL-BREF scores are responsive to change across wide cross-cultural variations in local healthcare, diagnostic criteria and treatment delivery.

We also aimed to assess whether changes in WHOQOL-BREF domains were relevant and meaningful in clinical and social settings. Highly significant effect sizes found in each domain demonstrated that when scores change, these changes are relevant. More importantly, relevant results were found even for the social and environmental QoL domains where the changes had been small. In clinical terms, this means that WHOQOL-BREF domains have the capacity to detect even small changes induced by treatments or events. These positive findings indicate that the instrument shows good performance in measurement terms, and additionally confirms that the WHOQOL-BREF possesses an 'essential aspect of validity' (Norman GR, p816)[26] among the many other properties previously tested. However, these score changes should also be interpreted in conjunction with subjective perceptions of change during treatment, care or events. This second piece of perceptual information is important when managing the WB of individual patients or clients. Normative country data can also assist with interpretation. More recent country data have been collected by some national WHOQOL centres since 2004,[15] and this reference data should be requested from the WHOQOL country centre concerned, when seeking permission to use a particular language version.

Table 2 Effect sizes for the four WHOQOL domains

	k	N	d	CI	Q	I²
Physical	24	2084	0.37	0.25 to 0.49	153.81*	85.0
Psychological	24	2084	0.22	0.14 to 0.30	66.39*	65.4
Social	24	2084	0.10	0.05 to 0.15	35.85**	35.8
Environment	24	2084	0.12	0.06 to 0.18	52.76*	56.4

* P<0.001, **P <0.01.
d, Cohen's d effect size; I², extent of heterogeneity; K, number of studies; n, number of participants; Q, significance of heterogeneity; QOL, quality of life.

Table 3 Potential moderating effects of the responsiveness of the WHOQOL-BREF

	k	Physical					Psychological					Social					Environment				
		β	SE	CI	I²	Adj R²	β	SE	CI	I²	Adj R²	β	SE	CI	I²	Adj R²	β	SE	CI	I²	Adj R²
Age	18	0.01	0.01	−0.00 to 0.03	83.57	23.42	0.01	0.00	−0.00 to 0.02	69.91	11.09	.00	0.00	−0.00 to 0.01	1.05	52.00	0.01	0.00	−0.00 to 0.01	43.92	28.37
Gender	23	−0.00	0.00	−0.01 to 0.01	86.07	−6.39	−0.00	0.00	−0.00 to 0.00	66.85	−8.13	−0.00	0.00	−0.00 to 0.00	35.65	−10.90	0.00	0.00	−0.00 to 0.00	56.13	−7.64
Interval	21	−0.00	0.00	−0.01 to 0.00	84.26	10.69	−0.00	0.00	−0.00 to 0.00	64.74	7.56	−0.00	0.00	−0.00 to 0.00	12.56	33.84	−0.00	0.00	−0.00 to 0.00	56.87	6.78

Adj R², percentage of heterogeneity explained by covariate; β, regression coefficient (increase in effect size per unit increase in the covariate); I², extent of heterogeneity; K, number of studies; QOL, quality of life.

A predominance of physical health conditions among the meta-analysis studies may explain the strongest results for the physical domain. Several mental health studies contributed fairly strong responsiveness evidence to the psychological domain, but psychological involvement connected with chronic physical conditions adds weight to this. As very few studies directly investigated social relationships (eg, women with violent partners),[38] and environmental interventions (eg, rehousing schizophrenic patients)[23] or events (eg, earthquake survival),[39] the apparent weakness of these domains may be due to limited information about them. New research should investigate selected settings where a priori, these two domains of QoL might be expected to respond significantly (positively or negatively) to these types of events. Although cultural variations exist, events that impact on social QoL could include divorce, marriage and bereavement. For refugees, internally displaced people, homeless, indigenous and jobless populations,[40] environmental QoL could be salient. A further implication of discovering that this environmental QoL domain is responsive, is that it can now be used to assess environmental WB in many SDGs both related to health and beyond (eg, SDG 9). Public health targets among the SDGs that require SWB assessment will benefit from access to environmental QoL information, and be reassured that it performs equally well as other domains, and across key sociodemographic features.

Small sample sizes limited many studies included in the present SR, so data from large international surveys would improve conclusions. Nevertheless, this heterogeneous, cross-cultural data strengthen global conclusions about generalisability. Although effect size calculations were limited to Cohen's d, future studies should compare WHOQOL-BREF results across the range of responsiveness indices.

Some potential applications of the WHOQOL-BREF to the SDGs include measuring the impact on QoL of living with HIV, tuberculosis (TB) or malaria, QoL costs to families from maternal mortality, assessing whether providing universal healthcare improves QoL, improvements to QoL in women from greater access to reliable contraception; the cost to parent's QoL when a child dies from a disease that is preventable by vaccination (eg, measles), and impact on family QoL of a stillbirth or a baby with spina bifida. For other non-health SDGs, empirical evidence of QoL changes potentially could evaluate population well-being after a peace treaty (eg, Syria), when social justice is restored (eg, Turkey), if food security is delivered (eg, Ethiopia), after sustainable transport is established (eg, Bangladesh), and as household poverty is reduced (eg, Bolivia).

Agreeing a set of common international items to evaluate all SDG targets remained a major challenge in April 2017, when the list of 230 indicators was disclosed.[7] Not one indicator lists SWB or QoL assessment, despite earlier inclusion.[5] Due to its generic nature, the QoL dimensions of the WHOQOL could monitor any SDG where people's

QoL perceptions need assessment. Instead of offering a standard core of indicators to be applied by every country and then adding a few culturally appropriate indicators to satisfy local needs, countries are now invited to select from the UN list. Consequently, idiosyncratic sets could be chosen by around 200 countries to address the same target. The resulting data will confound cross-national comparisons in 2030, as common indicators drawn from a standard international core will not be available. Nevertheless, the IAEG[7] has built in some flexibility of cultural adaptation; a strategy that will satisfy countries who feel constrained by international demands for uniformity. This IAEG process has synergy with WHOQOL procedures for generating 'national items' on important local issues that are assessed with the core WHOQOL-BREF measure; such adaptability is yet another feature to recommend its use in the SDGs.

A spectrum of affordable technological devices and internet communication networks now enables subjective QoL information to be collected remotely in many countries. Mobiles, computers and 'smart' phones are almost ubiquitous, but cheaper new devices will replace them by 2030. National participation in the SDGs depends on having local expertise in computing, electronics and statistics[2]; administration methods must necessarily be culturally acceptable and affordable in GDP terms, and may need additional resources from external development investors.[2] Strategies developed in culturally appropriate ways can support vulnerable groups who might otherwise be overlooked. Lack of 'know how' may be most acute in low-income countries and minority communities, and paradoxically, it is these sectors that have the most pressing need to bring information about their QoL to the attention of global policymakers. Ensuring that this situation is avoided will be key to SDG aims.

CONCLUSION

A 20-year track record of advanced cross-cultural research on the WHOQOL-BREF has resulted in a short, state-of-the-art measure with high-quality measurement performance. We show how WHOQOL-BREF domains demonstrate responsiveness to clinical and social change that is relevant and meaningful, and consolidates knowledge about its validity. As this versatile tool can assess adult QoL in many cultures and conditions, it is highly suitable for use in the many different contexts addressed by SDG targets, where knowing about changes to WB will be essential to comprehensive evaluation.

REFERENCES

1. United Nations. *Millennium development goals report: time for global action for people and planet*. New York: United Nations Publications, 2015.
2. Nations U. 2015. Transforming our World: the 2030 agenda for sustainable development. Resolution adopted by the 70th Session of the UN General Assembly. New York: 12/35. Report No. A /RES/70/1.
3. ICUN, UNEP, WWF. *World Conservation Strategy: living resource conservation for Sustainable Development*. Gland Switzerland: ICUN, UNEP, WWF, 1980.
4. Sachs JD, McArthur JW. The Millennium Project: a plan for meeting the millennium development goals. *Lancet* 2005;365:347–53.
5. Sustainable Development Solutions Network. 2015. A global initiative for the United Nations. Framework for the Sustainable Development Goals: landing a data revolution for the SDGs. Reportto the Secretary-General of the United Nations by the Leadership Council of the Sustainable Development Network, June 12 th, 2015; Annex 1: Beyond GDP – new measures for development;Wellbeing a cross-cutting issue: 69.
6. Skevington SM, Sartorius N, Amir M. Developing methods for assessing quality of life in different cultural settings. The history of the WHOQOL instruments. *Soc Psychiatry Psychiatr Epidemiol* 2004;39:1–8.
7. United Nations. 2017. Report of the Inter-Agency and Expert Group on the Sustainable Development Goal (IAEG-SDG) indicators. Final list of proposed SDG indicators (Annex IV) : United Nations Statistics Division. Report No. E/CN.3/2016/Rev1.
8. Diener E, Suh EM, Lucas RE, *et al*. Subjective well-being: three decades of progress. *Psychol Bull* 1999;125:276–302.
9. Camfield L, Skevington SM. On subjective well-being and quality of life. *J Health Psychol* 2008;13:764–75.
10. Skevington SM. Boehnke JM and the WHOQOL SRPB Group. How is subjective wellbeing related to quality of life? Do we need two constructs and both measures?. In Press.
11. Prescott-Allen R. The Wellbeing of Nations: a country by country index of quality of life and the environment, Island Press, Washington DC. *USA* 2001.
12. Easterlin RA, Sawangfa O. Happiness and economic growth: does the cross-section predict time trends? Evidence from developing countries. In: Diener E, Helliwell JF, Kahneman D, eds. *International Differences in Wellbeing* Oxford, UK Oxford University Press, 2010:166–216.
13. Bowden A, Fox-Rushby JA. A systematic and critical review of the process of translation and adaptation of generic health-related quality of life measures in Africa, Asia, Eastern Europe, the Middle East, South America. *Soc Sci Med* 2003;57:1289–306.
14. Skevington SM, Bradshaw J, Saxena S. Selecting national items for the WHOQOL: conceptual and psychometric considerations. *Soc Sci Med* 1999;48:473–87.
15. Skevington SM, Lotfy M, O'Connell KA. WHOQOL Group. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. *Qual Life Res* 2004;13:299–310.

16. The WHOQOL Group. Development of the WHOQOL-BREF quality of life assessment. *Psychol Med* 1998;28:551–8.

17. Llewellyn AM, Skevington SM. Using guided individualised feedback to review self-reported quality of life in health and its importance. *Psychol Health* 2015;30:301–17.

18. Llewellyn AM, Skevington SM. Evaluating a new methodology for providing individualized feedback in healthcare on quality of life and its importance, using the WHOQOL-BREF in a community population. *Qual Life Res* 2016;25:605–14.

19. The WHOQOL Group. The development of the World Health Organisation quality of life assessment instrument (The WHOQOL). In: Orley J, Kuyken W, eds. *Quality of Life Assessment: International Perspectives. Springer-Verlag, Berlin*, 1994:41–60.

20. Den Oudsten BL, Van Heck GL, De Vries J. The suitability of patient-based measures in the field of Parkinson's disease: a systematic review. *Mov Disord* 2007;22:1390–401.

21. Department of Mental Health and Addiction Services (DMHAS). *Annual report of Consumer Satisfaction Survey*. Connecticut (US): DMHAS, 2015.

22. Hawthorne G, Herrman H, Murphy B. Interpreting the WHOQOL-BREF: preliminary population norms and effect sizes. *Soc Indic Res* 2006.

23. Skevington SM, McCrate FM. Expecting a good quality of life in health: assessing people with diverse diseases and conditions using the WHOQOL-BREF. *Health Expect* 2012;15:49–62.

24. Diehr PH, Derleth AM, McKenna SP, *et al*. Synchrony of change in depressive symptoms, health status, and quality of life in persons with clinical depression. *Health Qual Life Outcomes* 2006;4:27–37.

25. Wyrwich KW, Bullinger M, Aaronson N, *et al*. Estimating clinically significant differences in quality of life outcomes. *Qual Life Res* 2005;14:285–95.

26. Norman GR, Wyrwich KW, Patrick DL. The mathematical relationship among different forms of responsiveness coefficients. *Qual Life Res* 2007;16:815–22.

27. Cohen J. *Statistical power for the behavioral sciences*. 2nd edn. New York: Academic Press, 1988.

28. Yen PN, Yang CC, Chang WP, *et al*. Late effects on the health-related quality of life in a cohort population decades after environmental radiation exposure. *Int J Radiat Biol* 2013;89:639–44.

29. Ardalan A, Mazaheri M, Vanrooyen M, *et al*. Post-disaster quality of life among older survivors five years after the Bam earthquake: implications for recovery policy. *Ageing Soc* 2011;31:179–96.

30. Guo HX. Chen H Teresa BK Chen Q, Au ML. Factors influencing the quality of life of elderly living in a prefabricated housing complex in the Sichuan earthquake area. *J Nurs* 2012;59:61–71.

31. Hammoudeh W, Hogan D, Giacaman R. Quality of life, human insecurity, and distress among Palestinians in the Gaza Strip before and after the Winter 2008-2009 Israeli war. *Qual Life Res* 2013;22:2371–9.

32. Akinyemi OO, Owoaje ET, Ige OK, *et al*. Comparative study of mental health and quality of life in long-term refugees and host populations in Oru-Ijebu, Southwest Nigeria. *BMC Res Notes* 2012;5:394.

33. Pabilonia W, Combs SP, Cook PF. Knowledge and quality of life in female torture survivors. *Torture* 2010;20:4–22.

34. Simonelli G, Leanza Y, Boilard A, *et al*. Sleep and quality of life in urban poverty: the effect of a slum housing upgrading program. *Sleep* 2013;36:1669–76.

35. Skevington SM. Conceptualising dimensions of quality of life in poverty. *J Community Appl Soc Psychol* 2009;19:33–50.

36. Aaronson N, Alonso J, Burnam A, *et al*. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res* 2002;11:193–2002.

37. StataCorp. *Stata statistical software: release 11*. College Station, Texas, 2009.

38. Alsaker K, Moen BE, Kristoffersen K. Health-related quality of life among abused women one year after leaving a violent partner. *Soc Indic Res* 2008;86:497–509.

39. Valenti M, Masedu F, Mazza M, *et al*. A longitudinal study of quality of life of earthquake survivors in L'Aquila, Italy. *BMC Public Health* 2013;13: 1143-–1150.

40. Chiu WT, Huang SJ, Hwang HF, *et al*. Use of the WHOQOL-BREF for evaluating persons with traumatic brain injury. *J Neurotrauma* 2006;23:1609–20.

41. O'Carroll RE, Smith K, Couston M, *et al*. A comparison of the WHOQOL-100 and the WHOQOL-BREF in detecting change in quality of life following liver transplantation. *Qual Life Res* 2000;9:121–4.

42. Hwang HF, Liang WM, Chiu YN, *et al*. Suitability of the WHOQOL-BREF for community-dwelling older people in Taiwan. *Age Ageing* 2003;32:593–600.

43. Taylor WJ, Myers J, Simpson RT, *et al*. Quality of life of people with rheumatoid arthritis as measured by the World Health Organization Quality of Life Instrument, short form (WHOQOL-BREF): score distributions and psychometric properties. *Arthritis Rheum* 2004;51:350–7.

44. van de Willige G, Wiersma D, Nienhuis FJ, *et al*. Changes in quality of life in chronic psychiatric patients: a comparison between EuroQol (EQ-5D) and WHOQoL. *Qual Life Res* 2005;14:441–51.

45. Ackerman IN, Graves SE, Bennell KL, *et al*. Evaluating quality of life in hip and knee replacement: psychometric properties of the World Health Organization quality of life short version instrument. *Arthritis Rheum* 2006;55:583–90.

46. Zhao L, Leung KF, Liu FB, *et al*. Responsiveness of the Chinese quality of life instrument in patients with congestive heart failure. *Chin J Integr Med* 2008;14:173–9.

47. Thakar S, Christopher S, Rajshekhar V. Quality of life assessment after central corpectomy for cervical spondylotic myelopathy: comparative evaluation of the 36-Item Short Form Health Survey and the World Health Organization Quality of Life-Bref. *J Neurosurg Spine* 2009;11:402–12.

48. Sartorio A, Agosti F, De Col A, *et al*. Concurrent comparison of the measurement properties of generic and disease-specific questionnaires in obese inpatients. *J Endocrinol Invest* 2014;37:31–42.

49. Oliveira IS, Costa LC, Manzoni AC, *et al*. Assessment of the measurement properties of quality of life questionnaires in Brazilian women with breast cancer. *Braz J Phys Ther* 2014;18:372–83.

50. Yeh DY, Kuo HC, Yang YH, *et al*. The responsiveness of patients quality of life to dental caries treatment-A prospective study. *PLoS One* 2016;11:e0164707.

51. Lin CY, Ou HT, Wu MH, *et al*. Validation of Chinese Version of Polycystic Ovary Syndrome Health-Related Quality of Life Questionnaire (Chi-PCOSQ). *PLoS One* 2016;11:e0154343.