

RESEARCH ARTICLE

Patient-Reported Outcomes for Quality of Life Assessment in Atrial Fibrillation: A Systematic Review of Measurement Properties

Dipak Kotecha^{1,2,4*}, Amar Ahmed¹, Melanie Calvert³, Mauro Lencioni⁴, Caroline B. Terwee⁵, Deirdre A. Lane^{1,2}

1 Institute of Cardiovascular Sciences, University of Birmingham, Birmingham, United Kingdom, **2** City Hospital, Sandwell and West Birmingham Hospitals NHS Trust, Birmingham, United Kingdom, **3** Institute of Applied Health Research, University of Birmingham, Birmingham, United Kingdom, **4** University Hospitals Birmingham NHS Trust, Birmingham, United Kingdom, **5** Department of Epidemiology and Biostatistics and the EMGO Institute of Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands

* d.kotecha@bham.ac.uk



OPEN ACCESS

Citation: Kotecha D, Ahmed A, Calvert M, Lencioni M, Terwee CB, Lane DA (2016) Patient-Reported Outcomes for Quality of Life Assessment in Atrial Fibrillation: A Systematic Review of Measurement Properties. PLoS ONE 11(11): e0165790. doi:10.1371/journal.pone.0165790

Editor: Terence J Quinn, University of Glasgow, UNITED KINGDOM

Received: May 26, 2016

Accepted: October 4, 2016

Published: November 1, 2016

Copyright: © 2016 Kotecha et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper, its supporting information files, and the original articles which comprise the systematic review.

Funding: AA was supported by the University of Birmingham and funded by the EW Kite Bequest. DK is funded by the National Institute for Health Research (NIHR) as part of the RAte control Therapy Evaluation in Atrial Fibrillation (RATE-AF) trial (www.birmingham.ac.uk/rate-af) and a Career Development Fellowship (CDF-2015-08-074). The opinions expressed are those of the authors and do

Abstract

Background

Atrial fibrillation is a large and growing burden across all types of healthcare. Both incidence and prevalence are expected to double in the next 20 years, with huge impact on hospital admissions, costs and patient quality of life. Patient wellbeing determines the management strategy for atrial fibrillation, including the use of rhythm control therapy and the clinical success of heart rate control. Hence, evaluation of quality of life is an emerging and important part of the assessment of patients with atrial fibrillation. Although a number of questionnaires to assess quality of life in atrial fibrillation are available, a comprehensive overview of their measurement properties is lacking.

Methods and Results

We performed a systematic review of the measurement properties of atrial fibrillation-specific health-related quality of life questionnaires. Methodological quality was assessed using the Consensus based Standards for selection of health Measurement Instruments (COSMIN) checklist, with measurement properties rated for quality against optimal criteria and levels of evidence. We screened 2,216 articles, of which eight articles describing five questionnaires were eligible for inclusion: Atrial Fibrillation 6 (AF6), Atrial Fibrillation Effect on QualiTY-of-Life (AFEQT), Atrial Fibrillation Quality of Life Questionnaire (AFQLQ), Atrial Fibrillation Quality of Life (AFQoL), and Quality of Life in Atrial Fibrillation (QLAF). Good reliability (internal consistency and test-retest reliability) was demonstrated for AF6, AFEQT, AFQLQ and AFQoL. Content, construct and criterion validity were positively rated only in AFEQT. Responsiveness was positively rated only in AFEQT, but with limited

not represent the NIHR or the UK Department of Health. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

Competing Interests: All authors have completed the ICMJE uniform disclosure form (www.icmje.org/coi_disclosure.pdf) and declare the following, all outside of the submitted work: DK is the lead for the Beta-blockers in Heart Failure Collaborative Group (BB-meta-HF) and the RATE-AF trial, and has received honoraria and research grants from Menarini, professional development support from Daiichi-Sankyo and speaker fees from AtriCure; AA none; MC has received personal fees from Astellas Pharma outside the submitted work, and has received travel awards from the European Society of Cardiology and the Atrial Fibrillation NETWORK (AFNET) to attend conferences presenting on the use of PROMs in cardiovascular research and clinical practice; ML has received travel grants from Pfizer, Bayer and Biosense Webster and has been on the speaker bureau for Bayer and Pfizer; CT has not received any funding but is one of the founders of the COSMIN initiative; DAL has received investigator-initiated educational grants from Bayer Healthcare, Bristol Myers Squibb (BMS) and Boehringer Ingelheim, has been on the speaker bureau for Boehringer Ingelheim, Bayer, and BMS/Pfizer and is a consultant for BMS and Boehringer Ingelheim.

evidence. Overall, AFEQT showed strong positive evidence for 2 of 9 measurement properties, compared to one for AFQoL and none for the remaining questionnaires.

Interpretation

Given the low ratings for many measurement properties, no single questionnaire can be recommended, although AFEQT performed strongest. Further studies to robustly assess reliability, validity and responsiveness of AF-specific quality of life questionnaires are required. This review consolidates the current evidence for quality of life assessment in patients with atrial fibrillation and identifies priority areas for future research.

Introduction

Atrial fibrillation (AF) is the most common cardiac arrhythmia and is predicted to double in prevalence over the next 20 years. It is not only associated with adverse prognosis, but accounts for considerable healthcare expenditure. [1] Numerous studies have identified a substantial reduction in health-related quality of life (QoL) in AF populations. [2] Patients with AF experience symptoms such as palpitations, exercise intolerance, dizziness and dyspnea, which frequently limit capacity to undertake daily activities. [3, 4]

Apart from anticoagulation to prevent strokes, current management of AF is focused on reducing symptoms and improving QoL. [5] Even well-established therapies in AF have proven to have little impact on death and other adverse outcomes, including rate and rhythm control therapy. [6–9] Although the fundamental reason for pursuing such therapies is to improve QoL, there is limited appreciation of how to measure QoL in clinical practice.

A variety of both generic and AF-specific questionnaires are available to assess patient-reported QoL. [10] Generic patient reported outcome measures (PROMs), such as the SF-36 and EQ-5D questionnaires, have the advantage of allowing comparison of QoL in patients with different diseases, but are less sensitive to the effects of a single disease on QoL. [11] With the growing burden of AF, the relatively high cost of treatment, and the appreciation of QoL as a treatment objective, there has been substantial interest in the development of AF-specific PROMs for use in both clinical research and routine practice. However, these have yet to enter routine clinical practice, in part due to concerns about their validity and responsiveness. [12]

Our objective was to perform a systematic review of the measurement properties of disease-specific PROMs measuring QoL in AF, assessing studies which designed and validated the questionnaires. We aim to provide clinicians with an understanding of whether these instruments would likely be of value in research and clinical practice. We utilized a comprehensive methodology from the Consensus based Standards for selection of health Measurement Instruments (COSMIN) group [13], which includes rigorous assessment of validity, reliability, and responsiveness of QoL questionnaires.

Methods

Eligibility criteria and search strategy

We included all studies that examined at least one or more measurement properties of a disease-specific PROM measuring QoL in patients with AF (aged 18 years and older). We permitted inclusion of studies with interviewer-based administration, as long as questionnaires were self-reported by patients. Studies were required to be published as a full-text article. Articles

were excluded if the article failed to specify the type of tool utilized. There was no restriction on language, date of publication or type of AF.

A systematic search was performed in Medline, EMBASE, PsycINFO and CINAHL from inception until the 15th February 2015 and in the Cochrane library from 1993 to 9th March 2015. The search strategy (see [S1 Appendix](#)) consisted of 3 filters composed of search terms for the following: (1) AF; (2) PROMs; and (3) measurement properties. The latter two filters were developed at the University of Oxford and VU University Medical Center [14] respectively (available at www.cosmin.nl). All filters were adapted for each database. Reference lists of the included studies were also manually searched in addition to articles that cited them.

This review was prospectively registered with the PROSPERO database of systematic reviews (http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015016600). In this paper, we report on our primary outcome which addresses the quality of AF-specific PROMs using the COSMIN methodology.

Data collection and synthesis

For each study, two investigators independently extracted and tabulated data on a standardized data extraction form. Discrepancies and missing data were resolved by group discussion, reference to the original publication and additional adjudication. Two studies published in Japanese [15, 16] were first translated by a Cardiologist fluent in both the Japanese and English languages.

The COSMIN checklist was used by two independent reviewers to assess the methodological quality of the included studies. [13] This tool distinguishes three main domains (reliability, validity and responsiveness), subdivided into nine measurement properties. For each measurement property, numerous standards are included in the form of questions to determine the quality of the study. Each item per measurement property is scored as excellent, fair, good or poor, with the lowest rating determining the overall quality score for that study (“worst score counts” method).

Measurement properties were rated for quality against criteria for optimal measurement properties. Each result was rated as either positive (+), negative (-) or indeterminate (?), providing that methodological quality was not poor. [17] [Table 1](#) summarizes the definitions of the nine measurement properties assessed and the preferred methods for their assessment, based on the COSMIN taxonomy, with detailed criteria presented in [S1 Table](#). [18, 19] Best evidence synthesis was performed by applying levels of evidence developed by Terwee *et al.* (www.cosmin.nl), combining the ratings of study quality with the strength of findings to determine an overall judgement for each questionnaire, as demonstrated in [S2 Table](#).

Results

Study characteristics

The results of the search strategy are outlined in [Fig 1](#). After removal of duplicates, the literature search yielded 2,216 articles of which eight were included in the final analysis. [15, 16, 20–25] A summary of studies excluded at the full-text stage is presented in [S3 Table](#).

We identified five questionnaires developed and validated for the assessment of quality of life in AF: Atrial Fibrillation 6 (AF6) [24, 25], Atrial Fibrillation Effect on Quality-of-Life (AFEQT) [20], Atrial Fibrillation Quality of Life Questionnaire (AFQLQ) [15, 16], Atrial Fibrillation Quality of Life (AFQoL) [21, 22], and Quality of Life in Atrial Fibrillation (QLAF). [23]

The characteristics of the questionnaires included in this review and the study populations in which their measurement properties were evaluated are summarized in [Table 2](#). The QoL

Table 1. Definitions of measurement properties.

Measurement Property	Definition	Quality criteria for positive rating *
1. Reliability		
Internal Consistency	The extent to which items of a questionnaire are interrelated in their underlying construct	Cronbach's alpha (+ if ≥ 0.70)
Test-retest reliability	The proportion of the total variance in the measurements which is due to true differences between patients.	Intra-class correlation coefficient or weighted Kappa (+ if ≥ 0.7 or Pearson's $r \geq 0.8$)
Measurement Error	The extent of change not attributable to the true alteration in the patient's quality of life	Minimal important change (MIC) (+ if MIC > smallest detectable change or MIC outside the limits of agreement)
2. Validity		
Content Validity	Assesses whether the questionnaire adequately reflects the construct of interest	Relevance to target population (+ if relevant to the construct measured and comprehensive)
Construct Validity		
• <i>Structural Validity</i>	Consistency between factor structure and the underlying construct	Factor analysis (+ if factors explain at least 50% of variance)
• <i>Hypothesis testing</i>	<ul style="list-style-type: none"> • Convergent: Determines whether expected similar domains between measurement tools are in fact similar • Discriminant: Determines whether expected dissimilar domains between measurement tools are in fact unrelated 	Correlation coefficient (+ if ≥ 0.5 with an instrument measuring the same construct, or $\geq 75\%$ in accordance with hypotheses and concordance with stated constructs)
• <i>Cross-cultural validity</i>	Assesses whether a translated questionnaire adequately reflects the original questionnaire	Comparison of different language versions (+ if original factor structure confirmed or no important differences)
Criterion Validity	Similarity between scores derived through the questionnaire against those from a gold standard	Correlation coefficient with an accepted gold standard measure (+ if ≥ 0.7)
3. Responsiveness		
Response to clinical change	The ability to detect change in quality of life over time	Correlation and hypothesis concordance (+ if correlation ≥ 0.50 with instrument measuring same construct or $\geq 75\%$ in accordance with hypotheses or area under the curve ≥ 0.70)

* For a full description of the criteria for measurement properties, see [S1 Table](#).

doi:10.1371/journal.pone.0165790.t001

questionnaires were assessed in a variety of geographical locations, however patients were of a similar mean age (62–67 years). Women were under-represented in all of the cohorts (20–43%). The type of AF varied in the individual studies, although on average there was an even split between paroxysmal and more persistent forms.

QoL instruments

Results of the assessment of validity for each instrument are detailed in [S4 Table](#) and for reliability and responsiveness in [S5 Table](#). It was not possible to rate any of the instruments in relation to measurement error, as statistical evaluations were not reported. Similarly, there were no data on cross cultural validity, and hence study quality and instrument ratings were not performed for these domains. A summary of findings is presented in [Fig 2](#). Best evidence synthesis to provide an overall appraisal of each questionnaire is detailed in [Table 3](#), limited by the availability of only one or two published articles for each questionnaire. Below, we concisely review the main results for the individual QoL instruments alphabetically.

1. AF6. AF6 was evaluated in 2 articles, both utilizing the same cohort of 111 patients recruited in Sweden. [[24](#), [25](#)]

Reliability: The quality of the study was fair for internal consistency and reliability, with both of these measurement properties rated positively. Internal consistency was only evaluated in

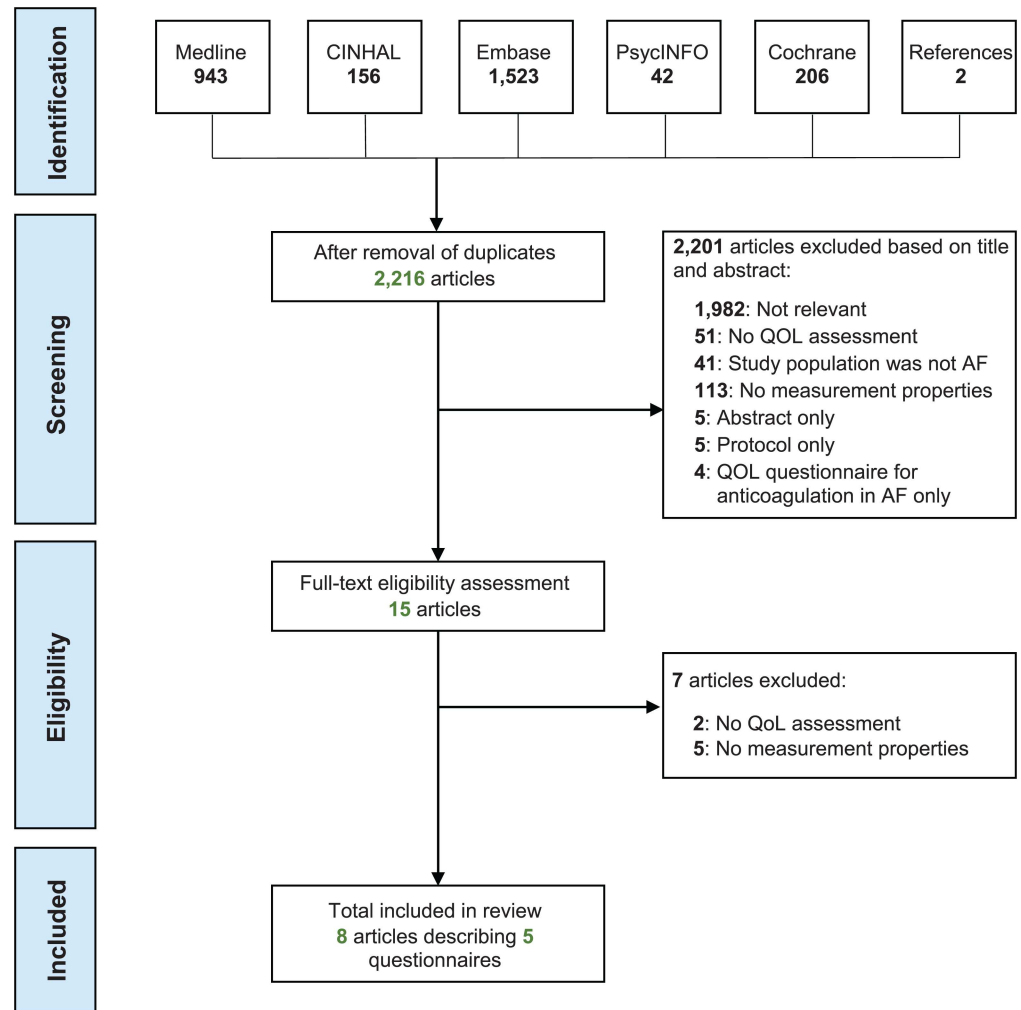


Fig 1. Selection of studies flowchart. Study inclusion flowchart. See [S1 Appendix](#) for search strategy.

doi:10.1371/journal.pone.0165790.g001

patients who underwent direct current cardioversion (which may not be representative of general AF patients and can variably affect QoL depending on success). Test-retest reliability was only assessed in 9 patients who had failed cardioversion, with 3 of the 6 items (limitations in daily life due to AF, discomfort due to AF and fatigue due to AF) associated with poor coefficient values.

Validity: Content validity and hypothesis testing of this instrument were rated as indeterminate due to poor ratings for reported methodology. For criterion validity, appropriate statistics were not reported. Structural validity assessment was deemed of fair quality overall, although statistical methods and handling of missing data were excellent and the instrument was positively rated for this measurement property.

Responsiveness: Only assessed in a single study of poor methodological quality, which did not describe the construct of the comparator (SF-36) or detail the handling of missing data or the underlying hypotheses. [25]

Synthesis: Two measurement properties was positively rated based on limited evidence.

2. AFEQT. AFEQT was evaluated in a single article composed of 213 patients recruited from hospitals across North America. [20] The population was predominantly of European descent and English-speaking.

Table 2. Characteristics of included studies.

Instrument	Population characteristics					Questionnaire characteristics		
	Geographical location	Sample size of studies	Mean age (years ± SD)	Women (%)	Type of AF (%)	Number of items	Domains	Response options
AF6[24, 25]	Sweden	111	67 ± 12	20%	Not reported	6	Dyspnea at rest, dyspnea on exertion, limitation in daily life due to AF, discomfort due to AF, fatigue due to AF, anxiety due to AF	10 point Likert scale
AFEQT[20]	Canada & US	213	62 ± 12	42%	Paroxysmal 66%, persistent 29%, permanent 5%	20	Symptoms, daily activities, treatment concerns, treatment satisfaction	7 point Likert scale
AFQLQ[15, 16]	Japan	40 & 172	64 ± 10	24%	Paroxysmal 57%, persistent 43%	26	Type and frequency of symptoms, severity of symptoms, psychological aspects, limitation in daily life	4–6 options of ranging severity dependent on domain
AFQoL[21, 22]	Spain	112 & 417	62 ± 12	35%	Paroxysmal 53%, permanent 47%	18	Psychological, physical, sexual activity	5 point Likert scale
QLAF[23]	Brazil	63	63 ± 12	43%	Paroxysmal 38%, persistent 32%, permanent 30%	22	Palpitations, chest pain, breathlessness, dizziness, drugs, direct current cardioversion, ablation	Letters assigned to text options and yes/no tick boxes

doi:10.1371/journal.pone.0165790.t002

Reliability: The quality of the study was excellent for internal consistency and fair for test-retest reliability, due to inadequate information regarding the environment of questionnaire administration. Both measurement properties were rated as positive.

Validity: The quality of the study on content validity was deemed methodologically excellent, whereas the quality of the study regarding structural validity, hypothesis testing and criterion validity were rated fair as the handling of missing data was not detailed. All of these validity aspects were rated positive.

Responsiveness: Although there was excellent description of hypotheses, comparators and methods, the overall quality of the study was rated fair, as responsiveness was only assessed in cohorts of patients who received an intervention (66 patients with medication adjustment and 76 with ablation). It was unclear how missing data were handled, and the expected or achieved correlations were not reported. The responsiveness of the instrument was however, rated as positive (data presented in S5 Table). Comparison with generic QoL measures (SF-36 and EQ-5D) resulted in smaller effect sizes, although the responsiveness of the AFEQT was comparable to the Toronto AF Symptoms Check List and Atrial Fibrillation Severity Scale.

Synthesis: Two measurement properties showed strong positive evidence, and four were positive based on limited evidence.

3. AFQLQ. AFQLQ was studied in two articles, both recruiting patients from Japan. [15, 16]

Reliability: Assessed in a single article, [16] quality was categorized as methodically fair for both internal consistency and test-retest reliability. Although the sample size was >100, both measurement properties were marked down due to inadequate information on missing data and potential differences in test conditions during resampling. Both measurement properties were rated positive.

Validity: Structural validity was investigated in a single article. [15] The study was methodologically fair as assessment was only performed in a small number of patients and the handling

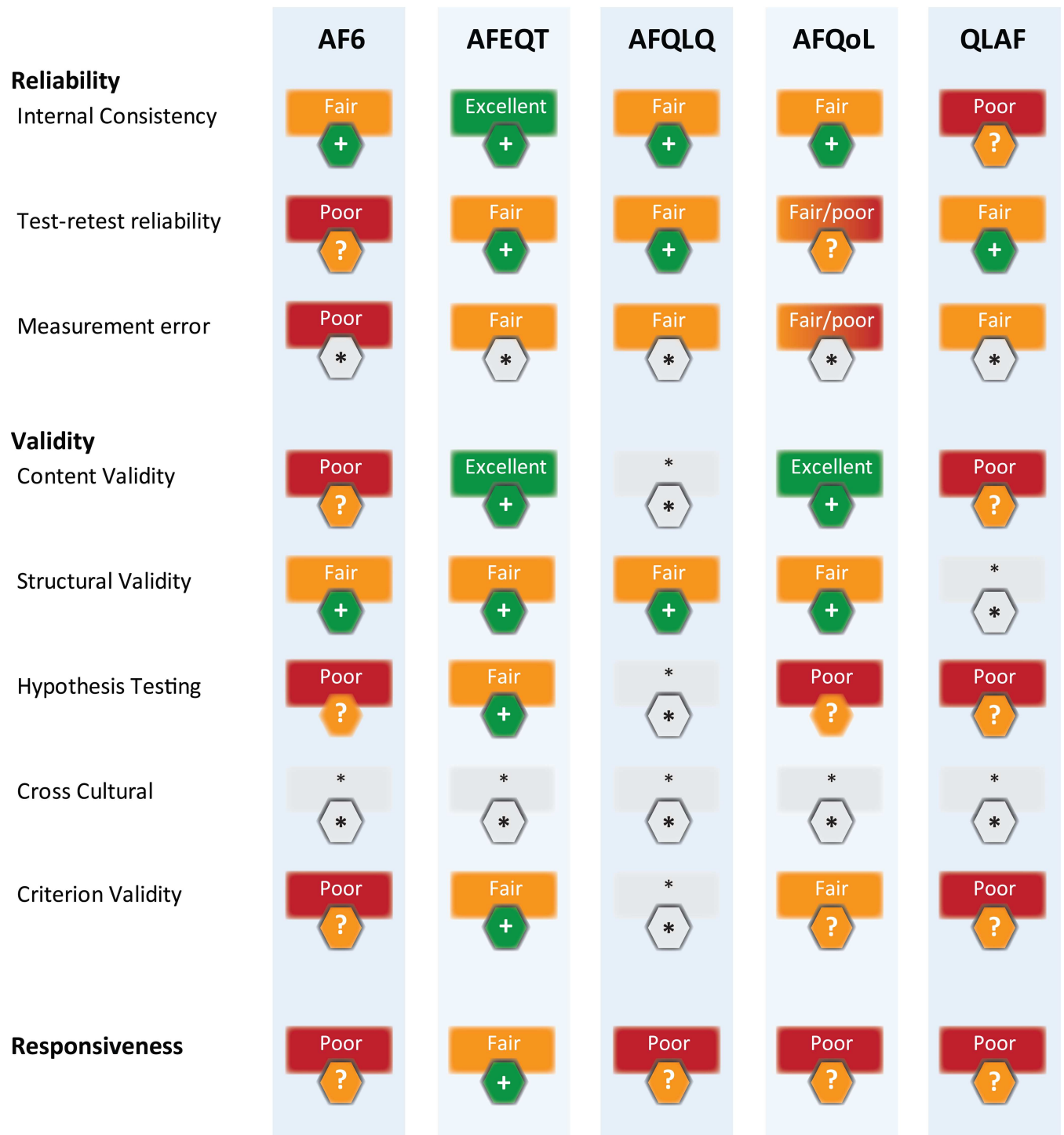


Fig 2. Summary of assessment. For each measurement property, the PROM is assessed for methodological quality of the study (excellent, good, fair or poor) and given an overall rating (positive [+], negative [-] or indeterminate/unknown [?]) for the results. A level of evidence was applied, combining the number and quality of the studies with the strength of findings. Studies that have poor quality are given an 'indeterminate/unknown' rating due to the limited level of evidence. For AFQoL, reliability was assessed differently in the two related studies. * Insufficient data were available to rate this criterion. See Table 1 for assessment criteria.

doi:10.1371/journal.pone.0165790.g002

Table 3. Synthesis of results.

Measurement property	AF6	AFEQT	AFQLQ	AFQoL	QLAF
Internal Consistency	+	+++	+	+	?
Test-retest reliability	?	+/-	+	+/-	+
Measurement error	n/r	n/r	n/r	n/r	n/r
Content Validity	?	+++	n/r	+++	?
Structural Validity	+	+	+	+	n/r
Hypothesis Testing	?	+	n/r	?	?
Cross Cultural	n/r	n/r	n/r	n/r	n/r
Criterion Validity	?	+	n/r	+/-	?
Responsiveness	?	+	?	?	?

“+++” or “- - -” = strong positive or negative evidence; “++” or “--” = moderate positive or negative evidence; “+” or “-” = limited positive or negative evidence; “+/-” = conflicting findings; “?” = unknown, due to poor methodological quality; n/r = not rated due to insufficient data. See [S2 Table](#) for assessment criteria.

doi:10.1371/journal.pone.0165790.t003

of missing data was not clearly reported. The instrument was positively rated for structural validity. No information was available on other validity aspects.

Responsiveness: Assessed in a single article [16], this study was rated as methodically poor. It was unclear if changes occurred in the patients during the interim period, either in treatment or symptoms, and the study failed to detail expected hypotheses.

Synthesis: Three positive measurement properties with limited evidence, however there was insufficient data to rate 5 out of 9 measurement properties.

4. AFQoL. AFQoL was assessed in two articles recruiting Spanish patients from hospitals across Madrid and Bilbao. [21, 22]

Reliability: An evaluation of internal consistency was rated as methodologically fair in both articles, whereas test-retest was rated fair in one [21] and poor in the other [22], due to the relatively small sample size, uncertainty if the test conditions were the same, and lack of details regarding the handling of missing data. The instrument was rated positive for internal consistency in both articles [21, 22], but indeterminate for test-retest reliability.

Validity: Content validity and structural validity were assessed in a single article. [22] The quality of content validity assessment was methodologically excellent, while structural validity was deemed fair due to omission of details on the handling of missing data. The instrument was rated positively for both of these measurement properties. The study on hypothesis testing was methodologically poor due to a lack of detail on the self-perceived health status scale employed, hence the instrument was rated indeterminate. Criterion validity was assessed in a single article [21], with fair study quality but an indeterminate rating as QoL was compared with patients post-myocardial infarction.

Responsiveness: Assessed in a single article of poor methodologically quality [21], as only correlations between baseline SF-36 and AFQoL were reported.

Synthesis: One measurement property showed strong positive evidence, and two were positive based on limited evidence.

5. QLAF. QLAF was assessed in a single article consisting of 63 patients recruited from centers across Brazil. [23]

Reliability: The assessment of internal consistency was rated as methodically poor as factor analysis was not performed. Study quality for test-retest reliability was fair, due to the small sample size and insufficient detail on the handling of missing items. The instrument was rated as positive for this measurement property.

Validity: Content validity was rated as methodologically poor, as the questionnaire was administered to a small target population and there was no patient involvement in the questionnaire

design. Prior hypotheses regarding expected correlations between the SF-36 and QLAF were not stated when undertaking hypothesis testing and appropriate statistics were not used during criterion validity assessment; thus both construct and criterion validity were rated as methodologically poor.

Responsiveness: Poor study quality due to a lack of appropriate statistical analysis.

Synthesis: One measurement property showed positive evidence based on limited evidence.

Discussion

This comprehensive systematic review identified five health-related QoL questionnaires used to assess AF populations, and evaluated their measurement properties using the COSMIN approach. To our knowledge, this is the first study to systematically evaluate the methodological quality of studies and the measurement properties of QoL questionnaires in AF patients. Our key finding was the lack of robust psychometric testing of any of the currently available AF-specific QoL instrument, particularly for measurement error. We found deficiencies in most instruments in relation to reliability, construct and criterion validity, but particularly in the responsiveness of questionnaires. The results raise important questions about the use of these instruments in research and daily clinical practice, and suggest the need for further evaluation before implementation.

Quality of life encompasses a person's perceptions of their physical and psychological state, social relationships and environment. [26] Many generic and disease-specific questionnaires exist to assess QoL in AF patients. The Short Form 36 (SF-36) is perhaps the most common generic QoL measure [4], and often used as a gold-standard comparator in the validation of other questionnaires (although as there is no 'true' gold-standard measure, criterion validity remains problematic to adequately appraise). Generic measures are useful to determine overall QoL in the population, particularly to quantify health utility or for comparing between diseases and populations. However, they may not be as sensitive to the effect of a single disease on QoL [11], whereas disease-specific questionnaires assess domains more relevant to the particular condition and may provide more targeted information to inform shared-decision making. [10] There is increasing interest in evaluating the effect of management strategies on QoL among AF patients, but for clinical confidence, the tool used should demonstrate good responsiveness, in addition to being a reliable and valid measure. Indeed, a recently developed QoL questionnaire for patients undergoing catheter ablation for cardiac arrhythmias has demonstrated responsiveness to clinical change. [27]

This review utilized the COSMIN approach for performing the systematic review. The COSMIN taxonomy was developed after an international Delphi study [13], using a consensus of terminology across validation studies and allowing standardized objective assessment of measurement properties. Further, the group have developed standards (the COSMIN checklist) for how measurement properties should be evaluated [28] and provide a structured protocol for undertaking such reviews to a high quality (<http://www.cosmin.nl/>). The COSMIN methodology has been incorporated into several other reviews to evaluate PROM assessment across a variety of diseases, including chronic obstructive pulmonary disease [29], osteoarthritis [30], cancer [31], multiple sclerosis [32] and Parkinson's Disease. [32] Whilst the standards within COSMIN are high, the thresholds for reliability are in line with the International Society for Quality of Life Research (ISOQOL) requirements. [33] It could be argued that high standards are needed to drive improvements in the methodology of development and validation of PROMs, and allow clinicians to have greater confidence in their use. As with any system based on scientific judgment, our results based on the COSMIN approach are only one component of evaluating quality of life questionnaires.

A specific limitation of our review was the inclusion of questionnaires developed and validated before the COSMIN checklist became widely available, which may have impacted on the scores assigned for methodological quality. This may be less of an issue in future reviews, where validation studies would ideally have incorporated the COSMIN checklist. There are other methodologies available for evaluating PROMs, for example the Federal Drug Administration scheme, which utilizes broadly similar assessment categories. [34] However, the COSMIN approach offers a more quantifiable appraisal of measurement properties. Secondly, the inclusion of two Japanese papers requiring translation meant these articles were assessed by a sole reviewer, which may impact on internal validity. We are also limited to the 'general' AF populations studied, whereas patients with specific comorbidities may respond differently to PROMs. For example, heart failure is a common comorbidity in AF [35] that is underrepresented in the studies evaluated.

The content of these questionnaires demonstrates some similarities among them, with 4 out of 5 examining the type and/or frequency of symptoms (the only exception being AFQoL), and all except QLAF assessing the impact of AF on daily activities, physical and psychological functioning, albeit to differing extent. Synthesis of results in our review showed that AFEQT demonstrated strong positive evidence for internal consistency and content validity and AFQoL showed strong positive evidence for content validity. All other questionnaires failed to provide sufficient evidence to make a strong positive rating. Neither measurement error nor cross-cultural validity were appropriately assessed in any study and there were limited data for responsiveness, which we were only able to rate for AFEQT. Where study quality was poor, or ratings indeterminate, this does not mean that a particular questionnaire performed badly, only that the evidence was limited. For example, AFQoL, in addition to AFEQT, demonstrated strong positive evidence for content validity, an important measurement property that is the foundation of a good QoL questionnaire.

Refinements to AF PROMs are ongoing, and AFQLQ was recently updated to a second version, demonstrating good intra- and inter-observer reproducibility and internal consistency in 40 AF patients. [36] However, even with full information on psychometric properties, clinical validity requires further appraisal in large-scale AF populations, a review of which is beyond the scope of this methodological assessment. Apart from AFEQT [37], there is limited information about the minimally important change seen with these questionnaires.

With the growing burden of AF and the equivocal benefit of rate or rhythm-control strategies in improving mortality [6–9], there is increasing importance of QoL assessment. Over 34 different QoL questionnaires have been utilized in published AF studies [10] emphasizing the current lack of consensus on the best instrument to assess QoL. Our assessment of methodological quality suggests that reported results for the available instruments should be treated with caution and that further, more robust validation is required to determine if these measures are appropriate to assess QoL in clinical populations. However, adequate measurement properties are just one aspect of whether a QoL tool is useful in research and clinical practice. Cost, ease of administration, time taken and patient acceptability are also important considerations. [29] Future studies are clearly required that address the detectable change in scores, as well as the responsiveness of the questionnaire to changes in patient symptoms and management strategies in the clinical setting. One such study that is due to commence shortly is the RATE Control Therapy Evaluation in Atrial Fibrillation (RATE-AF) randomized trial. [38]

Conclusion

Using the systematic COSMIN approach, we have identified clear deficiencies in the five available instruments for assessing disease-specific quality of life in patients with atrial fibrillation.

The strongest performing tool was AFEQT, but further evidence for test-retest reliability, measurement error and responsiveness are required before recommending routine clinical implementation.

Supporting Information

S1 Appendix. Search strategy for Medline.

(DOCX)

S2 Appendix. PRISMA checklist.

(PDF)

S1 Table. COSMIN criteria for measurement properties.

(DOCX)

S2 Table. COSMIN criteria for data synthesis.

(DOCX)

S3 Table. Table of excluded full-text studies.

(DOCX)

S4 Table. Validity assessment.

(DOCX)

S5 Table. Reliability and responsiveness assessment.

(DOCX)

Author Contributions

Conceptualization: DK.

Data curation: AA DAL DK.

Formal analysis: AA DAL DK.

Funding acquisition: DK.

Investigation: AA DAL.

Methodology: CT.

Project administration: DK.

Resources: DAL DK.

Supervision: DAL DK.

Validation: AA DAL DK.

Visualization: DK.

Writing – original draft: AA DAL DK.

Writing – review & editing: MC ML CT DK.

References

1. Stewart S. Cost of an emerging epidemic: an economic analysis of atrial fibrillation in the UK. *Heart*. 2004; 90(3):286–92. doi: [10.1136/hrt.2002.008748](https://doi.org/10.1136/hrt.2002.008748) PMID: [14966048](https://pubmed.ncbi.nlm.nih.gov/14966048/)

2. Thrall G, Lane D, Carroll D, Lip GY. Quality of life in patients with atrial fibrillation: a systematic review. *Am J Med.* 2006; 119(5):448 e1–19.
3. Waktare JE. Atrial fibrillation. *Circulation.* 2002; 106(1):14–6. PMID: [12093762](#)
4. Aliot E, Botto GL, Crijns HJ, Kirchhof P. Quality of life in patients with atrial fibrillation: how to assess it and how to improve it. *Europace.* 2014; 16(6):787–96. doi: [10.1093/europace/eut369](#) PMID: [24469433](#)
5. Kirchhof P, Benussi S, Kotecha D, Ahlsson A, Atar D, Casadei B, et al. 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS: The Task Force for the management of atrial fibrillation of the European Society of Cardiology (ESC). Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Endorsed by the European Stroke Organisation (ESO). *Eur Heart J.* 2016; doi: [10.1093/eurheartj/ehw210](#) PMID: [27567408](#)
6. Kotecha D, Holmes J, Krum H, Altman DG, Manzano L, Cleland JG, et al. Efficacy of beta blockers in patients with heart failure plus atrial fibrillation: an individual-patient data meta-analysis. *Lancet.* 2014; 384(9961):2235–43. doi: [10.1016/S0140-6736\(14\)61373-8](#) PMID: [25193873](#)
7. Ziff OJ, Lane DA, Samra M, Griffith M, Kirchhof P, Lip GY, et al. Safety and efficacy of digoxin: systematic review and meta-analysis of observational and controlled trial data. *BMJ.* 2015; 351:h4451. doi: [10.1136/bmj.h4451](#) PMID: [26321114](#)
8. Caldeira D, David C, Sampaio C. Rate versus rhythm control in atrial fibrillation and clinical outcomes: updated systematic review and meta-analysis of randomized controlled trials. *Arch Cardiovasc Dis.* 2012; 105(4):226–38. doi: [10.1016/j.acvd.2011.11.005](#) PMID: [22633297](#)
9. Al-Khatib SM, Allen LaPointe NM, Chatterjee R, Crowley MJ, Dupre ME, Kong DF, et al. Rate- and rhythm-control therapies in patients with atrial fibrillation: a systematic review. *Ann Intern Med.* 2014; 160(11):760–73. doi: [10.7326/M13-1467](#) PMID: [24887617](#)
10. Reynolds MR, Ellis E, Zimetbaum P. Quality of life in atrial fibrillation: measurement tools and impact of interventions. *J Cardiovasc Electrophysiol.* 2008; 19(7):762–8. doi: [10.1111/j.1540-8167.2007.01091.x](#) PMID: [18266667](#)
11. de Vries M, Ouwendijk R, Kessels AG, de Haan MW, Flobbe K, Hunink MG, et al. Comparison of generic and disease-specific questionnaires for the assessment of quality of life in patients with peripheral arterial disease. *J Vasc Surg.* 2005; 41(2):261–8. doi: [10.1016/j.jvs.2004.11.022](#) PMID: [15768008](#)
12. Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? *Injury.* 2011; 42(3):236–40. doi: [10.1016/j.injury.2010.11.042](#) PMID: [21145544](#)
13. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res.* 2012; 21(4):651–7. doi: [10.1007/s11136-011-9960-1](#) PMID: [21732199](#)
14. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res.* 2009; 18(8):1115–23. doi: [10.1007/s11136-009-9528-5](#) PMID: [19711195](#)
15. Yamashita T, Kumagi K, Koretsune Y, Mitamura H, Okumura K, Ogawa S, et al. A new method for evaluating quality of life specific to patients with atrial fibrillation: Atrial fibrillation quality of life questionnaire (AFQLQ). *Jpn J Electrocardiology.* 2003; 23:332–43.
16. Yamashita T, Komatsu K, Kumagi K, Uno K, Niwano S, Fujiki A, et al. Internal consistency and reproducibility of atrial fibrillation specific quality of life questionnaire (AFQLQ). *Jpn J Electrocardiology.* 2005; 25:488–94.
17. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007; 60(1):34–42. doi: [10.1016/j.jclinepi.2006.03.012](#) PMID: [17161752](#)
18. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010; 63(7):737–45. doi: [10.1016/j.jclinepi.2010.02.006](#) PMID: [20494804](#)
19. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res.* 2010; 19(4):539–49. doi: [10.1007/s11136-010-9606-8](#) PMID: [20169472](#)
20. Spertus J, Dorian P, Bubien R, Lewis S, Godejohn D, Reynolds MR, et al. Development and validation of the Atrial Fibrillation Effect on Quality-of-Life (AFEQT) Questionnaire in patients with atrial fibrillation. *Circ Arrhythm Electrophysiol.* 2011; 4(1):15–25. doi: [10.1161/CIRCEP.110.958033](#) PMID: [21160035](#)
21. Arribas F, Ormaetxe JM, Peinado R, Perulero N, Ramirez P, Badia X. Validation of the AF-QoL, a disease-specific quality of life questionnaire for patients with atrial fibrillation. *Europace.* 2010; 12(3):364–70. doi: [10.1093/europace/eup421](#) PMID: [20056594](#)

22. Badia X, Arribas F, Ormaetxe JM, Peinado R, de Los Terreros MS. Development of a questionnaire to measure health-related quality of life (HRQoL) in patients with atrial fibrillation (AF-QoL). *Health Qual Life Outcomes*. 2007; 5:37. doi: [10.1186/1477-7525-5-37](https://doi.org/10.1186/1477-7525-5-37) PMID: [17610734](https://pubmed.ncbi.nlm.nih.gov/17610734/)
23. Braganca EO, Filho BL, Maria VH, Levy D, de Paola AA. Validating a new quality of life questionnaire for atrial fibrillation patients. *Int J Cardiol*. 2010; 143(3):391–8. doi: [10.1016/j.ijcard.2009.03.087](https://doi.org/10.1016/j.ijcard.2009.03.087) PMID: [19395073](https://pubmed.ncbi.nlm.nih.gov/19395073/)
24. Harden M, Nystrom B, Kulich K, Carlsson J, Bengtson A, Edvardsson N. Validity and reliability of a new, short symptom rating scale in patients with persistent atrial fibrillation. *Health Qual Life Outcomes*. 2009; 7:65. doi: [10.1186/1477-7525-7-65](https://doi.org/10.1186/1477-7525-7-65) PMID: [19604399](https://pubmed.ncbi.nlm.nih.gov/19604399/)
25. Harden M, Nystrom B, Bengtson A, Medin J, Frison L, Edvardsson N. Responsiveness of AF6, a new, short, validated, atrial fibrillation-specific questionnaire—symptomatic benefit of direct current cardioversion. *J Interv Card Electrophysiol*. 2010; 28(3):185–91. doi: [10.1007/s10840-010-9487-3](https://doi.org/10.1007/s10840-010-9487-3) PMID: [20461545](https://pubmed.ncbi.nlm.nih.gov/20461545/)
26. World Health Organization. Measuring quality of life: the World Health Organization quality of life instruments. Available at: http://www.who.int/mental_health/media/68.pdf [Accessed 01 Feb 2016]. 1997.
27. White J, Withers KL, Lencioni M, Carolan-Rees G, Wilkes AR, Wood KA, et al. Cardiff cardiac ablation patient-reported outcome measure (C-CAP): validation of a new questionnaire set for patients undergoing catheter ablation for cardiac arrhythmias in the UK. *Qual Life Res*. 2015; doi: [10.1007/s11136-015-1194-1](https://doi.org/10.1007/s11136-015-1194-1) PMID: [26659900](https://pubmed.ncbi.nlm.nih.gov/26659900/)
28. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol*. 2010; 10:22. doi: [10.1186/1471-2288-10-22](https://doi.org/10.1186/1471-2288-10-22) PMID: [20298572](https://pubmed.ncbi.nlm.nih.gov/20298572/)
29. Weldam SW, Schuurmans MJ, Liu R, Lammers JW. Evaluation of Quality of Life instruments for use in COPD care and research: a systematic review. *Int J Nurs Stud*. 2013; 50(5):688–707. doi: [10.1016/j.ijnurstu.2012.07.017](https://doi.org/10.1016/j.ijnurstu.2012.07.017) PMID: [22921317](https://pubmed.ncbi.nlm.nih.gov/22921317/)
30. Dobson F, Hinman RS, Hall M, Terwee CB, Roos EM, Bennell KL. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. *Osteoarthritis Cartilage*. 2012; 20(12):1548–62. doi: [10.1016/j.joca.2012.08.015](https://doi.org/10.1016/j.joca.2012.08.015) PMID: [22944525](https://pubmed.ncbi.nlm.nih.gov/22944525/)
31. Wheelwright S, Darlington AS, Hopkinson JB, Fitzsimmons D, White A, Johnson CD. A systematic review of health-related quality of life instruments in patients with cancer cachexia. *Support Care Cancer*. 2013; 21(9):2625–36. doi: [10.1007/s00520-013-1881-9](https://doi.org/10.1007/s00520-013-1881-9) PMID: [23797577](https://pubmed.ncbi.nlm.nih.gov/23797577/)
32. Elbers RG, Rietberg MB, van Wegen EE, Verhoef J, Kramer SF, Terwee CB, et al. Self-report fatigue questionnaires in multiple sclerosis, Parkinson's disease and stroke: a systematic review of measurement properties. *Qual Life Res*. 2012; 21(6):925–44. doi: [10.1007/s11136-011-0009-2](https://doi.org/10.1007/s11136-011-0009-2) PMID: [22012025](https://pubmed.ncbi.nlm.nih.gov/22012025/)
33. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res*. 2013; 22(8):1889–905. doi: [10.1007/s11136-012-0344-y](https://doi.org/10.1007/s11136-012-0344-y) PMID: [23288613](https://pubmed.ncbi.nlm.nih.gov/23288613/)
34. US Department of Health and Human Services Food and Drug Administration. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. 2009; [Accessed 01-Jan-2016]; Available from: www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf.
35. Kotecha D, Piccini JP. Atrial fibrillation in heart failure: what should we do? *Eur Heart J*. 2015; 36(46):3250–7. doi: [10.1093/eurheartj/ehv513](https://doi.org/10.1093/eurheartj/ehv513) PMID: [26419625](https://pubmed.ncbi.nlm.nih.gov/26419625/)
36. Moreira RS, Bassolli L, Coutinho E, Ferrer P, Braganca EO, Carvalho AC, et al. Reproducibility and Reliability of the Quality of Life Questionnaire in Patients With Atrial Fibrillation. *Arq Bras Cardiol*. 2016; 106(3):171–81. doi: [10.5935/abc.20160026](https://doi.org/10.5935/abc.20160026) PMID: [27027365](https://pubmed.ncbi.nlm.nih.gov/27027365/)
37. Dorian P, Burk C, Mullin CM, Bubien R, Godejohn D, Reynolds MR, et al. Interpreting changes in quality of life in atrial fibrillation: how much change is meaningful? *Am Heart J*. 2013; 166(2):381–7.e8. doi: [10.1016/j.ahj.2013.04.015](https://doi.org/10.1016/j.ahj.2013.04.015) PMID: [23895823](https://pubmed.ncbi.nlm.nih.gov/23895823/)
38. Institute of Cardiovascular Sciences; University of Birmingham, Chief Investigator Kotecha D. Rate Control Therapy Evaluation in Permanent Atrial Fibrillation (RATE-AF). National Library of Medicine (US). 2015; NLM Identifier: NCT02391337 [Accessed 20-Nov-2015]; Available from: <https://clinicaltrials.gov/ct2/show/NCT02391337>