Patient reported outcome measures for facial skin cancer: A systematic review and evaluation of the quality of their measurement properties

Running Title: A systematic review of patient reported outcome measures for facial skin cancer

Authors

Thomas D. Dobbs MA, BM BCh, MRCS[1,2]*, Harsh Samarendra BA, BM BCh[3], Sarah Hughes BSc, MHSc, MRCSLT[4,5], Hayley A Hutchings BSc, PhD[4], Iain Whitaker MBBChir PhD FRCS(Plast)[1,2]

Affiliations

(1) Reconstructive Surgery and Regenerative Medicine Research Group (ReconRegen), Institute of Life Science 2, Swansea University Medical School, Swansea, UK

(2) The Welsh Centre for Burns and Plastic Surgery, Morriston Hospital, Swansea, UK

(3) Oxford University Medical School, Oxford, UK

(4) Patient and Population Health and Informatics, Institute of Life Science 2, Swansea University Medical School, Swansea, UK

(5) Abertawe Bro Morgannwg University Health Board, Princess of Wales Hospital, Bridgend, UK

Accepted for publication in British Journal of Dermatology:

Dobbs TD, Samarendra H, Hughes S, Hutchings HA, Whitaker I. Patient reported outcome measures for facial skin cancer: a systematic review and evaluation of their methodological and psychometric qualities. *British Journal of Dermatology* 2018. DOI 10.1111/bjd.17342.

Corresponding Author

Mr Thomas Dobbs

The Welsh Centre for Burns and Plastic Surgery,

Morriston Hospital,

Swansea.

SA6 6NL

e: tomdobbs@doctors.org.uk

Conflicts of Interest and Funding

All authors declare they have no conflicts of interest. This work has received no specific funding. Mr Thomas Dobbs is funded by the Welsh Clinical Academic Training Fellowship.

Word Count: 3032

Figures: 1

Tables: 6

Key Words: Facial skin cancer; patient-reported outcome measures; PROMs; systematic review; COSMIN

Summary

Introduction

Skin cancer is the commonest malignancy worldwide, often occurring on the face. Both the condition and treatment can lead to scarring and facial disfigurement, affecting a patient's health-related quality of life (HRQoL), which can be measured using patient-reported outcome measures (PROMs). This systematic review identifies PROMs for facial skin cancer and appraises their methodological quality and psychometric properties using up-to-date methods.

Methods

MEDLINE, EMBASE, PsychINFO, Cochrane and CINAHL were systematically searched in accordance with PRISMA guidelines, identifying all PROMs designed for or validated in facial skin cancer. Methodological quality and evidence of psychometric properties were assessed using the COnsensus-based Standards for the Selection of Health Measurement INstruments (COSMIN) checklist and criteria proposed by Terwee et al. A best evidence synthesis and assessment of instrument focus on post-resection reconstruction was also performed.

Results

Twenty-four studies on 11 PROMs were included. Methodological quality and psychometric evidence was variable, with the Patient Outcome of Surgery – Head/Neck (POS-H/N), Skin Cancer Index (SCI), Skin Cancer Quality of Life Impact Tool (SCQLIT) and Essers et al demonstrating the greatest level of validation. None scored well in their relevance to post-skin cancer reconstruction of the face.

Discussion

This systematic review critically appraises PROMs for facial skin cancer using internationally accepted criteria. The identified PROMs demonstrate a variation in the quality of validation performed, with a need to improve this across all PROMs in the field. Only through improving the quality of PROMs available and their focus on the post-treatment aesthetic and functional outcome will we be able to truly appreciate the concerns of our patients' and improve the management of facial skin cancer.

Bulleted Summary

1. **What is already known about this topic?**

Patient-reported outcome measures (PROMs) are important in both research and daily clinical practice. This is especially true in facial skin cancer, where both the condition and the resulting aesthetic outcome of treatment are important. PROMs for facial skin cancer exist, however their validity against the contemporary international consensus have yet to be reported. The relevance of these PROMs to patients' views of treatment outcomes is yet to be investigated.

2. **What does this study add?**

This systematic review provides a comprehensive assessment of the validity of those PROMs used for facial skin cancer using current best practice assessment tools, helping clinicians and researchers to select the most appropriate PROM to use. Each PROM is also assessed for relevance to the post-treatment aesthetic outcome, with a recommendation that further validated items are required to adequately assess this important area of skin cancer treatment.

Introduction

Skin cancer is the commonest malignancy worldwide(1), affecting 1 in 5 Americans during their lifetime(2). The incidence of non-melanoma skin cancers (NMSC) in England is 98.85/100,000 person-years and predominantly affects the face(3,4). 70,000 new diagnoses of NMSC were made in the United Kingdom in 2013(5), presenting a significant and growing health burden. Although skin cancer mortality is low, particularly for NMSC(5,6), the diagnosis is often psychologically damaging, including anxiety over the cancer diagnosis(7) and concerns over visible scarring, especially on the face(8), affecting health-related quality of life (HRQoL).

HRQoL has been given a number of definitions(9), but broadly represents an individual's perception of the effects of an illness and/or treatment on physical, psychological and social aspects of their life(10). One method for assessing HRQoL is the use of patient-reported outcome measures (PROMs). PROMs are standardised, validated questionnaires that are completed by patients and capture one or more aspects of their health and wellbeing(11,12). They are considered by the UK Department of Health as the current best method for quantifying a patient's clinical experience. Currently only four conditions have routine PROM data collected at a national level in the UK(12), although PROM data collection in many different cancer registries and dermatological trials is now commonplace(13,14).

Previous reviews have demonstrated a number of PROMs used in the assessment of patients with both skin cancer generally(15,16) and facial skin cancer(17). However, none have used current 'gold-standard' methodology for assessing the methodological quality of included studies, or the quality of those PROMs' measurement properties. Furthermore, given the burden associated with

cosmetic outcomes in post-skin cancer facial reconstruction, no review has yet assessed available PROMs for their focus on this. In an era of core outcome sets (COS)(18,19), where agreed upon minimum sets of outcomes when reporting research are expected, it is important that PROMs are appraised for their validity. If validation, or relevant items for the condition of interest are lacking, it is important that this is identified and rectified before inclusion in a COS.

The objectives of this systematic review are therefore to: (1) identify PROMs that have been designed for and/or validated in patients with facial skin cancer, (2) assess the methodological quality of the included studies, (3) assess the psychometric properties of those identified PROMs, (4) to make an assessment of the focus of each PROM on the reconstructive aspect of patient care and (5) to make recommendations that could lead to the development of a facial skin cancer COS.

Methods

Search Strategy and Selection Criteria

A systematic review protocol was developed in accordance with the Preferred Reporting for Items for Systematic Reviews and Meta-Analyses-Protocols (PRISMA-P)(20,21) and registered with PROSPERO (CRD42016043181).

The search strategy was constructed in line with PRISMA guidelines(22), the Cochrane handbook(23) and guidance from Terwee et al(24). To identify all papers that discussed some aspect of PROM development or validation for facial skin cancer, three separate constructs were explored; target condition, target body area and measurement instrument (e.g. PROM). Key words and MeSH terms were selected where available and searches were performed in; MEDLINE (Ovid), Embase (Ovid), PyschINFO (Ovid), Cochrane and CINAHL (EBSCO). An example search strategy can be seen in

Supplementary Figure 1. Grey literature and reference lists were also searched using Google, Google Scholar and known PROMs based websites. Searches were performed by two independent researches (TD and HS) on the same day in August 2016, with results uploaded to the reference management software package, EndNote® Version X7 (Clarivate Analytics). The search strategy was re-run prior to submission in January 2018 to identify any further studies that matched the inclusion criteria. Duplicates were removed using the functionality in EndNote®, with all references transferred to the online programme Covidence (www.covidence.org) for title and abstract screening. References were screened by two independent reviewers (TD and HS) according to the inclusion and exclusion criteria (*Table 1*), with all remaining articles downloaded in full-text format and re-screened. Discrepancies were discussed between the two reviewers with a third reviewer (HH) consulted if required.

Assessment of the Methodological Quality of included studies

The methodological quality of included studies was assessed using the COnsenus-based Standards for the Selection of Health Measurement INstruments (COSMIN)(25,26). The COSMIN checklist contains 9 main sections each assessing a different measurement property: internal consistency, reliability, measurement error, content validity, construct validity (structural validity and hypothesis testing), cross-cultural validity, criterion validity and responsiveness. An updated checklist with a 4-category rating scale (4-excellent, 3-good, 2-fair, 1-poor) was used(27). Each paper included in the review was compared against the 98-items in the checklist, and for those where evidence was presented in the paper, a score on the 4-category scale was given. One is only able to assess criterion validity where the PROM in question was compared to a longer version. Any paper describing criterion validity but not actually assessing

against a ‘gold standard’ or long-version was not assessed for criterion validity. The final rating for methodological quality in any given area of assessment is considered to be the lowest score (i.e. if a property such as internal consistency is scored ‘excellent’ in one question, but ‘poor’ in another, the methodological quality for that property is considered to be ‘poor’).

The COSMIN checklist has good inter-rater agreement and reliability(28), however to account for bias and subjectivity when rating studies it is considered good practice to compare results between two independent reviewers. A randomly-selected sample of 30% of the included studies were assessed by two reviewers (TD and SH) and compared using intraclass coefficient (ICC)(29), Cohen’s Kappa(30) and percentage agreement. If agreement was low in this sample, all included studies would be doubly assessed.

Assessment of Psychometric Properties

The psychometric quality of each PROM was assessed using criteria developed by Terwee et al(31) and updated in 2016(32). *Supplementary figure 2* describes the measurement properties that are assessed according to these criteria. Each criterion is rated as criteria met (+), criteria not met (-), or not all information present (?).

Data analysis and best evidence synthesis

Data were collated in Excel for Mac (V14.5.7) and presented as tables and narrative synthesis. Inter-rater reliability statistics were calculated for the COSMIN analysis using the Statistical Package for the Social Sciences (SPSS) software V.22 (IBM Corp., New York, USA).

A best evidence synthesis was performed by applying the levels of evidence summary as described by Furlan et al(33) to the combined results of the COSMIN and Terwee et al assessments. The Outcome Measures in Rheumatology (OMERACT) criteria were then used to categorise each instrument into A) instrument meets all requirements and is recommended for use, B) instrument meets two or more required items and therefore has potential for use, C) instrument has low quality in at least one area and is not recommended for use and D) instrument has almost no validatation(34). This method has previously been used by Gerbens et al in the dermatology literature(35).

Assessment of Reconstructive Relevance

The focus of each PROM on reconstruction post-skin cancer has never been assessed before and therefore there is no framework to work from. We therefore performed a subjective assessment of the included questions based on specialist knowledge of the topic area by the authors. As a reconstructive PROM was not the aim of the original scale developers we have performed this assessment separately and did not let this influence the COSMIN analysis when judging content validity.

Results

4886 articles were independently reviewed by two reviewers. With the addition of articles identified during reference searching a total of 24 studies were finally included (*Figure 1*)(7,36-58). Of those articles included, 11 different PROMs were identified: 2 generic PROMs (SF-36 and FACT-G) and 9 skin cancer-specific (FACT-M, POS-H/N, SCI, SCQoL, aBCCdex, SCQOLIT, FACE-Q, DLQI, Essers et al). As per the inclusion criteria, all PROMs included demonstrated some aspect of validation

in the facial skin cancer population. A summary of identified PROMs and included papers describing aspects of design or validation are presented in *Table 2*. A more detailed assessment of each instrument is presented in *Supplementary Figure 3*.

Methodological quality of those included studies

Raw individual category scores for each PROM are presented in *Table 3.* Of the 11 PROMs included, there was a range of methodological quality, with only one paper scoring in all 8 of the COSMIN categories (FACT-M). The spread of ratings between the 4 categories (excellent, good, fair and poor) was relatively even, with 28% being 'excellent', 18% 'good', 14% 'fair' and 40% 'poor'. The content validity for all bar 1 condition-specific PROMs (Essers et al) demonstrated 'excellent' methodology. Of the other categories, internal consistency and structural validity are the next two most commonly reported on and appropriately investigated areas of PROM development and validation in the identified studies.

ICC of 0.844 (0.796 – 0.88), Kappa of 0.648 ($p < 0.005$) and a percentage agreement of 97.84% was observed between the two reviewers, demonstrating good agreement.

Psychometric properties of included patient-reported outcome measures

The results of the psychometric evaluation are shown in *Table 4*. Of the 11 PROMs assessed, none scored positively in all domains. The PROMs with the lowest scoring psychometric measurement properties as assessed using criteria produced by Terwee et al were SF-36, FACT-G and FACE-Q skin cancer module. The FACE-Q skin cancer module was only described in outline in one paper(54), hence the scores noted in *Table 4*.

Content validity and internal consistency are the two most commonly reported on and well-designed aspects of PROMs validation papers. Seven out of 9 condition-specific PROMs showed 'appropriate assessment of content validity', demonstrating appropriate use of commonly used methods to generate items specific to the patient group(59). Good internal consistency, as demonstrated as having a Cronbach's alpha of 0.70 – 0.95, was shown in 8 of the 9 condition-specific PROMs.

The presentation of data in the included studies required to assess the other criteria of Terwee et al was, however, more sporadic. Overall, the SCI showed the greatest number of positive ratings across all domains.

Best evidence synthesis

A summary of the best evidence synthesis using the method outlined can be seen in *Table 5*. Using the OMERACT filter no PROMs met the criteria for an 'A' graded PROM, 4 PROMs were considered to be a 'B' graded PROM, 4 were 'C' grade PROMs, 2 were 'D' grade PROMs and 1 was un-gradable.

Focus on reconstructive aspects in each questionnaire

An assessment of the questions included in each questionnaire was made for their relevance to and focus on the reconstructive aspects and cosmesis of facial skin cancer. A summary of the questions that hold some relevance to reconstruction for each PROM is shown in *Table 6*.

Discussion

This systematic review has been designed to identify all PROMs that are validated for use in patients with facial skin cancer. At a time when the use of PROMs is being encouraged in both research and clinical use, it is important that only those PROMs that show evidence of validation are used. In the ideal world these would be validated in the exact population in which they were being implemented, however in practice this is often too time-consuming and expensive. Previous systematic reviews on this topic(16,17) have demonstrated many similar PROMs to this review, however we have assessed the methodological quality of these studies using internationally accepted criteria to minimise the risk of bias. This was performed using the COSMIN checklist, the current 'gold standard' for appraising and reporting the methodological rigour of studies reporting on instrument design and validation(26). It is now routinely accepted across the systematic review literature and has been used extensively in orthopaedics(60), paediatrics(61), neurology(62) and dermatology(35). A further update to the COSMIN methodology has been published, although this was after this review was performed(63). We also assessed the quality of the psychometric properties of the included PROMs using Terwee et al(32)'s criteria and performed a best evidence synthesis.

Of the two generic instruments identified, SF-36 and FACT-G, only rudimentary validation was provided in 1 paper(43). Both instruments are well established in the literature for their general use, however due to poor evidence of validation in the facial skin cancer population their use in this setting is difficult to recommend. This is mainly due to the instruments initially being designed for a different population to the one studied here and therefore they lack face and content validity. For example, the issues affecting a facial skin cancer population are likely to

be very different to those affecting the population groups used to design the SF-36. There was a range of quality with respect to design and validation across the 9 condition-specific PROMs identified. After removing FACE-Q from the analysis due to only very preliminary work being available, of the remaining 8 condition-specific PROMs internal consistency was measured in 7, with a range of ratings seen. Reliability was less frequently reported, but in a similar manner to internal consistency there was a range of ratings from poor to good. Measurement error and criterion validity were the most poorly reported, with only 3 PROMs demonstrating evidence of measurement error assessment. This may be due to the need for the instrument to be administered twice in order to calculate measurement error, increasing the time for data collection(59). Unfortunately, measurement error is an important concept required to design high quality prospective studies using these instruments. Evidence for content validity was excellent in all but 1 condition-specific PROM (Essers et al), with all condition-specific PROMs attempting to include representative patients in their design and validation. Structural validity and hypothesis testing were broadly done well. Criterion validity was poorly reported and in those reporting it, poorly done. This is due largely to the lack of a 'gold standard' comparator instrument. Finally, responsiveness was reported in 6 of the condition-specific PROMs. Ratings were either poor or fair and in a similar manner to measurement error, the need for at least two administrations in a longitudinal design may be why this is being poorly performed.

Combining the results of the COSMIN and Terwee et al analysis into a best evidence synthesis identified 4 PROMs that are currently the most appropriate for inclusion in a COS for facial skin cancer: POS-H/N, SCI, SCQOLIT and Essers et al. All of these still have deficiencies in their validation however (*Table 5*) and further

studies are advised. Furthermore, the FACE-Q skin cancer module has the potential to be a well-designed and validated instrument, but further studies are awaited.

This is the first systematic review on the subject to assess each PROM for their focus on the post-resection reconstruction of facial skin cancer. The results show that this is poorly addressed, even in PROMs designed specifically for facial skin cancer. Questions relating to the degree of scarring, how noticeable it is, physical symptoms such as pain and itch and psychological concerns all featured, but no single instrument adequately addressed this area. This is an important finding. In an era where skin cancer is treatable the long term sequelae of the treatment given is important, especially where this results in visible and potentially disfiguring scarring on the face. The only way in which the medical community will be able to improve the treatment offered is by asking patients what they think, through the medium of PROMs. It is therefore important that PROMs exist which include relevant and valid items relating to issues such as the reconstruction if they are to be included in a facial skin cancer COS. A COS for basal cell carcinomas is already in creation(64) and the CSG-COUSIN group(65) plan many more in the dermatology world. We therefore hope and implore that these take into account areas such as aesthetic and functional outcomes of reconstructive surgery.

The use of the COSMIN checklist is a strength of this study, however despite being validated and well accepted in the literature, there are limitations associated with it. Firstly, scoring of each item in the checklist is reliant on author judgment and therefore can be subjective. Secondly, the checklist is extensive and while this means it is considered to be the ‘gold-standard’ it is potentially difficult for the non-health outcome specialist to use.

In this systematic review we tried to control for inter-rater reliability issues by two independent reviewers assessing a randomly selected selection of papers. An intra-

class coefficient (ICC) score of 0.844 (considered ‘good’ by Koo and Li(66)), Kappa statistic of 0.648 ($p < 0.005$) (‘moderate agreement’(67)) and percentage agreement of 97.84% validated our inter-rater reliability and therefore COSMIN scores. While this assessment provides some reassurance when using COSMIN, we appreciate that it is feasible that another review team may score items differently.

Another strength of this systematic review is the use of a validated and highly sensitive search strategy, using guidance from the Cochrane group(23) and Terwee et al(24). We used a broad search strategy to identify all relevant studies demonstrating some aspect of design or validation of a PROM for facial skin cancer. However, this could also be a limitation in that we only included those studies that demonstrated aspects of design or validation. Studies that used a PROM in the facial skin cancer population but did not assess validation were excluded, potentially missing PROMs, which if they were validated, may be useful in this population group.

Conclusion

This systematic review has identified a number of different PROMs relevant to the facial skin cancer population. The identified PROMs demonstrated variable psychometric validation and all poorly addressed the reconstructive aspects of facial skin cancer. While POS-H/N, SCI, SCQOLIT and Essers et al all show potential, further validation work is required before they could be confidently included in a COS.

In order to move forward and improve our understanding of patients’ views on facial skin cancer and the difference between treatment options, it is important that these deficiencies in validation studies are addressed. Furthermore, additional items, either as an addition to a current PROM or included in an entirely new PROM, are required to specifically address the reconstruction and aesthetic outcomes of facial skin

cancer. It is hoped that in time the tools will exist to confidently assess our patients' views on their facial skin cancer and treatment outcomes, reducing the psychological and social burden associated with this disease.

Reference

1. Geller AC, Annas GD. Epidemiology of melanoma and nonmelanoma skin cancer. Seminars in Oncology Nursing. W.B. Saunders 2003;**19(1)**:2–11.

2. Understanding skin cancer. Available at www.skincancer.org (last accessed 23 November 2017).

3. Lomas A, Leonardi-Bee J, Bath-Hextall F. A systematic review of worldwide incidence of nonmelanoma skin cancer. Br J Dermatol 2012;**166(5)**:1069-80.

4. Franceschi S, Levi F, Randimbison L, La Vecchia C. Site distribution of different types of skin cancer: New aetiological clues. Int J Cancer 1996;**67(1)**:24–8.

5. Cancer Research UK statistics. 2016 Apr 7;:1–1. Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/skin-cancer/incidence (last accessed 15 October 2017).

6. Weinstock MA, Bogaars HA, Ashley M, Litle V, et al. Nonmelanoma skin cancer mortality. A population-based study. Arch Dermatol 1991;**127(8)**:1194–7.

7. Körner A, Garland R, Czajkowska Z, Coroiu A, et al. Supportive care needs and distress in patients with non-melanoma skin cancer: Nothing to worry about? Eur J Oncol Nurs 2016;**20**:150–5.

8. Sobanko JF, Sarwer DB, Zvargulis Z, Miller CJ. Importance of Physical Appearance in Patients With Skin Cancer. Dermatol Surg 2015;**41(2)**:183.

9. Karimi M, Brazier J. Health, Health-Related Quality of Life, and Quality of Life: What is the Difference? Pharmacoeconomics 2016;**34(7)**:645–9.

10. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research; U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research; U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes 2006;**4(1)**:79.

11. McGrail K, Bryan S, Davis J. Let's all go to the PROM: the case for routine patient-reported outcome measurement in Canadian healthcare. Healthc Pap. 2011;**11(4)**:8–18.

12. Devlin NJ, Appleby J. Getting the most out of PROMS. Putting health outcomes at the heart of NHS decision-making. Health Economics 2010. Available at http://www.kingsfund.org.uk/publications/proms.html (last accessed 23 November 2017).

13. Kotronoulas G, Kearney N, Maguire R, Harrow A, et al. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. J Clin Oncol 2014;**32(14)**:1480–501.

14. Brass D, Fouweather T, Stocken DD, Macdonald C, et al. An observer blinded randomised controlled pilot trial comparing localised immersion psoralen ultraviolet A (PUVA) with localised narrowband ultraviolet V (NBUVB) for

the treatment of palmar hand eczema. Br J Dermatol 2017;**13**:doi 10.1111/bjd.16238.

15. Lee EH, Klassen AF, Nehal KS, Cano SJ et al. A systematic review of patient-reported outcome instruments of nonmelanoma skin cancer in the dermatologic population. J Am Acad Dermatol 2013;**69(2)**:e59–e67.

16. Gibbons E, Casañas i Comabella C, Fitzpatrick R. A structured review of patient-reported outcome measures for patients with skin cancer, 2013. Br J Dermatol 2013;**168(6)**:1176–86.

17. Bates AS, Davis CR, Takwale A, Knepil GJ. Patient-reported outcome measures in nonmelanoma skin cancer of the face: a systematic review. Br J Dermatol 2013;**168(6)**:1187–94.

18. Williamson P, Altman D, Blazeby J, Clarke M, et al. Driving up the quality and relevance of research through the use of agreed core outcomes. J Health Serv Res Policy 2012;**17(1)**:1–2.

19. Porter ME, Larsson S, Lee TH. Standardizing Patient Outcomes Measurement. N Engl J Med 2016;**374(6)**:504–6.

20. Moher D, Shamseer L, Clarke M, Ghersi D, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015;**4(1)**:1.

21. Shamseer L, Moher D, Clarke M, Ghersi D, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015;**349**:g7647–7.

22. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. Ann Intern Med 2009;**151(4)**:264–9.

23. Higgins J, Green S. Cochrane handbook for systematic reviews of interventions version 5.1. 0 [updated March 2011]. The Cochrane Collaboration; 2011.

24. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res 2009;**18(8)**:1115–23.

25. Mokkink LB, Terwee CB, Patrick DL, Alonso J, 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.

26. Mokkink LB, Terwee CB, Patrick DL, Alonso J, 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.

27. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, et al. 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.

28. Mokkink LB, Terwee CB, Gibbons E, Stratford PW, Alonso J, Patrick DL, et al. Inter-rater agreement and reliability of the COSMIN (COnsensus-based Standards for the selection of health status Measurement Instruments)

Checklist. BMC Med Res Methodol 2010;**10(1)**:82.

29. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;**86(2)**:420–8.

30. Cohen J. A coefficient of agreement for nominal scales. Educational and psychological measurement 1960;**20(1)**:37-46.

31. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;**60(1)**:34–42.

32. Prinsen CAC, Vohra S, Rose MR, Boers M, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" – a practical guideline. Trials 2016;**17(1)**:449.

33. Furlan AD, Pennick V, Bombardier C, van Tulder M, et al. Updated method guidelines for systematic reviews in the Cochrane Back Review Group. Spine 2009;**34**:1929-41.

34. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. J Rheumatol 1998;**25**:198-99.

35. Gerbens LAA, Prinsen CAC, Chalmers JR, Drucker AM, et al. Evaluation of the measurement properties of symptom measurement instruments for atopic eczema: a systematic review. Allergy 2017;**72**:146-63.

36. Rhee JS, Matthews BA, Neuburg M, Burzynski M, et al. Creation of a Quality of Life Instrument for Nonmelanoma Skin Cancer Patients. Laryngoscope 2005;**115(7)**:1178–85.

37. Cormier JN, Davidson L, Xing Y, Webster K, et al. Measuring quality of life in patients with melanoma: development of the FACT-melanoma subscale. J Support Oncol 2005;**3(2)**:139–45.

38. Cormier JN, Ross MI, Gershenwald JE, Lee JE, et al. Prospective assessment of the reliability, validity, and sensitivity to change of the functional assessment of cancer Therapy-Melanoma questionnaire. Cancer. 2008;**112(10)**:2249–57.

39. Askew RL, Xing Y, Palmer JL, Cella D, et al. Evaluating minimal important differences for the FACT-Melanoma quality of life questionnaire. Value Health 2009;**12(8)**:1144–50.

40. Swartz RJ, Baum GP, Askew RL, Palmer JL, et al. Reducing patient burden to the FACT-Melanoma quality-of-life questionnaire. Melanoma Res 2012;**22(2)**:158–63.

41. Winstanley JB, Saw R, Boyle F, Thompson J. The FACT-Melanoma Quality-of-Life Instrument: comparison of a five-point and four-point response scale using the Rasch measurement model. Melanoma Res. 2013;**23(1)**:61–9.

42. Cano SJ, Browne JP, Lamping DL, Roberts AHN, et al. The Patient Outcomes of Surgery-Head/Neck (POS-head/neck): a new patient-based outcome measure. J Plast Reconstr Aesthet Surg 2006;**59(1)**:65–73.

43. Rhee JS, Loberiza FR, Matthews BA, Neuburg M, et al. Quality of Life Assessment in Nonmelanoma Cervicofacial Skin Cancer. Laryngoscope 2003;**113(2)**:215–20.

44. Matthews BA, Rhee JS, Neuburg M, Burzynski ML, et al. Development of the facial skin care index: a health-related outcomes index for skin cancer patients. Dermatol Surg 2006;**32(7)**:924–34.

45. Rhee JS, Matthews BA, Neuburg M, Logan BR, et al. Validation of a Quality-of-Life Instrument for Patients With Nonmelanoma Skin Cancer. Arch Facial Plast Surg 2006;**8(5)**:314–8.

46. Rhee JS, Matthews BA, Neuburg M, Logan BR, et al. The Skin Cancer Index: Clinical Responsiveness and Predictors of Quality of Life. Laryngoscope 2007;**117(3)**:399–405.

47. de Troya-Martín M, Rivas-Ruiz F, Blázquez-Sánchez N, Fernández-Canedo I, et al. A Spanish version of the Skin Cancer Index: a questionnaire for measuring quality of life in patients with cervicofacial nonmelanoma skin cancer. Br J Dermatol 2015;**172(1)**:160–8.

48. Vinding GR, Christensen KB, Esmann S, Olesen AB, et al. Quality of Life in Non-Melanoma Skin Cancer—The Skin Cancer Quality of Life (SCQoL) Questionnaire. Dermatol Surg 2013;**39(12)**:1784–93.

49. Vinding GR, Esmann S, Olesen AB, Hansen LB, et al. Interpretation of the Skin Cancer Quality of Life Score: A Validated Quality of Life Questionnaire for Non-Melanoma Skin Cancer. Dermatology. 2014;**229(2)**:123–9.

50. Mathias SD, Chren M-M, Colwell HH, Yim YM, et al. Assessing Health-Related Quality of Life for Advanced Basal Cell Carcinoma and Basal Cell Carcinoma Nevus Syndrome. JAMA Dermatol 2014;**150(2)**:169–8.

51. Mathias SD, Chren MM, Crosby RD, Colwell HH, et al. Reliability and validity of the Advanced Basal Cell Carcinoma Index (aBCCdex). Br J Dermatol 2015;**173(3)**:713–9.

52. Burdon-Jones D, Thomas P, Baker R. Quality of life issues in nonmetastatic skin cancer. Br J Dermatol 2010;**162(1)**:147–51.

53. Burdon-Jones D, Gibbons K. The Skin Cancer Quality of Life Impact Tool (SCQOLIT): a validated health-related quality of life questionnaire for non-metastatic skin cancers. J Eur Acad Dermatol Venereol 2013;**27(9)**:1109–13.

54. Lee EH, Klassen AF, Lawson JL, Cano SJ, et al. Patient experiences and outcomes following facial skin cancer surgery: A qualitative study. Australas J Dermatol 2015;**57(3)**:e100-4.

55. Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)—a simple practical measure for routine clinical use. Clin Exp Dermatol 1994;**19(3)**:210–6.

56. Blackford S, Roberts D, Salek MS, Finlay A. Basal cell carcinomas cause little handicap. Qual Life Res 1996;**5(2)**:191–4.

57. Essers BAB, Nieman FHM, Prins MH, Krekels GAM, et al. Determinants of satisfaction with the health state of the facial skin in patients undergoing surgery for facial basal cell carcinoma. Patient Educ Couns 2006;**60(2)**:179–86.

58. Essers B, Nieman F, Prins M, Smeets N, et al. Perceptions of facial aesthetics in surgical patients with basal cell carcinoma. J Eur Acad Dermatol Venereol 2007;**21(9)**:1209–14.

59. Streiner DL, Norman GR, Cairney J. Health Measurement Scales. Fifth Edition. Oxford University Press 2015.

60. Gagnier JJ, Mullins M, Huang H, Marinac-Dabic D et al. A Systematic Review of Measurement Properties of Patient-Reported Outcome Measures Used in Patients Undergoing Total Knee Arthroplasty. J Arthroplasty 2017;**32(5)**:1688–1697.e7.

61. Speyer R, Cordier R, Parsons L, Denman D, et al. Psychometric Characteristics of Non-instrumental Swallowing and Feeding Assessments in Pediatrics: A Systematic Review Using COSMIN. Dysphagia 2017;**31(1)**:1–14.

62. Haywood KL, Mars TS, Potter R, Patel S, et al. Assessing the impact of headaches and the outcomes of treatment: A systematic review of patient-reported outcome measures (PROMs). Cephalalgia 2017; doi: 10.1177/0333102417731348

63. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018;**27(5)**:1147-1157.

64. Schlessinger DI, Iyengar S, Yanes AF, Lazaroff JM, et al. Development of a core outcome set for clinical trials in basal cell carcinoma: study protocol for a systematic review of the literature and identification of a core outcome set using a Delphi survey. Trials 2017;**18(1)**:490.

65. Schmitt J, Deckert S, Alam M, Apfelbacher C, et al. Report from the kick-off meeting of the Cochrane Skin Group Core Outcome Set Initiative (CSG-COUSIN). Br J Dermatol 2016;**174(2)**:287-95.

66. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 2016;**15(2)**:155–63.

67. McHugh ML. Interrater reliability: the kappa statistic. Biochemia Medica 2012;**22(3)**:276–82.

68. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Medical care 1992;**30(6)**:473–83.

69. Brazier JE, Harper R, Jones NM, O'Cathain A, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. BMJ 1992;**305(6846)**:160–4.

70. McHorney CA, Ware JE, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Medical care 1993;**31(3)**:247–63.

71. Cella DF, Tulsky DS, Gray G, Sarafian B, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. J Clin Oncol 1993;**11(3)**:570–9.

72. Rhee JS, Matthews BA, Neuburg M, Smith TL, et al. Skin Cancer and Quality of Life: Assessment With the Dermatology Life Quality Index. Dermatol Surg 2004;**30(4)**:525–9.

73. Nunes DH, FrÖde TS. Quality of Life in Basal Cell Carcinoma Patients in Brazil: A Pilot Cross Sectional Study. Dermatol Surg 2013;**39(4)**:620–6.

74. Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984;**11(1)**:1–47.

Table 1 – Inclusion and exclusion criteria used when screening identified studies.

Inclusion criteria	1) Head and neck skin cancer population 2) Papers discussing some aspect of PROM development or validation 3) English only articles
Exclusion criteria	1) Questionnaires not developed or validated in patients with head and neck skin cancer 2) Oropharyngeal head and neck cancer population 3) Questionnaires developed to assess nodal or distant metastatic disease 4) General oncology questionnaires unless specifically validated in a head and neck skin cancer population

Table 2 – Summary of patient-reported outcome measures (PROMs) and corresponding papers identified using the inclusion and exclusion criteria of this systematic review. The number of items in each questionnaire and domains assessed are documented.

PROM	Papers included	Generic or condition-specific	Number of items	Domains
SF-36	Rhee et al, 2003(43)	Generic	36	• Vitality • Physical functioning • Bodily pain • General health perception • Physical role functioning • Emotional role functioning • Social role functioning • Mental health
FACT-G	Rhee et al, 2003(43)	Generic	27	• Physical • Social/family • Emotional • Functional well-being
FACT-M	Cormier et al, 2005(37) Cormier et al, 2008(38) Askew et al, 2009(39) Swartz et al, 2012(40) Winstanley et al, 2013(41)	Condition-specific	24 (in FACT-M subscale) 18 in reduced version	• Physical well-being • Emotional well-being • Social well-being
POS-H/N	Cano et al, 2006(42)	Condition-specific	15 (6 pre-operatively and 9 post-operatively)	• Psychological functioning and cosmetic appearance • Satisfaction
SCI	Rhee et al, 2005(36) Matthews et al, 2006(44) Rhee et al, 2006(45) Rhee et al, 2007(46) de Troya-Martin et al, 2015(47) Korner et al, 2016(7)	Condition-specific	15	• Emotion • Social • Appearance
SCQoL	Vinding et al, 2013(48) Vinding et al, 2014(49)	Condition-specific	9	• Function • Emotions • Control
aBCCdex	Mathias et al, 2014(50) Mathias et al, 2015(51)	Condition-specific	26	• Worry about future lesions • Mental health

				• Social/Relationships • Lesion symptoms • Life impact
SCQOLIT	Burdon-Jones et al, 2010(52) Burdon-Jones et al, 2012(53)	Condition-specific	10	• Psychosocial • Physical
FACE-Q	Lee et al, 2015(54)	Condition-specific	N/A	N/A
DLQI	Finlay et al, 1994(55) Blackford et al, 1996(56)	Generic skin PROM	10	• Symptoms and feelings • Daily activities • Leisure • Work and school • Personal relationships • Treatment
Esser et al	Essers et al, 2006(57) Essers et al, 2007(58)	Condition-specific	22	• Worrying about facial health • Susceptibility for facial BCC • Fear of developing a new BCC

Table 3 – Individual category scores for each study for all included patient-reported outcome measures (PRO
point scale. Each domain is made up of a number of questions as part of the COSMIN checklist, with the low
the overall methodological quality for that domain in the paper assessed.

PROM	Paper	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypotheses testing	C
SF-36	Rhee et al, 2003	Poor	--	--	--	--	--	
FACT-G	Rhee et al, 2003	Poor	--	--	--	--	--	
FACT-M	Cormier et al, 2005	--	--	--	Excellent	--	--	
	Cormier et al, 2008	Poor	Poor	Poor	Poor	Poor	Good	
	Winstanley et al, 2012	Excellent	--	--	--	Excellent	--	
	Swartz et al, 2012	Excellent	Poor	--	--	Excellent	--	
POS-Head/Neck	Cano et al, 2005	Excellent	Poor	Excellent	Excellent	--	Poor	

FSCI/SCI	Rhee et al, 2005	Poor	--	--	Excellent	--	--
	Matthews et al, 2006	Poor	Poor	--	Excellent	Poor	--
	Rhee et al, 2006	Good	--	--	--	--	Good
	Rhee et al, 2007	--	--	--	--	--	--
	de Troya-Martin, 2012	Good	Good	--	Excellent	Good	Poor
SCQoL	Vinding et al, 2013 *IRT	Poor	--	--	Excellent	Excellent	Fair
	Vinding et al, 2014	--	--	--	--	--	Fair
aBCCdex	Mathias et al, 2014	--	--	--	Excellent	--	--
	Mathias et al, 2015	Good	Good	--	--	Good	Fair
SCQOLIT	Burdon-Jones et al, 2009	--	--	--	Excellent	--	--
	Burdon-Jones et al, 2012	Good	Good	Poor	Excellent	Poor	Good

FACE-Q Skin cancer module	Lee et al, 2015	--	--	--	--	--	--
DLQI	Finlay et al, 1993	--	Fair	--	Excellent	Poor	--
	Blackford et al, 1996	--	--	--	--	--	Poor
Essers et al	Essers et al, 2006 and 2007	Good	--	--	Fair	Fair	--

All domains are scored according to the COSMIN checklist with 4-point scale(27). Potential categories inclu

(--) indicates domains not measured in a study. * refers to the use of Item Response Theory, rather than Class

Table 4 - Individual category scores assessing psychometric properties for each study for all included patient

(PROMs) as developed by Terwee et al(31,32).

PROM	Paper	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypotheses testing	C
SF-36	Rhee et al, 2003	-						
FACT-G	Rhee et al, 2003	-						
FACT-M	Cormier et al, 2005				+			
	Cormier et al, 2008	+			+		+	
	Winstanley et al, 2012	+				+		
	Swartz et al, 2012	-						
POS-Head/Neck	Cano et al, 2005	+	+		+		+	

FSCI/SCI	Rhee et al, 2005						
	Matthews et al, 2006	+			+	-	
	Rhee et al, 2006	+					+
	Rhee et al, 2007						
	de Troya-Martin, 2012	-	+		+		+
SCQoL	Vinding et al, 2013 *IRT	+			+	+	-
	Vinding et al, 2014						-
aBCCdex	Mathias et al, 2014				+		
	Mathias et al, 2015	+	-		+		+
SCQOLIT	Burdon-Jones et al, 2009				+		
	Burdon-Jones et al, 2012	+	+		+		+

FACE-Q Skin cancer module	Lee et al, 2015				
DLQI	Finlay et al, 1993			+	
	Blackford et al, 1996				+
Essers et al	Essers et al, 2006 and 2007	+		+	

Each criterion is assessed as either; positive rating (+), negative rating (-), or indeterminate rating (?). (Blank

where no evidence is presented.

Table 5 – Best evidence synthesis and grading according to the OMERACT filter

PROM	Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypothesis testing	Criterion validity	Responsive
SF-36	?							
FACT-G	?							
FACT-M	±	?	?	±	±	++	?	-
POS-H/N	+++	?	±	+++		?		?
SCI	±	±		+++	±	±		+
SCQoL	?			+++	+++	-		-
aBCCdex	++	--		+++	±	+		?
SCQOLIT	++	++	?	+++	?	++		?
FACE-Q								
DLQI		-		+++	?	?		
Essers et al	++			+	±			

Positive rating for measurement property (+++ consistent findings in multiple studies of good methodologica

quality / ++ consistent findings in multiple studies of fair methodological quality or one good study / + one st

Negative rating for measurement property (--- consistent findings in multiple studies of good methodological

quality / -- consistent findings in multiple studies of fair methodological quality or one good study / - one stu

indicated indeterminate due to poor quality study. ± indicates conflicting evidence.

OMERACT filter using categories of A, B, C, D as discussed in the methods. Category of B/C where a PRO

Table 6 – Assessment of each questionnaire for a focus on questions relating to reconstruction and the post-treatment aesthetics

PROM	Questions with a focus relevant to reconstruction	Global summary of focus on reconstruction
SF-36	No questions relevant to reconstruction	Absent
FACT-G	No questions relevant to reconstruction	Absent
FACT-M	Four items show some relevance *- I feel numbness at my surgical site* *- I have pain at my melanoma site or surgical scar* *- I worry about the appearance of surgical scars* *- I have swelling as a result of surgery*	Poor
POS-H/N	Post-surgical questionnaire attempts to address aspects of the operation and outcomes *- are the results of the operation on your head/neck skin growths – better/about/worse than expected?* *- if a friend has a similar head/neck skin growths that you had before your operation would you recommend the same operation you had?*	Average
SCI	Two items relating to scarring *- worried about how large the scar will be?* *- thought about how noticeable the scar will be to others?*	Average
SCQoL	No focus on the treatment or reconstructive aspect. One question with a vague reference to aesthetics *- during the past week, I have used such things as make-up or clothing to hide my skin cancer from others*	Poor
aBCCdex	Items relevant to appearance *- your appearance changing due to surgery or procedures* Three items relating to the lesion *- bleeding from lesion(s)* *- oozing or pus from lesions(s)* *- sensitive/tender skin around lesion(s)* However, no questions with a focus on the reconstruction	Poor
SCQOLIT	One item relating to disfigurement and one relating to discomfort following the treatment *- over the last week, how much have you been bothered about any disfigurement or scarring, in respect to your skin cancer or its treatment?* *- over the last week, how much skin discomfort or inconvenience have you experienced, in respect to your skin cancer or its treatment?*	Poor/Average

FACE-Q	No specific questionnaire items have yet to be published but one of the aims of the new skin cancer module is to address areas around facial aesthetics	Absent
DLQI	No questions relevant to reconstruction	Absent
Esser et al	No questions relevant to reconstruction	Absent

Figure Legend:

Figure 1 – PRISMA flow diagram demonstrating the identification and screening of studies for inclusion