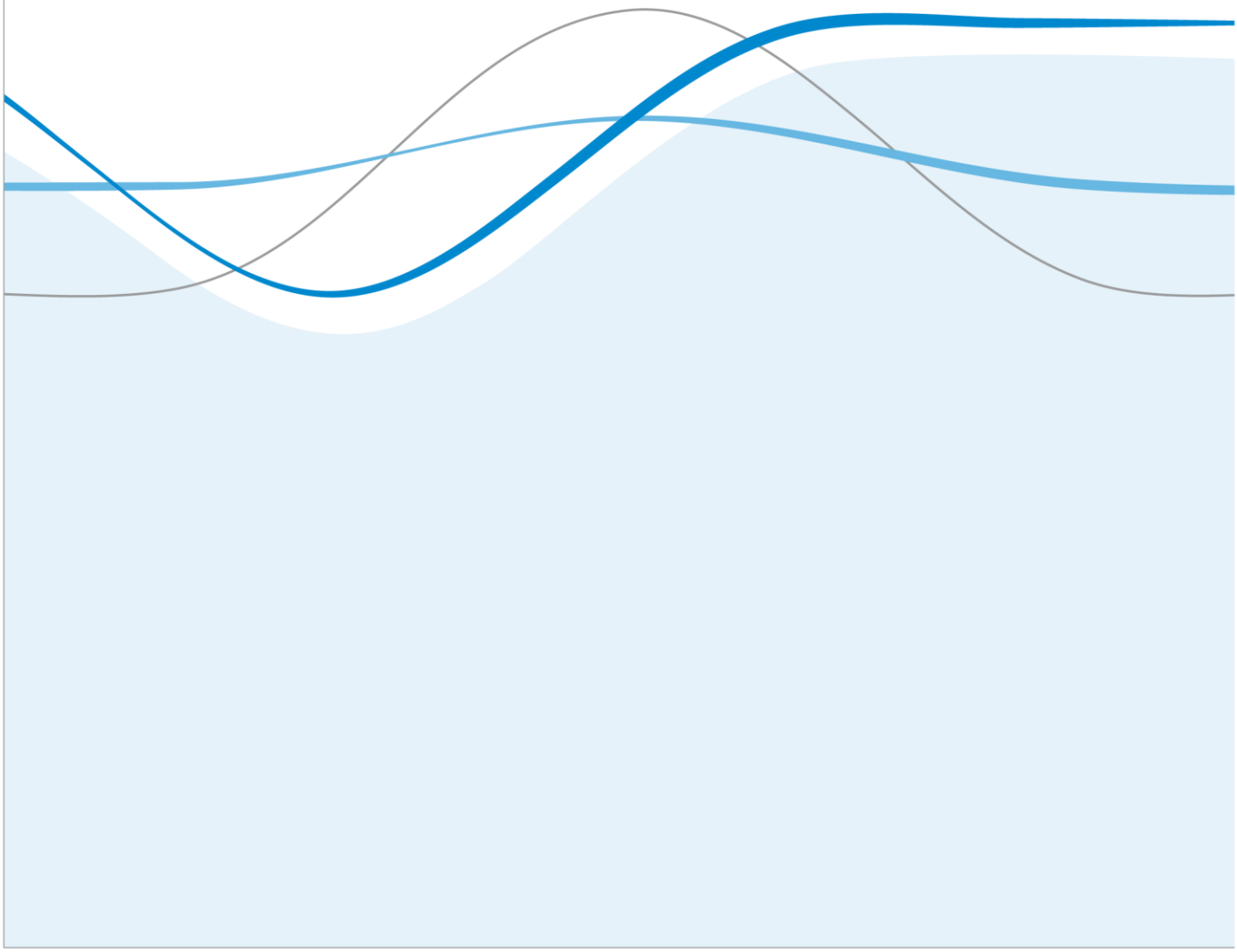
Office of Health Economics Consulting

Consulting Report

Quality of Life in Long-term Cancer Survivors: Implications for Future Health Technology Assessments in Oncology

June 2018

Patricia Cubi-Molla, David Mott, Koonal Shah, Mike Herdman, Yvonne Summers, Nancy Devlin



Quality of Life in Long-term Cancer Survivors: Implications for Future Health Technology Assessments in Oncology

Patricia Cubi-Molla[a], David Mott[a], Koonal Shah[a], Mike Herdman[a], Yvonne Summers[b] and Nancy Devlin[a]

[a]Office of Health Economics; [b]The Christie and Manchester University Foundation Trusts

June 2018

Please cite this report as:

Cubi-Molla, P., Mott, D., Shah, K., Herdman, M., Summers, Y. and Devlin, N. (2018). *Quality of Life in Long-term Cancer Survivors: Implications for Future Health Technology Assessments in Oncology*. OHE Consulting Report, London: Office of Health Economics. Available at: https://www.ohe.org/publications/quality-life-long-term-cancer-survivors-implications-future-health-technology

OHE Consulting Ltd
(a registered company number 09853113)
Southside, 7th Floor
105 Victoria Street
London SW1E 6QT
United Kingdom

For further information please contact
Koonal Shah
Principal Economist
Tel: +44 (0)20 7747 8856
Or: +44 (0)7920 496832
kshah@ohe.org

About OHE Consulting Reports

Many of the studies OHE Consulting performs are proprietary and the results are not released publicly. Studies of interest to a wide audience, however, may be made available, in whole or in part, with the client's permission. They may be published by OHE alone, jointly with the client, or externally in scholarly publications. Publication is at the client's discretion.

Studies published by OHE as OHE Consulting Reports are subject to internal quality assurance and undergo a rapid review by a member of the OHE Editorial Panel.

Funding and acknowledgements

This study was funded by Roche Products Ltd. The authors are grateful to the project steering group (Carolina Tsuchiya, Ajay Patel, Matt Hodgson, Eleftherios Sideris, Jessica Purchase, Ning Yu) for their comments on earlier draft versions and advice throughout the project. The authors are also grateful to Martin Buxton for his comments on an early draft version.

Table of Contents

EXECUTIVE SUMMARY

Background

Cancer survival rates have improved dramatically in recent decades due in part to pharmaceutical advances, with a growing range of increasingly effective and targeted medicines being developed, such as immunotherapies. In the economic modelling of such treatments, the question arises of which utilities should be assigned to patients who show a long-term, durable response. In recent critiques of economic models in this area by the National Institute for Health and Care Excellence (NICE), the idea that long-term cancer survivors (LTCS) who have received such treatments could report quality of life (QoL) scores which are similar to, or higher than, those of equivalent general population samples has not been viewed as credible. This literature review examines whether there is evidence to support the assumption that the QoL of LTCS can be similar to that of age/sex-matched population samples.

Methods

A search strategy was devised, following significant testing, to identify studies in electronic databases of published articles which assessed QoL in LTCS treated with immunotherapy. However, as the testing indicated that there may not be very many studies with this type of data, additional searches were devised to cover recent conference abstracts, and inclusion criteria were expanded to include studies that explored QoL in LTCS regardless of treatment type. Records were screened using a two-stage approach (using two sets of inclusion and exclusion criteria) to determine which should be retained for full text analysis.

Results

A total of 20 papers were included in the review, representing 23 studies. The studies covered a range of countries and different types of cancers (breast, lung, head and neck, colorectal, ovarian and prostate), with the majority (n=13; 57%) collecting data from more than 100 LTCS. LTCS included in the studies were more likely to have experienced early-stage cancer relative to late-stage cancer. The most common treatment interventions experienced by LTCS were chemotherapy, radiotherapy and surgery. Only one study contained LTCS that had been treated with immunotherapy, but this was only the case for four patients. None of the included studies provided data on health state utility values.

Of the included studies, just under half (n=11; 48%) compared the QoL of a sample of LTCS and the QoL of a control group. The type of control group varied but the majority of the studies used age and/or sex-adjusted samples of the general population (n=6). Four of the 11 studies generated evidence that suggests that QoL in LTCS was better than that of the control group, with one suggesting that QoL in LTCS was 'similar or better'. One study suggested that QoL in LTCS was similar to the control group, with another providing mixed evidence. Three studies generated evidence to suggest that QoL in LTCS was 'similar or worse' compared to the control group, with only one study providing evidence that QoL was worse for LTCS.

Discussion

The majority of studies which have made explicit comparisons between LTCS and the general population suggest that levels of QoL in the two groups are quite similar. This

could provide some evidence to support an argument for applying general population utilities to LTCS in economic models for cancer treatments.

It may be the case that cancer treatment is effective in restoring patients to a high level of QoL in the long-term. Other possible explanations for this finding could include 'adaptation', where cancer patients adapt to their condition, potentially due to changing their internal standards regarding their QoL (known as 'response shift'). However, no studies explored this in detail. Alternatively, individuals with cancer may improve their lifestyle following diagnosis which could improve their QoL in the long-term. However, evidence of lifestyle change is mixed in the included studies.

There are also potential limitations in the data. This could include selection bias, where a fully representative sample cannot be recruited. This could be the case if less healthy LTCS were less likely to have been recruited, and to have responded, to the studies. In addition, some of the sample sizes were quite small and some of the reporting and analysis was of low quality (e.g. no statistical significance testing in some studies).

Conclusion

Notwithstanding the very limited evidence base, of the studies reporting comparisons with population norms, the majority provided evidence that the QoL of LTCS is similar to that of the general population. However, it is unclear how generalisable the results are, due to the different study settings. Therefore, care is required in asserting that general population utilities could be used to represent LTCS in economic models. More directed research is likely required in a broad range of cancer and treatment types to provide a stronger foundation for such an assertion. In particular, there is need for more data on the QoL of LTCS who were treated with immunotherapy for advanced/metastatic disease.

1. INTRODUCTION

Cancer survival rates have improved dramatically in recent decades due in part to pharmaceutical advances, with a growing range of increasingly effective and targeted medicines being developed (Paterson, 2017). In particular, the use of immunotherapies to treat several types of cancer has shown very encouraging results in recent years, with some treatments leading to impressive long-term survival rates (McDermott et al., 2016; Brahmer et al., 2017).

In the economic modelling of such treatments, the question arises of which utilities should be assigned to patients who show a long-term, durable response to such treatment. In recent critiques of economic models in this area by NICE, the idea that long-term survivors who have received such treatments could report quality of life (QoL) scores which are similar to or higher than equivalent general population samples has not been viewed as credible (NICE, 2015; NICE, 2017).

A common concern is whether health economics models assume appropriate QoL outcomes for cancer survivors who have received treatment (particularly immunotherapy) and survive a long time (what constitutes 'a long time' may vary across cancer types). This literature review examines whether there is evidence to support this view and to support the assumption that the QoL of long-term cancer survivors (LTCS) recovers to healthy levels (e.g. comparable to age/sex-matched population norms).

The research questions for the study can therefore be defined as follows:

- Are utility gains in health economic models being appropriately estimated, particularly in the immunotherapy area where the treatments can result in very long-term survival or have the *potential* to constitute a tantamount cure?
- Where patients receiving treatment survive a long time, are there arguments for asserting that their QoL can legitimately be modelled at the pre-progression or baseline levels that are comparable to population norms?

2. METHODS

2.1. Sources of data

The source of data for the review was an electronic search of the MEDLINE and Embase databases (accessed via Ovid), carried out in October 2017. MEDLINE was used to identify full text articles in the life sciences and biomedical literature. Embase was used to identify conference abstracts, many of which do not appear in MEDLINE.

The MEDLINE search was restricted to records published since the year 2001, which is recognised as the launch year of the first targeted therapy (reflecting the overlap between targeted therapy and immunotherapy, and the fact that both work differently from standard chemotherapy). The Embase search was restricted to records published since the year 2015, reflecting an assumption that any worthwhile results presented at conferences before that time would now be available as full articles (and therefore be picked up by the MEDLINE search). Both searches were for English language records only.

2.2. Search strategy

An iterative approach was used to develop and refine the search strategy.

2.2.1. Search focusing on immunotherapy

A preliminary, exploratory search combined the terms *cancer* AND *immunotherapy* AND *quality of life*. This resulted in a very large number of records, and an informal review of the first 40-50 titles suggested that the vast majority of these were not relevant to the research question. Specifying the type of cancer in the search (e.g. *lung cancer*) greatly reduced the number of hits but very few of the resulting records appeared to be relevant.

Terms related to *cancer* (e.g. *tumour*), *immunotherapy* (e.g. *immune-oncology*) and *quality of life* (e.g. *utility*) were added to increase the sensitivity of the search. These also included specific names of two widely-used QoL measures, *EQ-5D* and *SF-36*. Brand and generic names of relevant immunotherapy drugs were also added, as alternatives to *immunotherapy*. Finally, the term *survival* (and closely related terms; see below) were added.

Spelling variants, plurals and closely related terms were captured via the use of wildcard symbols (e.g. *surviv** was used to represent related terms such as *survival*, *survivor*, *survivors* and *survivorship*).

This process resulted in the following search strategy, which was used to search both the MEDLINE and Embase databases for title and abstract terms:

> (cancer* OR neoplasm* OR tumor* OR tumour* OR carcinoma)
>
> AND (immunotherap* OR immune-oncology OR targeted therap* OR EGFR inhibitor* OR atezolizumab OR Tecentriq OR nivolumab OR Opdivo OR pembrolizumab OR Keytruda OR durvalumab OR Imfinzi OR avelumab OR Bavencio OR ipilimumab OR Yervoy OR PD-1 pathway inhibitor* OR PD-L1 pathway inhibitor* OR PD-1 inhibitor* OR PD-L1 inhibitor* OR ALK inhibitor*)
>
> AND surviv*
>
> AND (utilit* OR quality of life OR QOL OR health-related quality of life OR HRQOL OR health status OR EQ-5D OR SF-36)

2.2.2. Parallel search focusing on long-term survivors

At the onset of the study it was acknowledged that the literature on QoL in cancer survivors following immunotherapy (as sought using the search strategy described above) may not include much evidence on long-term outcomes. Indeed, prefixing *surviv** with *long-term* dramatically reduced the number of hits.

It was therefore deemed useful to conduct a parallel search focusing on long-term cancer survival outcomes, but not necessarily linked to immunotherapy. This involved adding the terms *long term*, *long standing*, *long-term*, *long-standing*, *disease-free* and *remission*. This process resulted in the following search strategy, which was used to search the MEDLINE database:

> (cancer* or neoplasm* or tumor* or tumour* or carcinoma)
>
> AND (long term or long-term or long standing or long-standing or disease-free or remission)
>
> AND surviv*
>
> AND (utilit* or quality of life or QOL or health-related quality of life or HRQOL or health status or EQ-5D or SF-36)

This search was restricted to keywords found in titles only, since extending to abstracts increased the number of records to an unmanageable number.

2.2.3. Parallel search focusing on conference presentations

It was noted that many key immunotherapy trials would not yet have sufficiently mature data to appear in published journal articles, but emerging results may have been reported at recent conferences. A further parallel search was therefore conducted, focusing on terms related to *immunotherapy*, terms related to *quality of life*, and terms representing the names of key trials already known to the authors and/or the project steering group. This process resulted in the following search strategy, which was used to search the Embase database for title and abstract terms:

> (immunotherap* OR immune-oncology OR targeted therap* OR EGFR inhibitor* atezolizumab OR Tecentriq OR nivolumab OR Opdivo OR pembrolizumab OR Keytruda OR durvalumab OR Imfinzi OR avelumab or Bavencio or ipilimumab OR Yervoy OR PD-1 pathway inhibitor* OR PD-L1 pathway inhibitor* OR PD-1 inhibitor* OR PD-L1 inhibitor* OR ALK inhibitor*)
>
> AND (IMmotion* OR PCD4989g OR FIR OR BIRCH OR POPLAR OR OAK OR IMPower* OR CheckMate* OR IMvigor* OR Keynote* OR PACIFIC)
>
> AND (utilit* OR quality of life OR QOL OR health-related quality of life OR HRQOL OR health status OR EQ-5D OR SF-36)

The final search strategy is shown in Table 1.

Table 1. Final search strategy

	Search Terms	Restrictions	Database
1	(cancer* OR neoplasm* OR tumor* OR tumour* OR carcinoma) AND (immunotherap* OR immune-oncology OR targeted therap* OR EGFR inhibitor* OR atezolizumab OR Tecentriq OR nivolumab OR Opdivo OR pembrolizumab OR Keytruda OR durvalumab OR Imfinzi OR avelumab OR Bavencio OR ipilimumab OR Yervoy OR PD-1 pathway inhibitor* OR PD-L1 pathway inhibitor* OR PD-1 inhibitor* OR PD-L1 inhibitor* OR ALK inhibitor*) AND surviv* AND (utilit* OR quality of life OR QOL OR health-related quality of life OR HRQOL OR health status OR EQ-5D OR SF-36)	Titles/Abstracts 2001-Present No restriction by record type English language Humans	MEDLINE (via Ovid)
2	(cancer* OR neoplasm* OR tumor* OR tumour* OR carcinoma) AND (immunotherap* OR immune-oncology OR targeted therap* OR EGFR inhibitor* OR atezolizumab OR Tecentriq OR nivolumab OR Opdivo OR pembrolizumab OR Keytruda OR durvalumab OR Imfinzi OR avelumab OR Bavencio OR ipilimumab OR Yervoy OR PD-1 pathway inhibitor* OR PD-L1 pathway inhibitor* OR PD-1 inhibitor* OR PD-L1 inhibitor* OR ALK inhibitor*) AND surviv* AND (utilit* OR quality of life OR QOL OR health-related quality of life OR HRQOL OR health status OR EQ-5D OR SF-36)	Titles/Abstracts 2015-Present Conference Abstracts English language Humans	Embase (via Ovid)
3	(cancer* or neoplasm* or tumor* or tumour* or carcinoma) AND (long term or long-term or long standing or long-standing or disease-free or remission) AND surviv* AND (utilit* or quality of life or QOL or health-related quality of life or HRQOL or health status or EQ-5D or SF-36)	Title only 2001-Present No restriction by record type English language Humans	MEDLINE (via Ovid)
4	(immunotherap* OR immune-oncology OR targeted therap* OR EGFR inhibitor* atezolizumab OR Tecentriq OR nivolumab OR Opdivo OR pembrolizumab OR Keytruda OR durvalumab OR Imfinzi OR avelumab or Bavencio or ipilimumab OR Yervoy OR PD-1 pathway inhibitor* OR PD-L1 pathway inhibitor* OR PD-1 inhibitor* OR PD-L1 inhibitor* OR ALK inhibitor*) AND (IMmotion* OR PCD4989g OR FIR OR BIRCH OR POPLAR OR OAK OR IMPower* OR CheckMate* OR IMvigor* OR Keynote* OR PACIFIC) AND (utilit* OR quality of life OR QOL OR health-related quality of life OR HRQOL OR health status OR EQ-5D OR SF-36)	Titles/Abstracts 2015-Present Conference Abstracts English language Humans	Embase (via Ovid)

2.3. Selection of studies for inclusion

2.3.1. Initial criteria

The following eligibility criteria were originally proposed:

1. Participants
   Include:
   - Studies in which participants can be considered LTCS (defined initially as an individual who has survived for three or more years following treatment initiation)
   - Studies in which participants developed cancer in adulthood (i.e. over 18 years of age) of any ethnicity, sex, socioeconomic status, occupation and associated morbidities
   - Studies of any type of cancer

   Exclude:
   - Studies focusing exclusively on short survival durations
   - Studies focusing on childhood cancers

2. Interventions
   Include:
   - Studies of any intervention intended for the treatment of cancer, particularly immunotherapy drugs

3. Comparators
   Include:
   - Studies using any comparator, or no comparator

4. Outcomes
   Include:
   - Studies reporting evidence on health-related QoL, as assessed using any measure of overall health status or QoL, including condition-specific measures

5. Study design and publication type
   Include:
   - Studies which provide evidence of QoL valuations/ratings in cancer survivors
   - Original quantitative (interventional or observational) or qualitative research, including randomised clinical trials, quasi-experimental trials, longitudinal cohorts, cross-sectional surveys, interviews and focus group studies
   - Reviews of the literature (systematic or otherwise)

   Exclude:
   - Studies which do not provide evidence of QoL valuations/ratings in cancer survivors
   - Case reports, position papers, and studies that do not report or analyse data

6. Setting
   Include:
   - Studies carried out in any location and setting

7. Language

Include:

- English language publications

Exclude:

- Publications in any language other than English

2.3.2. Revised criteria

An initial screening of the titles and abstracts of records identified by the literature search (see section 3.1) indicated that very few studies reported long-term QoL outcomes for cancer survivors *following immunotherapy*. Further, many potentially eligible studies: focused on early stage (as opposed to advanced/metastatic) cancer; focused on non-solid tumours; or were not original research studies.

Based on these initial findings, it was judged that the eligibility criteria should be revised in order to permit the selection of studies that were more closely relevant to the overarching research question. It was acknowledged that the revised criteria constituted a shift in focus away from immunotherapy (since few relevant long-term immunotherapy studies had been identified) and towards long-term QoL outcomes for cancer survivors more generally.

The revised eligibility criteria are:

1. Participants

   Include:

   - Studies in which at least 25% of participants had advanced (stages III/IV) cancer[1]
   - Studies in which patients had solid tumours

   Exclude:

   - Studies in which fewer than 25% of participants had advanced cancer[1]
   - Studies failing to report the proportion of participants who had advanced cancer
   - Studies focusing on non-solid tumours

2. Interventions

   Include:

   - Studies of any intervention intended for the treatment of cancer, particularly immunotherapy drugs

3. Comparators

   Include:

   - Studies using any comparator, or no comparator

4. Outcomes

   Include:

   - Studies reporting evidence on health-related QoL, as assessed using any measure of overall health status or QoL, including condition-specific measures

   Exclude:

   - Studies failing to report evidence on health-related QoL

[1] Specifically, 25% of the participants whose cancer stage is known/reported

5. Study design
   Include:
   - Studies which provide evidence of QoL valuations/ratings in cancer survivors
   - Original quantitative (interventional or observational) or qualitative research, including randomised clinical trials, quasi-experimental trials, longitudinal cohorts, cross-sectional surveys, interviews and focus group studies

   Exclude:
   - Any studies that do not report or analyse data
   - Literature reviews, network meta-analyses and cost-effectiveness analyses

6. Length of follow-up
   Include:
   - Studies with at least two years of follow-up data

   Exclude:
   - Studies with less than two years of follow-up data
   - Studies that failed to report the length of follow-up

7. Setting
   Include:
   - Studies carried out in any location and setting

8. Research question
   Exclude:
   - Studies with a research question deemed to be irrelevant (subjective judgement)

9. Format and data availability
   Exclude:
   - Conference summaries
   - Studies for which a full text article was unavailable and the abstract did not contain sufficient information

Records were screened in three stages: (1) checking whether they could be excluded based on the title; (2) checking whether they could be excluded based on the abstract; and (3) checking whether they could be excluded based on the full text, if available.

2.3.3. Screening of records

An initial screening of titles and abstracts applying the original criteria (see section 2.3.1) was conducted by two members of the OHE Consulting team, working independently of each other. Discrepancies between reviewers were discussed and resolved through consensus.

A second screening of titles, abstracts and full texts applying the revised criteria (see 2.3.2) was conducted by two members of the project steering group. Only records retained following the initial screening were taken forward to the second screening.

2.4. Data extraction

Data were extracted by two members of the OHE Consulting team, using an extraction form comprising the following fields (see Appendix A):

- Full citation
- Type of publication
- Country
- Study objective(s) [using authors' own terminology where possible]
- Study type
- Sample size [full sample; key subgroups]
- Sample characteristics
- Cancer type
- Cancer stage [at time of study, if reported; at time of diagnosis]
- Intervention(s) [intervention being assessed in study; any prior interventions reported]
- Any reference to immunotherapy? [yes/no]
- Length of survival
- Outcome measure(s)
- Length of study follow-up
- Main results [relating to QoL]
- Comparison with QoL at general population [if made]
- Comparison with QoL at baseline [if made]
- Key verbatim quotes of relevance to the research question

A pilot test was conducted whereby both reviewers independently extracted data for the same record and compared their outputs. The results of the pilot test confirmed the shared understanding and consistency of approach between the reviewers.

3. RESULTS

3.1. Literature search output

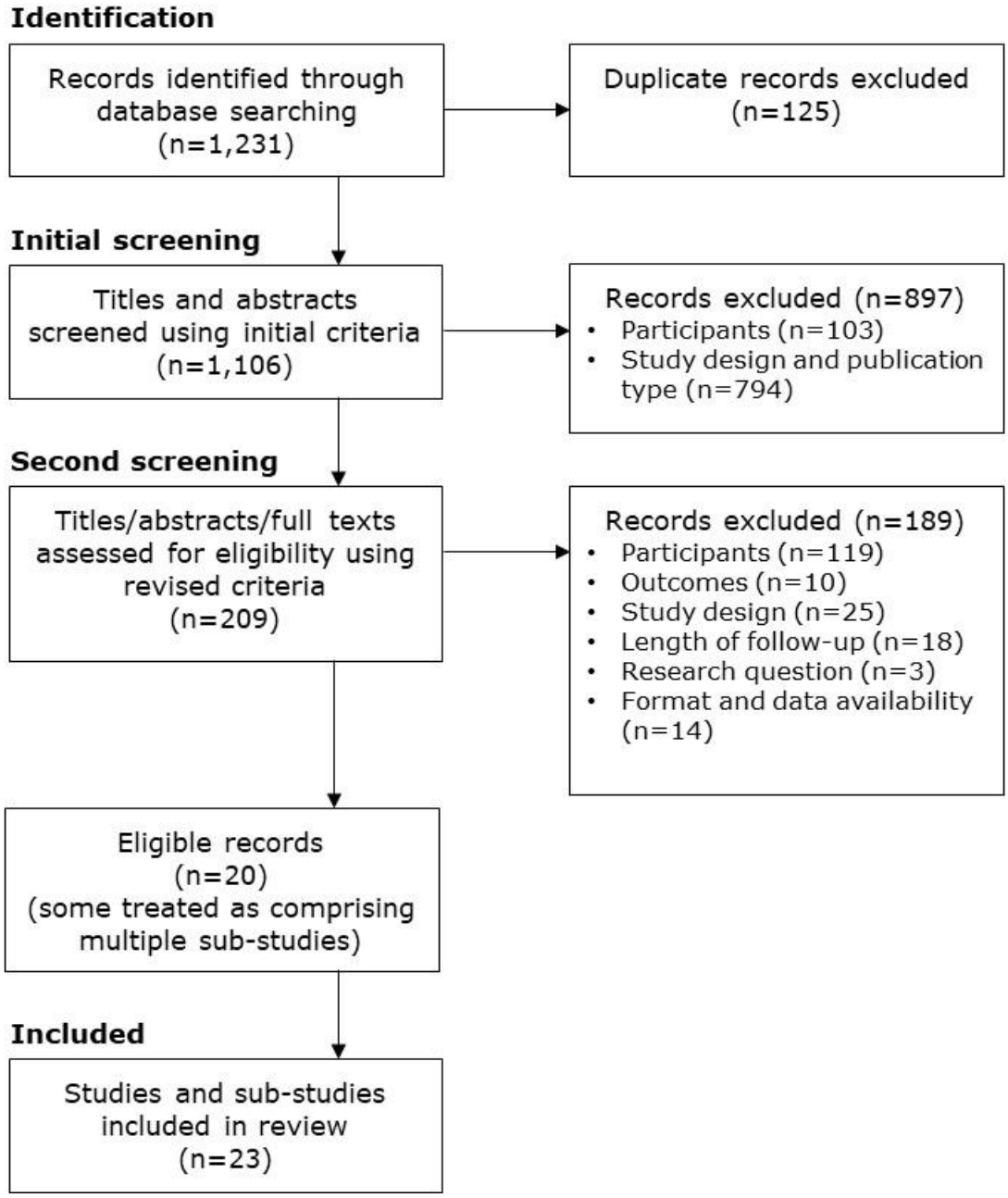
The final search strategy yielded 1,231 results, of which 125 duplicates were excluded, leaving 1,106 records. Based on an initial screening of titles and abstracts applying the original criteria (see section 2.3.1), 794 records were excluded for failing to meet the *Study design and publication type* criterion and a further 103 records were excluded for failing to meet the *Participants* criterion (all due to the shortness of survival durations), leaving 209 records.

Based on a second screening of the remaining 209 records applying the revised criteria (see 2.3.2), seven records could not be accessed, 17 records were excluded based on titles, 61 records were excluded based on abstracts, and 104 records were excluded based on full texts. This meant that a total of 20 records were included in the review. Figure 1 illustrates the search and selection process, providing the reasons for exclusion of studies at each stage.

There were two cases in which a given record reported separate results, analyses and discussions for multiple cancer types (paper 1: colon cancer and rectal cancer; paper 5: colorectal cancer; prostate cancer; breast cancer). In these cases, we treated each record as comprising multiple sub-studies, where each study is reported as a standalone study in our review. Hence, our review includes 23 studies and sub-studies derived from

20 published articles. In all cases, the sub-studies each had samples sizes of more than 100.

Figure 1. Literature search and selection flow diagram



3.2. General results

A summary of the studies can be found in Table 2. Half of the papers report results from studies performed in the USA while the majority of the remaining studies were performed in European countries, i.e. France (studies 1a, 1b), Spain (studies 5a, 5b, and 5c), Austria (study 6), Sweden (study 7), Germany (studies 8 and 9), the Netherlands (studies 12 and 19), and Norway (study 13). Study 1 was performed in Brazil. The sample size of the population under study ranges from 11 (study 6) to 1,937 (study 17) LTCS. The studies usually included patients with different cancer stages at the time of diagnosis (information not reported in two studies: 8 and 11). The mean length of survival (LoS) is in general greater than five years, but was lower in study seven (3 years) and study 19 (4.5 years).[2]

The primary objective of each study is shown in Table 2. Most of the studies had as their stated objective that of exploring and evaluating the QoL of LTCS. Two studies (5 and 12) aimed to explore the evolution of the QoL of LTCS across different points in time, and only one paper (study 8) directly addressed the comparison of QoL in LTCS with adjusted population norms. The remaining studies (3, 9, 13, 16, and 17) had other stated objectives such as exploring how physical activity levels impact the QoL of long-term survivors.

Table 3 describes the distribution characteristics of key variables for the 23 studies.

All the studies included in the final selection are journal articles, and most published in the last eight years. A total of nine studies (3, 6, 7, 9, 12, 16-19), or 39%, can be classified as longitudinal and all were prospective, i.e. they collected information from the same patient cohort at different points in time. Of those studies analysing cross-sectional data, only study 20 can be classified as a prospective study design (the data were obtained from a large longitudinal study, but only observations corresponding to a single cross section are explored in the paper). There were six (26%) population-based studies (1a, 1b, 2, 9, 13, and 14). In the other studies, patients were recruited using convenience sampling techniques in several hospitals or patient-related units. Most of the studies recruited well-defined cohorts, typically cancer patients diagnosed in one or more centres during a time period of more than a year (studies 3-5, 7, 8, 12-17, and 19). Sampling methods in the remaining studies (6, 10, 11, 18, and 20) were more *ad hoc*. Information about patient enrolment was not reported for study 19.

Thirteen studies (56%) collected data from more than 100 cancer LTCS, with two (9%) studies analysing data from more than 500 LTCS (16, 17). Three studies (4, 10, and 19) had a sample size of between 50 and 100 patients, and the remaining seven studies (30%) collected data from fewer than 50 patients.

[2] The mean LoS has been estimated in the following studies: 1a, 1b, 10, 11, 15-17, 19, and 20 (the estimates are provided in italics in Table 2). *Methods*: the mean has been derived as the mean of the different LoS categories adjusted by the frequencies in studies 1a, 1b, 16, and 17. The mean LoS has been assumed to equal the (reported) median in studies 10, 11, 15, and 20. For study 19 the mean LoS is computed as the midpoint between the minimum and maximum LoS (the latter categories reported in the papers). In study 4 the mean LoS cannot be computed, since only the minimum LoS is reported (5 years).

Table 2. Summary of the included studies (n=23 from n=20 papers)

No.	Authors (date)	Country	Study type[1]	Cancer type[2]	Sample size	Stage I-II n (%)[3]	Stage III-IV n (%)[3]	LoS-mean (years)[4]	Primary objective[5]
1a	Caravati-Jouvenceaux et al. (2011)	France	CS -R	GI	344	231 (67%)	78 (23%)	*13.6*	Explore how colon cancer impacts QoL over the long term
1b		France	CS -R	GI	198	133 (67%)	34 (17%)	*13.6*	Explore how rectal cancer impacts QoL over the long term
2	Carmichael *et al.* (2013)	USA	CS -R	GG	28	0 (0%)	28 (100%)	5.1	Characterize the HRQoL of long-term survivors with metastatic renal cell carcinoma
3	Clark *et al.* (2008)	USA	Long	L	272	204 (75%)	68 (25%)	5.0	Explore how physical activity level impacts the QoL of long-term lung cancer survivors
4	Duke et al. (2005)	USA	CS -R	H&N	86	41 (48%)	42 (49%)	N/R	Analyse the dental status in long-term H&N cancer survivors
5a	Ferro et al. (2014)	Spain	CS -R	GI	134	84 (63%)	33 (25%)	7.5	Ascertain and compare the health status between the colorectal cancer survivors at 5 and 10 years
5b		Spain	CS -R	GG	105	51 (49%)	18 (17%)	7.5	Ascertain and compare the health status between the prostate cancer survivors at 5 and 10 years
5c		Spain	CS -R	B	344	287 (83%)	27 (8%)	7.5	Ascertain and compare the health status between the breast cancer survivors at 5 and 10 years
6	Greimel et al. (2011)	Austria	Long	GG	11	3 (27%)	8 (73%)	10	Examine the long-term QoL of patients with epithelial ovarian cancer
7	Hammerlid et al. (2001)	Sweden	Long	H&N	135	66 (49%)	55 (41%)	3.0	Examine the HRQoL of a large group of H&N cancer survivors 3 years after diagnosis
8	Hartung et al. (2016)	Germany	CS -R	GG	164	85 (52%)	79 (48%)	11.6	Compare long-term HRQoL in germ cell tumour survivors and age-adjusted men
9	Jansen et al. (2011)	Germany	Long	GI	483	328 (67%)	159 (33%)	5.4	Investigate the prevalence of benefit finding and post-traumatic growth

No.	Authors (date)	Country	Study type[1]	Cancer type[2]	Sample size	Stage I-II n (%)[3]	Stage III-IV n (%)[3]	LoS-mean (years)[4]	Primary objective[5]
10	Lutgendorf et al. (2017)	USA	CS -R	GG	56	0 (0%)	55 (98%)	*14.0*	Examine QoL, survivorship concerns, and lifestyle factors among long-term survivors of advanced-stage epithelial ovarian cancer
11	Meisel et al. (2012)	USA	CS -R	B	18	0 (0%)	18 (100%)	*7.2*	Explore the QoL and psychosocial issues in patients who are living long term with metastatic breast cancer
12	Oskam et al. (2013)	Netherlands	Long	H&N	27	6 (22%)	20 (74%)	9.2	Evaluate changes in HRQoL from baseline and short-term follow-up to long-term follow-up in advanced oral and oropharyngeal cancer patients
13	Osthus et al. (2011)	Norway	CS -R	H&N	139	85 (61%)	54 (39%)	6.3	Study HRQoL survival predictions of successfully treated H&N squamous cell carcinoma patients
14	Payakachat et al. (2013)	USA	CS -R	H&N	47	0 (0%)	47 (100%)	8.4	Review QoL and outcomes after 5 years posttreatment for head and neck cancer survivors
15	Phipps et al. (2008)	USA	CS -R	GI	30	23 (77%)	7 (23%)	*7.0*	Investigate current physical and psychosocial problems as reported by long-term colon cancer survivors
16	Rausch et al. (2012)	USA	Long	L	1149	694 (60%)	447 (39%)	*5.0*	Identify SNPs [single nucleotide polymorphisms] related to pain in lung cancer survivors
17	Nes et al. (2012)	USA	Long	L	1937	1130 (58%)	687 (35%)	*5.6*	Examine whether change in physical activity level from diagnosis to follow-up would be associated with change in QoL in a large sample of lung cancer LTCS
18	Vainshtein et al. (2015)	USA	Long	H&N	40	0 (0%)	40 (100%)	6.5	Report long-term HRQoL outcomes in patients with locally advanced OPC
19	van der Schroeff et al. (2007)	Netherlands	Long	H&N	57	20 (35%)	37 (65%)	*4.5*	Evaluate and compare treatment outcome and QoL of older and younger H&N cancer patients
20	Vartanian et al. (2009)	Brazil	CS -P	H&N	273	0 (0%)	273 (100%)	*5.2*	Evaluate the acceptance of treatment and the QoL of LTCS of advanced H&N cancer patients

[1]CS -R: cross-sectional retrospective; CS -P: cross-sectional prospective; Long: longitudinal

[2]GI: gastrointestinal; GG: genitourinary and gynecologic; L: lung; B: breast; H&N: head and neck

[3]% of patients at stages III or IV at diagnosis, over the total sample (which in some papers includes patients whose stage was "unknown")

[4]*Figures in italic*: estimated

[5]QoL: quality of life; HRQoL: health-related quality of life; LTCS: long-term survivors; N/R: not reported

Table 3. Distribution of the key variables in the included studies

Variable	Freq.	%
Publication type		
Journal article	23	100%
Study type (1)		
Cross sectional	14	61%
Prospective	1	4%
Retrospective	13	57%
Longitudinal	9	39%
Study type (2)*		
Population-based	6	26%
Centre-based	16	70%
Year of publication		
Before 2010	6	26%
In 2010 or after	17	74%
Country		
Europe	12	52%
USA	10	43%
Other	1	4%
Sample size		
≤50	7	30%
51-100	3	13%
101-500	11	48%
>500	2	9%
Average sample size by cancer stage		
Stage I+II	161	
Stage III+IV	106	
Stage unknown	14	
Over half of the patients in stage III+IV	8	35%
Cancer type		
Breast	2	9%
Gastrointestinal	5	22%
Genitourinary and gynaecologic	5	22%
Head and neck	8	35%
Lung	3	13%

Variable	Freq.	%
Average length of survival (years)		
≤5	4	17%
5.1-10	14	61%
>10	4	17%
Unknown	1	4%
Inclusion of PROMs		
EORTC QLQ-C30	10	43%
EORTC QLQ-H&N35	4	17%
SF-36/SF-8	10	43%
FACT-G	1	4%
FACT-Kidney (FKSI-15)	1	4%
FACT-Head and Neck	1	4%
FACT-O	1	4%
FACT-Breast	1	4%
LASA	2	9%
UWQOL	3	13%
NHP	3	13%
CES-D	2	9%
Other	8	35%
Examination of QoL at different points in time		
Increasing over time	4	17%
Constant over time	3	13%
Decreasing over time	0	0%
Mixed evidence	1	4%
No comparison made	16	70%
Comparison with population norms		
Better than general population	4	17%
Better than or similar to general population	1	4%
Similar to general population	1	4%
Similar to or worse than general population	3	13%
Worse than general population	1	4%
Mixed evidence	1	4%
No comparison made	12	52%

Variable	Freq.	%
Intervention		
Chemotherapy	18	78%
Radiotherapy	18	78%
Surgery	18	78%
Immunotherapy	1	4%
Other	6	26%
Unknown	1	4%
Not reported	1	4%
Population mean age		
18-65	11	48%
65+	12	52%
Definition of "length of survival"		
Years since diagnosis	18	78%
Years since diagnosis with metastasis	2	9%
Years since treatment completion	3	13%

Variable	Freq.	%
Suggested explanatory factors of the results		
Response shift	6	26%
Borderline significance	2	9%
Selection bias	4	17%
Adaptation	2	9%
Long-term complications or sequelae		
Bowel problems	3	13%
Fatigue	3	13%
Social relations problems	3	13%
Sexual functioning	3	13%
Swallowing	3	13%
Other	3	13%

QoL: quality of life; EORTC QLQ-C30: European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire; EORTC QLQ-H&N35: Head and Neck 35 cancer module; SF-36: Short form (36) health survey; SF-8: Short form (8) health survey; FACT-G: Functional Assessment of Cancer Therapy-General; FACT-Kidney: Functional Assessment of Cancer Therapy–Kidney Symptom Index; FACT-Head and Neck: Functional Assessment of Cancer Therapy Additional Concerns for Head and Neck; FACT-O: Functional Assessment of Cancer Therapy-Ovarian; FACT-Breast: Functional Assessment of Cancer Therapy-Breast; LASA: Linear analogue self-assessments; UWQOL: University of Washington Quality of Life scale; NHP: Nottingham Health Profile; CES-D: Center for Epidemiological Studies Depression scale

*It was not possible to classify one of the studies in this manner.

Patients enrolled in the studies were usually in early stages of cancer at the time of diagnosis, with an average of 161 patients per study in stage I or stage II, compared to 106 patients per study in later stages. Only eight (35%) of the studies reported more patients in later stages than in early stages (2, 6, 10, 12, 14, 18-20). On average, cancer stage was reported as unknown in 14 patients per study. This figure is derived from six studies (1a, 1b, 10, 12, 16, and 17), which reported cancer stage as "unknown" for 35, 31, 1, 1, 8 and 120 patients, respectively.

In Tables 2 and 3 (and the analyses the follow), patients in stages III and IV are clustered together as 'stage III+IV'. This is because nine studies (39%) did not report data for stages III and IV separately, and those that did tended to have too few patients in their respective sub-samples to permit any meaningful comparisons between patients in these different stages.

Eight of the studies analysed (35%) focused on head and neck cancer (detailed as oropharynx, oral cavity, nasopharynx, larynx or other, in study 4; oral or oropharyngeal, in study 12; head and neck squamous cell carcinoma in study 13; nasopharynx, oral cavity, oropharynx, pharynx, hypopharynx, or larynx in study 14; oropharyngeal, in p18). Five studies (22%) analysed data from cancer types classified as gastrointestinal (colon, in study 1a and study 15; rectal, in study 1b; colorectal, in study 5a and study 9). Five studies focused on genitourinary and gynecologic cancer (renal cell carcinoma, in study 2; prostate, in study 5b; epithelial ovarian cancer in study 6; germ cell tumours in study 8; epithelial ovarian, peritoneal, or fallopian tube cancer, in study 10). Lung cancer was studied in three (13%) cases (3, 15, and 16), and breast cancer in the two remaining studies (5c and 11).

The type of cancer intervention was reported in all studies but one (study 16). The most common interventions were chemotherapy, radiotherapy and surgery, with each intervention named in 18 (78%) studies. In 14 studies, all three interventions were reported. Immunotherapy was only reported in study 2, with four patients receiving the treatment). Other interventions (such as homeopathy) were also mentioned in six studies (2, 5a, 5b, 5c, 11, and 17).

Eleven (48%) studies reported a population mean age below 65 (the minimum mean age was 44 in study 8).[3]

LoS is not consistently defined across the different papers. Most of the studies (18; 78%) define LoS as number of years since diagnosis.[4] Two studies (9%) referred to LoS as number of years since the patient was diagnosed with metastasis (2 and 11). In three studies (13%) the authors describe LoS as the number of years since treatment

[3] Some age means were not directly reported in the study and have been estimated. *Estimates:* 55 in study 19; 64 in study 5c; 66 in studies 9 and 11; 70 in studies 1a, 1b and 16; 72 in study 5b; 73 in study 5a. *Methods:* For studies 1a and 1b: assumed homogeneous split by cancer type for each age group (applying frequency weights: 63% to study 1a and 37% to study 1b). For mean age, assumed reference age: 18-54: 40, 55-64: 60, 65-74: 70, 75+: 80. For study 5a: typo in Table 2, age in years. 36 patients are inputted to the category >=80 (blank), to match the total and % with that provided elsewhere in the paper. For studies 5a, 5b, and p5c: assumed homogeneous split by cancer type for each age group (age category 60-69 split 50/50 between <65 and 65+). For mean age, assumed reference ages: 40-49: 44.5, 50-59: 54.5, 60-69: 64.5, 70-79: 74.5, 80+: 85. For studies 10, 11, 15, and 20: mean assumed to be equal to (reported) median. For study 9: reference age 40 and 75.

[4] Study 6: the definitions of long- and short-term survivors are not consistent throughout the paper. Short-term survivors are referred to as "Patients who died within 5 years post-diagnosis", or "those who deceased within 5 years post-treatment". Long-term survivors were referred to as "Patients who were alive at least 10 years after their initial diagnosis", "patients who were free of disease more than 10 years after diagnosis" or "alive 10 years post-treatment". In this report, we have considered the definition related to "post-diagnosis", as we consider it to represent the most logical choice based on the authors' recruitment criteria.

completion (4, 14, and 18). The average LoS was extracted (or estimated) from the different studies and is shown in Table 2. Eighteen studies (78%) had an average LoS of over 5 years.

A total of 31 different patient-reported outcome measures (PROMs) were used in the studies. The most frequently used measures were the Short-Form (36 or 8) Health Survey (SF-36 or SF-8)[5] and the European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30), which were used in 10 (43%) studies. The two instruments were used together in studies 1a, 1b, and 7. The head-and-neck-specific version of the EORTC QLQ-C30 (the EORTC QLQ-H&N35 Head and Neck 35 cancer module) was used in four studies (7, 12, 14, and 19). The Functional Assessment of Cancer Therapy measures (the generic QoL measure FACT-G, as well as condition-specific versions FACT-Kidney, FACT-Head and Neck, FACT-O, FACT-Breast) were used in four studies (2, 4, 10, and 11). No preference-based utility value was mentioned in any of the papers, even though the SF-36 has a well-established set of utility values linked to the questionnaire (SF-6D).

Nine studies (35%) investigated the evolution of patient QoL over time. Four studies (17%) found that the overall QoL of survivors improved over time (1b, 6, 16, and 20). Three studies concluded that QoL remained constant (12, 18, and 19), while one study (17) reported mixed results.

Eleven studies compared QoL reported by LTCS with population norms or averages from samples of the general population. Based on our own judgement, the different studies have been classified depending on the overall result of the comparison. The categories are summarised in Table 3. We judge four studies (17%) to provide evidence that patients' QoL was better than that of the general population (5a-5c and 18). Study 10 refers to better or similar QoL in both groups, and study 1a shows similar QoL. Studies 1b, 7, and 8 conclude that patients report QoL which is similar to or worse than that reported by the general population. Only study 15 reported that the QoL of LTCS was definitively worse overall than that of the population norm. Study 6 reported mixed results. In the following section, the results of these eleven studies are analysed in more detail.

3.3. Comparison of QoL in LTCS and population norms

In this section, we explore in more detail various characteristics of the studies that compared the QoL in LTCS with general population norms, paying special attention to the following:

- What kind of population sample is used as control
- Whether the samples are adjusted (by gender, age and/or additional factors)
- The use of statistical tools to test the significance of the results
- Whether the differences are clinically significant
- Whether the comparison is made for every item or on the aggregate score level
- Overall strengths and weaknesses of each study

Table 4 provides an overview of the studies reviewed in this section.

[5] SF-8 was used in studies 8 and 16.

Table 4. Summary of the studies that make comparisons with a control group (n=11 from n=8 papers)

No.	Paper	Cancer	N	Mean LoS*	Mean age	% Late stage	Control	Adjusted?	Statistical significance?
1a	Caravati-Jouvenceaux et al. (2011)	GI	344	*13.6*	*70*	23%	No prior cancer	Yes	Yes
1b		GI	198	*13.6*	*70*	17%	No prior cancer	Yes	Yes
5a	Ferro et al. (2014)	GI	134	7.5	*73*	25%	General population	Yes	No
5b		GG	105	7.5	*72*	17%	General population	Yes	No
5c		B	344	7.5	*64*	8%	General population	Yes	No
6	Greimel et al. (2011)	GG	11	10	56	*73%*	No prior cancer	Yes	No
7	Hammerlid et al. (2001)	H&N	135	3	62	41%	General population	Yes	Yes
8	Hartung et al. (2016)	GG	164	11.6	44	48%	General population	Yes	Yes
10	Lutgendorf et al. (2017)	GG	56	*14*	66	98%	General population	N/R	No
15	Phipps et al. (2008)	GI	30	*7*	69	23%	General population	Yes	No
18	Vainshtein et al. (2015)	H&N	40	6.5	63	100%	General population	N/R	No

In italics: estimated (not reported in the paper)
N/R: not reported; LoS: Length of survival since diagnosis; GI: gastrointestinal; GG: genitourinary and gynecologic; B: breast; H&N: head and neck
*: Length of survival since diagnosis (except for paper 18, where length of survival is since treatment completion)
% Late stage: % of patients at stages III or IV at diagnosis, over the total sample
Control: group for comparison
Adjusted: whether the mean has been adjusted at least by one factor (age or sex)
Statistical significance: whether the authors use statistical significance criteria

A total of 11 studies (extracted from eight papers) compared the QoL of LTCS and a representative control group. This control group is usually the general population, but in three studies (1a, 1b, and 6) only individuals with no prior cancer diagnosis were considered. The comparison was adjusted by at least one factor in all but two studies (10 and 18), where no adjustment is reported. Only four studies (1a, 1b, 7, and 8) included the term "statistically significant" in the analysis.

A description of the key variables in these papers (similar to the description provided in Table 3) is shown in Table 5. Only two studies (1a and 1b) were population-based. Studies 6, 7, and 18 are longitudinal, and most of the studies (n=9; 82%) were published recently (in or after 2010). All studies include chemotherapy as a cancer treatment used by part of the sample, and radiotherapy is mentioned in all studies except study 6. No paper focusing on lung cancer is included in this group. All the studies use diagnosis as a reference point for measuring LoS, except for study 18 where the number of years since treatment completion is counted.

Nine of the 11 studies used SF-36 or SF-8 to measure the QoL of survivors, and five studies use EORTC QLQ-C30. This similarity allows for a comparison of results across papers, down to item level when possible. In this way, it can be seen whether differences between patient and population groups are systematic across the same items or dimensions (such as physical health vs mental health, or general vs. condition-specific factors). Since study 10 does not report SF-36 or EORTC QlQ-C30, results on FACT-G are also included in the table. Items in the three measures are scored from 0 to 100, with higher values indicating better QoL.

Table 6 provides an overall summary of the comparisons. In each cell, a symbol of +, = or – indicates whether the average QoL of LTCS is better than, similar to or worse than the QoL of the comparison group. Studies 1a and 1b have been split into three columns, describing the results for 5-year, 10-year and 15-year LTCS. In the table, we indicate if the result shown in each cell is statistically significant and if it is clinically significant (using a difference of 10 points or greater, as suggested in Osoba et al., 1998)[6].

Overall, studies that apply statistical significance criteria show that the differences between LTCS and population QoL means are not statistically different for most of the items; and any statistically significant differences are rarely clinically significant. In one case (study 1b), the dimension of "social functioning" showed similar results when using SF-36 but significantly worse results for 5-year and 10-year LTCS when using the equivalent dimension from the EORTC QLQ-C30.

The individual studies are described in more detail in the next sections.

3.3.1. Caravati-Jouvenceaux et al. (2011) - Study A

A sample of 344 colon cancer survivors ("considered as cured") was drawn from population-based cancer registries. Three patient groups were defined, corresponding to three survival periods: 5 (4-6), 10 (9-11), and 15 (14-16) years after diagnosis (group size not reported in the paper). Individuals who had experienced relapse, metastasis, or

[6] Studies 6 and 18 use the threshold of 10 points to denote clinically significant differences. Studies 1 and 7 use 10 points to denote the existence of a moderate or large difference (differences between 5 and 10 points are also considered as clinically significant but not clinically meaningful).

another type of cancer or undergone treatment in the previous 5 years were excluded from the sample.

The study compares patients' QoL with that of a sample of the general population who had no prior history of cancer (n=768) using the EORTC QLQ-C30, the SF-36 and the MFI. The comparison with the population sample was adjusted by age group, gender, marital status, living alone, level of education, employment status, income, comorbid conditions and length of hospital stay.

The authors state that:

> *"In colon cancer survivors, we observed clinically small but statistically significantly higher scores on the diarrhea scale (QLQ-C30) than in controls at 5 years and 10 years after diagnosis (data not shown). There were no clinically significant differences in other QOL scores between colon cancer survivors and controls at any time point after diagnosis."* (p.1631)

Statistically significant differences are only reported for the QLQ-C30 item of "*diarrhoea*"; they are not reported in any other item of the PROs included in this study. Data regarding the results were not shown.

3.3.2. Caravati-Jouvenceaux et al. (2011) - Study B

A sample of 198 rectal cancer survivors ("considered as cured") was drawn from population-based cancer registries. Three patient groups were defined, corresponding to three survival periods: 5 (4-6), 10 (9-11), and 15 (14-16) years after diagnosis (group size not reported on the paper). Individuals who had experienced relapse, metastasis, or another type of cancer or undergone treatment in the previous 5 years were excluded from the sample.

The study compares the QoL measures with respect to a sample of the general population who had no prior history of cancer (n=413). The comparison was adjusted by age group, gender, marital status, living alone, level of education, employment status, income, comorbid conditions and length of hospital stay.

Detailed results are reported. 5-year survivors report statistically significant lower scores in physical functioning (SF-36), and social functioning and diarrhoea (EORTC QLQ-C30). 10-year survivors report lower scores in vitality (SF-36) and role functioning, social functioning, fatigue, constipation and diarrhoea. 15-year survivors only report (statistically significant) worse scores in the item diarrhoea (EORTC QLQ-C30). No statistically significant differences are found in other items and survivor groups.

The results suggest that only one group of LTCS (15-year survivors) reach levels of QoL that are comparable with those of the general population, with most of the QoL items not showing significant differences when comparing survivors and the general population (though some cancer- or treatment-related complications may persist over time, as diarrhoea was a robust finding). The authors fail to report the exact number of patients surviving at 15 years (we only know there are a total of 99 patients for both colon and rectal, but not how they are distributed between the two cancer locations). Pooling 10- and 15-year survivors may have shown stronger results, but this was not done by the authors.

3.3.3. Ferro et al. (2014) - All Three Studies

In these studies, the patient sample was identified from the "Hospital Discharge Minimum Basic Data Set" in four hospitals in Spain. The authors collect data on baseline

variables, but QoL data were only collected from LTCS. The comparison with the general population is not described in detail in the paper, in the sense that the authors did not clarify which data they were using, if patient and control groups were matched, or if they used any test to assess the statistical significance of any differences. However, the authors state that:

> *"In any case, it is worth underlining the overall high quality of life reported by these patients in comparison to the general Spanish population aged 60 years or older. Indeed, the quality-of-life scores of the survivors in our study exceeded the values in all dimensions as well as in the mental component of the SF-36 scale, while physical wellbeing was similar between the two groups"* (p.134)

Thus, we assume the authors compared their patient data to some existing SF-36 data for the Spanish general population, referenced in the paper. The comparison to the population aged 60 or above seems logical, since the estimated average ages of those surviving colorectal, prostate and breast cancer are 73, 72 and 64 years, respectively.

3.3.4. Greimel et al. (2011)

This is a longitudinal study where QoL data were collected pre-treatment (baseline), 1 year post-diagnosis and 10 years post-treatment, from a sample of 33 patients (drawn from 50 consecutive patients at the Department of Gynaecology at the Medical University, Austria). This is one of the few longitudinal studies found, but unfortunately the sample size for LTCS was only 11.

The QoL of LTCS was compared to that of females from a general population survey with-out a history of cancer (n=1139). The comparison is reported in the paper by means of a bar graph, and exact numbers of the normative reference data are not shown. The authors report that:

> *"Long-term ovarian cancer patients had comparable QoL scores in all functioning scales and symptom scales, except dyspnoea. Survivors had significantly higher scores on the dyspnoea scale (10 points higher than in the reference sample)."* (p.3009)

However, the authors use the criteria of clinical significance and it is not clear if any test of statistical significance was used to support the conclusions.

Table 5. Distribution of the key variables in the studies that make comparisons with a control group

Variable	Freq.	%
Publication type		
Journal article	11	100%
Study type (1)		
Cross sectional	8	73%
Prospective	0	0%
Retrospective	8	73%
Longitudinal	3	27%
Study type (2)		
Population-based	2	18%
Centre-based	9	82%
Year of publication		
Before 2010	2	18%
In 2010 or after	9	82%
Country		
Europe	8	73%
USA	3	27%
Other	0	0%
Sample size		
≤50	3	27%
51-100	1	9%
101-500	7	64%
>500	0	0%
Average sample size by cancer stage		
Stage I+II	88	
Stage III+IV	36	
Stage unknown	16	
More than half of the patients in stage III+IV	3	27%
Cancer type		
Breast	1	9%
Gastrointestinal	4	36%
Genitourinary and gynaecologic	4	36%
Head and neck	2	18%
Lung	0	0%

Variable	Freq.	%
Average length of survival (years)		
<5	1	9%
5.1-10	6	55%
10+	4	36%
Unknown	0	0%
Inclusion of PROMs		
EORTC QLQ-C30	4	36%
EORTC QLQ-H&N35	1	9%
SF-36/SF-8	9	82%
FACT-G	1	9%
FACT-Kidney (FKSI-15)	1	9%
FACT–Head and Neck	0	0%
FACT-O	1	9%
FACT-Breast	0	0%
LASA	0	0%
UWQOL	1	9%
NHP	3	27%
CES-D	1	9%
Other	5	45%
Examination of QoL at different points in time		
Increasing over time	3	27%
Constant over time	1	9%
Decreasing over time	0	0%
Mixed evidence	0	0%
No comparison made	7	64%
Comparison with population norms		
Better than general population	4	36%
Better than or similar to general population	1	9%
Similar to general population	1	9%
Similar to or worse than general population	3	27%
Worse than general population	1	9%
Mixed evidence	1	9%
No comparison made	0	0%

Variable	Freq.	%
Intervention		
Chemotherapy	11	100%
Radiotherapy	9	82%
Surgery	10	91%
Immunotherapy	0	0%
Other	1	9%
Unknown	1	9%
Not reported	0	0%
Population mean age		
18-65	5	45%
65+	6	55%
Definition of "length of survival"		
Years since diagnosis	10	91%
Years since diagnosis with metastasis	0	0%
Years since treatment completion	1	9%

Variable	Freq.	%
Suggested explanatory factors of the results		
Response shift	5	45%
Borderline significance	2	18%
Selection bias	4	36%
Adaptation	1	9%
Long-term complications or sequelae		
Bowel problems	3	27%
Fatigue	3	27%
Social relations problems	3	27%
Sexual functioning	3	27%
Swallowing	2	18%
Other	3	27%

QoL: quality of life; EORTC QLQ-C30: European Organization for the Research and Treatment of Cancer Quality of Life Questionnaire; EORTC QLQ-H&N35: Head and Neck 35 cancer module; SF-36: Short form (36) health survey; SF-36: Short form (8) health survey; FACT-G: Functional Assessment of Cancer Therapy-General; FACT-Kidney: Functional Assessment of Cancer Therapy–Kidney Symptom Index; FACT-Head and Neck: Functional Assessment of Cancer Therapy Additional Concerns for Head and Neck; FACT-O: Functional Assessment of Cancer Therapy-Ovarian; FACT-Breast: Functional Assessment of Cancer Therapy-Breast; LASA: Linear analogue self-assessments; UWQOL: University of Washington Quality of Life scale; NHP: Nottingham Health Profile; CES-D: Center for Epidemiological Studies Depression scale.

3.3.5. Hammerlid et al. (2001)

This study compared EORTC QLQ-C30, EORTC QLQ-H&N35 and SF-36 outcomes for 232 individuals who were diagnosed with head and neck cancer in a hospital in Sweden during 1993–95 with population norm scores on the same instrument. The study was longitudinal, with QoL data collected at baseline and 3 years after diagnosis (only 135 patients contacted at that point). However, no comparison of QoL scores over time (3-year compared to baseline) is reported at the paper. Instead, the authors compare QoL outcomes of the 3-year survivors with the QoL of (age- and gender-matched) general population. However, note that the reference used for the EORTC QLQ-C30 is a Norwegian population norm.

The results are mixed. On one hand, seven of the eight SF-36 health domains were not significantly different between cancer survivors and the Swedish SF-36 normative database reference values, suggesting that both populations report similar values in measures of general health status. On the other hand, scores reported in cancer-related limitations or problems (such as swallowing, dry mouth or mucus production) are significantly worse for LTCS than for those reported by the general (Norwegian) population. The authors note concerns about the differences between generic and condition-specific measures. One reading of these results could be that, although the cancer patients are more bothered by specific cancer- or treatment-related long term issues, these do not translate into any large effect on their overall (generic) QoL.

The paper explores the relation between LTCS and population norms by sex and age group (<65/65+). Female LTCS (n=42) scored the same as or better than the reference group on each of the SF-36 items, though the difference was only significant only for vitality. In contrast, male LTCS (n=93) reported statistically significant worse scores in four of the scales (role-physical functioning, role-emotional functioning, physical functioning and general health). Age was not found to affect the comparison of SF-36 scores.

The sample size of the study is reasonable, and the methodology used by the authors appears to be robust, the major limitation being the use of a foreign population norm in the case of the EORTC QLQ-C30 measure.

3.3.6. Hartung et al. (2016)

A total of 164 patients were enrolled from the outpatient department of a hospital in Germany. The LoS of those patients ranged from 1 to 35 years after diagnosis (mean=11.6 years). Results for the cancer survivors on SF-8 were compared with an age-adjusted population norm. The comparison showed mixed results. On one hand, there were no statistically significant differences between groups on the physical component summary (SF-8) or on five dimensions of the questionnaire (physical functioning, role limitation resulting from physical problems, bodily pain, role limitation resulting from emotional problems and social functioning). However, patients reported statistically significantly poorer scores on the mental health, vitality, and general health perception dimensions and items and on the mental component summary score. These differences were mainly observed in the 21 to 50-year age group, though that could have been partly the result of small sample sizes in the other age groups.

Table 6. Comparison of scores on individual PROM dimensions between patients and the general population

PROMs, Dimensions/items		Studies												
		1a			1b			5abc	6	7	8	10	15	18
		5y	10y	15y	5y	10y	15y							
Total	**FACT-G**											+		
	SF-36/SF-8													
Physical health	Physical functioning	=	=	=	-*	=*	=*	+		=*	=*		-^	+
	Role limitation resulting from physical problems	=	=	=	=*	=*	=*	+		-*^	=*		-^	+
	Bodily pain	=	=	=	=*	=*	=*	+		=*	=*		-^	+
	General health perception	=	=	=	=*	=*	=*	+		=*	-*		-^	+
	Physical component summary	=	=	=	-*	=*	=*	=			=*			
Mental health	Vitality	=	=	=	=*	-*	=*	+		=*	-*		-	+
	Social functioning	=	=	=	=*	=*	=*	+		=*	=*		-^	+
	Role limitation resulting from emotional problems	=	=	=	=*	=*	=*	+		=*	=*		-^	+^
	Mental health	=	=	=	=*	=*	=*	+		=*	-*		-^	+^
	Mental component summary	=	=	=	=*	=*	=*	+			-*			
	EORTC QLQ-C30													
Global	**Global health status**	=	=	=	=*	=*	=*		+					
Functioning scales	Physical functioning	=	=	=	=*	=*	=*		+					
	Role Functioning	=	=	=	=*	-*	=*		+					
	Emotional functioning	=	=	=	=*	=*	=*		+					
	Cognitive functioning	=	=	=	=*	=*	=*		+					
	Social functioning	=	=	=	-*^	-*^	=*		-					
C30 items	Fatigue	=	=	=	=*	-*	=*		+					
	Nausea and vomiting	=	=	=	=*	=*	=*		+					
	Pain	=	=	=	=*	=*	=*		+	-*^				
	Dyspnoea	=	=	=	=*	=*	=*		-^					
	Insomnia	=	=	=	=*	=*	=*		+					
	Appetite loss	=	=	=	=*	=*	=*		+					
	Constipation	=	=	=	=*	-*^	=*		+					
	Diarrhoea	-*	-*	=*	-*^	-*^	-*^		+					
	Financial difficulties	=	=	=	=*	=*	=*		+					
Other (condition-specific) items	Swallowing					-*								
	Senses								-*^					
	Social eating								-*^					
	Problems with teeth								-*^					
	Opening mouth wide								-*^					
	Dry mouth								-*^					
	Mucus production								-*^					
	Cough								=*					
	Feeling ill								=*					

Every item is scored between 0 and 100. ^: differences clinically significant (10 or more points); *: test of statistical significance applied (95% confidence); - if worse score for LTCS, = if no differences between LTCS and controls, + if better scores for LTCS; 5y, 10y, 15y: LTCS 5 years, 10 years, 15 years after diagnosis. SF-8 only reported in study 8.

3.3.7. Lutgendorf et al. (2017)

The 56 participants in this study were recruited from five academic medical centres and the Ovarian Cancer Research Fund Alliance. Patients were eligible if they were at least 8.5 years from diagnosis with stage III (89%) or stage IV (9%) epithelial ovarian cancer, providing a good sample of LTCS in a late stage of cancer.

The authors collected QoL data from this set of LTCS using several measures. The authors only compare the results from the non-site-specific measure FACT-G to population norms, and provide few details of the analysis. A mean QoL of 86.7 (SD approximately 11) was estimated for FACT-G, and the authors report that this figure approximates *"(…) normative FACT-G scores of the U.S. population (80.1 ± 18.1)"* (p.104).

3.3.8. Phipps et al. (2008)

Participants in this study were colon cancer patients who were disease-free at least 5 years since diagnosis. The 30 cancer LTCS were identified through the Albert Einstein Medical Center Tumor Registry in USA and through doctor referral. The median age of the LTCS patients was 69, 50% were white and 50% African American, 50% male, and all had a previous curative resection.

The authors found that "*across all SF-36 health dimensions, colon cancer survivors in our study had substantially worse functional status than individuals of similar ages in the general population*" (p.255). Mean published SF-36 scores of the US population aged 55-64 years and 65-74 years were used for the comparison. The authors do not clarify if the means were adjusted by sex or race, but from the way the results are presented (a figure with mean results by subscale), it seems unlikely that any adjustment was made (except for age). The authors do not show how comparable the population under analysis is with respect to the broader patient population, as they acknowledge ("*our sample of 30 US survivors may not be representative of all colon cancer survivors*"; p.258). The small sample size (n=30) is an additional weakness of the study.

3.3.9. Vainshtein et al. (2015)

A sample of 40 LTCS patients with stages III or IV at diagnosis of head and neck cancer was drawn from a prospective study. Several QoL instruments (including SF-36) were administered on average 6.5 years after treatment completion.

The mean score reported by LTCS was higher than the population mean on all SF-36 dimensions, and in two categories (role limitation-emotional and mental health) the difference between means was more than 10 points. However, the authors do not specify whether the population means were adjusted to the characteristics of the LTCS, and they do not report statistical testing to corroborate their results. The authors state that "*long-term overall physical and mental health mean scores for the cohort were comparable in each HRQOL domain to the US population norms"* (p.927).

3.4. Exploring the evolution of QoL in LTCS over time

In this section, we analyse in more detail the papers that investigated the evolution of patient QoL over time (the time points are usually among the following: baseline, 1 year, 3 years, 5 years, 10 years and/or 15 years after diagnosis).

3.4.1. Caravati-Jouvenceaux et al. (2011) Study B

The authors conclude (though the results are not that clear at this point and only partially reported) that the QoL of the survivors increases over time. In particular, 15-year survivors report better QoL than 10-year survivors on most instrument dimensions. However, the significance of the differences over time was not tested using any statistical method. In addition, this was not a longitudinal study, and thus the three groups of survivors (5-, 10- and 15-year) were drawn from three different cohorts. Finally, the response rate of the 15-year group was the lowest among the three groups (27%), so that the most impaired survivors may not have been represented in the sample. For all these reasons, we do not consider this to be a robust result.

3.4.2. Greimel et al. (2011)

The longitudinal structure of the data allows the authors to investigate the evolution of QoL scores at the baseline, 1 year and 10 years after diagnosis. Statistical significance is applied here. The comparison of average QoL at baseline and 10-year follow-up shows a better QoL for LTCS (except with respect to dyspnoea, diarrhoea and financial difficulties). The comparison of QoL at 1 year and 10 years after diagnosis still shows in general better QoL for LTCS, but with more exceptions, such as the fatigue, pain or appetite loss items, as well as a statistically significantly lower reported score for global health status (at 10 years compared to 1 year). In general, it could be said that the overall QoL of a LTCS appears to increase over time. The methodology used for this analysis seems more robust than in other papers (e.g. the same individuals form all the time cohorts, and Bonferroni corrections are applied when testing the differences); however, the very small sample size (n=11) definitely means the results should be interpreted with caution.

3.4.3. Hartung et al. (2016)

This paper does not explore directly the evolution of patient QoL over time, but the authors run a multivariate regression model of the physical and mental component scores (SF-8 summary measures) of patients as a function of LoS (among other variables). For the physical component score, the reported coefficient is positive (0.17) and significant (95% confidence interval [0.02-0.34]), suggesting that an additional year of survival may increase the physical component score of the SF-8 measure by 0.17 points. LoS is not a significant explanatory factor for the mental component score (coefficient not reported).

3.4.4. Oskam et al. (2013)

The objective of this study was to evaluate changes in HRQoL from baseline (at time of diagnosis) and short term (6 and 12 months) follow-up to long-term follow-up (8–11 years) in a homogeneous sample of head and neck cancer patients. This is a prospective, longitudinal study, with a sample of 27 patients.

The authors compare mean scores of the EORTC QLQ-C30 and EORTC QLQ-H&N35 subscales of the same patients at baseline, 6 months, 12 months and long-term (8–11 years) follow up. Significant differences in social functioning (EORTC QLQ-C30) and social contact, speech, dry mouth, sticky saliva, cough (EORTC QLQ-H&N35) were found between LTCS and baseline, resulting in lower averages for the group of LTCS. Significant differences are also reported in the EORTC QLQ-H&N35 item "pain", reflecting better (less pain) scores for LTCS. No significant differences were found in the other 14+7 items.

The authors remark that at long-term follow-up, the need for supportive care was lower than at the baseline, and was limited to a dental hygienist and a physical therapist. However, note that the group of LTCS is of small size (n=26) so the results should be interpreted with caution.

3.4.5. Nes et al. (2012)

This is a prospective patient follow-up study which examines the effect of a change in the physical activity level on the QoL of LTCS. Patients are divided into three groups, according to time since diagnosis: <3 years (n=714), 3-5 years (n=426), >5 years (n=797). The sample size is relatively large for this type of study (1937 participants), and the authors use tests of statistical significance to add robustness to the results. The main drawback of this paper is the fact that the results are reported separately for the two groups of analysis: those who reported a decreased physical activity level from baseline, and those who reported increased levels. The main results of the study were:

> *"(…) patients reporting decreased physical activity level from baseline to follow-up reported a significant decrease in mental ($p < .001$), physical ($p < .001$), emotional ($p < .001$), social ($p < .001$), and spiritual ($p < .001$) well-being, as well as in overall QOL ($p < .001$)"* (p.613)

> *"Patients reporting increased physical activity from baseline to follow-up, in contrast, reported increase in their mental ($p = .005$), physical ($p = .003$), emotional ($p = .02$), social ($p < .001$), and spiritual ($p = .02$) well-being, as well as in overall QOL ($p < .001$)."* (p.613)

3.4.6. Vainshtein et al. (2015)

A sample of 40 LTCS patients with stages III or IV of head and neck cancer at diagnosis was drawn from a prospective longitudinal study. Several QoL instruments (SF-36 among them) were collected before treatment and at 1, 3, 6, 12, 18, and 24 months after chemo-radiation therapy, and mean follow-up after treatment completion was 6.5 years.

The authors show that the average QoL of LTCS remained stable (compared with before treatment) using the HNQoL summary score, and it was statistically, but not clinically meaningfully, worse by UWQOL summary score. A further finding was that most of the condition-specific symptoms remained stable or continued to improve 2 years after treatment completion.

3.4.7. van der Schroeff et al. (2007)

This longitudinal, prospective study describes the results from a 3 to 6-year follow-up in a cohort of 57 LTCS, recruited at the University Medical Centre of Utrecht in the Netherlands. The patients enrolled in the study had newly diagnosed squamous cell carcinoma without distant metastasis at baseline. The results are described by age group: 45-60 versus 70+ years.

The authors investigated changes in the reported QoL of patients at baseline, 12month and 3-6-year follow-up, and found that effects varied across different measures and groups. In general, the majority of the QLQ-C30 scores were not found to differ significantly between the two age groups. A pattern of poorer physical functioning in elderly patients was observed; in particular, the older patients scored significantly worse on swallowing and speech compared to baseline. The authors do not compare the results with a sample of (age-adjusted) general population.

3.4.8. Vartanian et al. (2009)

This study is part of a longitudinal prospective study; however, the authors only report the results corresponding to patients' responses to a follow-up questionnaire. The study focuses on a cohort of patients in Brazil, with the objective being to evaluate the "acceptance of treatment and the quality of life of long-term survivors of advanced head and neck cancer who had undergone major surgical procedures" (p. 376).

The authors state that 74% of the patients contacted (possibly not all of them LTCS) reported that their health status was the same as or better than that before the treatment.

4. DISCUSSION

4.1. Summary of the findings

Twenty-three studies (from 20 papers) were included in this review. The studies of most interest in relation to the aims of this review are those that compared QoL in LTCS with QoL in control groups. Of the included studies, just under half (n=11; 48%) included some form of comparison, but the type of comparator differed between the studies. Six of the studies used a representative general population sample as a comparator and adjusted by at least one factor (e.g. age or sex) in order to make comparisons. A further two studies used a general population sample but did not report any adjustments. The remaining three studies all adjusted their control group by at least one factor but rather than recruiting a representative general population sample they recruited a sample of individuals that had not experienced cancer. We consider the use of matched general population comparators to be the most rigorous approach to making control group comparisons. Such an approach was used in just over half of the studies reporting comparisons of LTCS and control groups.

The comparisons provided a range of different results in relation to the research question; however, overall, the QoL of LTCS was better than or similar to the QoL of the comparator group in six of the 11 studies that made comparisons. Taking Table 6 into consideration, four of the studies generated evidence that suggests that QoL in LTCS was better than that of the control group (studies 5a-5c and 18), with one suggesting that QoL in LTCS was similar or better (study 10). One study suggested that QoL in LTCS was similar to the control group (study 1a), with another providing mixed evidence (study 6). Three studies generated evidence to suggest that QoL in LTCS was similar or worse compared to the control group (studies 1b, 7 and 8), with only one study providing evidence that QoL was worse for LTCS (study 15).

Very few of the 11 studies used statistical significance criteria, which is a concern. Of the studies that *did* apply statistical significance criteria (n=4; 36%), it appears that the differences between LTCS and control group QoL means are not statistically different for most of the QoL items. However, on some domains explored in these four studies, QoL is statistically significantly worse for LTCS than the comparator group. In some cases the differences may also be clinically significant (study 7 in particular). In addition, there is no instance of LTCS having statistically significantly better QoL on a single domain in any of these four studies. Therefore, whilst this review provides some evidence to suggest that QoL in LTCS is similar to that of the general population, it does not provide definitive evidence and it is clear that there is variation across different domains of QoL (as illustrated by Table 6).

Three studies report results about the correlation between QoL instruments and cancer stage at diagnosis. Study 7 compared SF-36 results for early (I+II) and late (III+IV) stage head and neck cancer patients three years after diagnosis. The authors found a few clinically significant differences (early stage patients scored higher on the bodily pain and mental health scales), but none were statistically significant. A linear regression analysis in study 8 found that patients in later stages are more likely to report higher values in the physical component score in the SF-8 instrument. No statistically significant results could be established for the mental component score. Study 15 did not find stage of colon cancer at diagnosis to be correlated with any dimension of the QOL-CS score.

There were four studies in which only patients in later cancer stages were included in the sample (studies 10, 14, 18 and 20). Study 14 shows good QoL for head and neck LTCS who did not report delayed complications (more than 5 years after treatment). About 60% of the head and neck cancer patients contacted in study 20 reported good or excellent global QoL, with patients in stage IV reporting worse QoL than those in stage III. Study 10 (50 epithelial ovarian cancer patients in stage III and 5 in stage IV at diagnosis) and study 18 (6 head and neck cancer patients in stage III and 34 in stage IV at diagnosis) reported better or similar QoL for these LTCS compared to the general population. However, neither study reported statistical testing nor provided details of how the comparisons with population norms was established (see sections 3.3.7. and 3.3.9. for further details). Further, these studies did not report separate QoL results for stage III and stage IV, presumably due to the small numbers of patients involved. Across all the studies included in this review, only study 9 (colorectal cancer) directly compares patients in Stage IV at time of diagnosis with those at early stages. The authors suggest that patients in Stage IV at time of diagnosis are more likely to experience moderate-to-high *benefit finding* and *post-traumatic growth* (two constructs of positive consequences of cancer) at 5 years after diagnosis, compared to early stages. However, these results are not statistically significant. No comparison is established in terms of QoL.

4.2. Possible explanations for the results

It may seem counterintuitive to observe that QoL in LTCS is equivalent to the QoL of the general population. There are several possible explanations, beyond that of LTCS simply having a good level of QoL, which could be put forward to explain these results. These explanations include: adaptation; lifestyle improvements; selection bias; and data limitations. This section provides a discussion around each of these possible explanations, drawing upon evidence and discussion from across all of the 23 included studies where possible (i.e. not only those that made control group comparisons).

4.2.1. Adaptation

Several of the studies included in this review discuss the concepts of "adaptation" (n=2) and "response shift" (n=5). Adaptation refers to the idea that cancer patients, or patients more generally, adapt to their conditions over time. This phenomenon has been observed in the health state valuation literature and is explored in numerous studies (Ubel et al., 2003; Cubí-Mollá et al., 2017; Ogorevc et al., 2017). Response shift is a potentially concerning component of adaptation. This involves individuals changing their internal standards, their values and their conceptualisation of QoL (Sprangers & Schwartz, 1999). In other words, LTCS might not rate their QoL on the same scale as individuals from the general population, potentially casting doubt on the legitimacy of QoL measurement in patients. However, it has been suggested that response shift is one

of many possible mechanisms that could influence adaptation (Stiggelbout & de Vogel-Voogt, 2008). It may well be the case that adaptation is a legitimate response to experience and that individuals' judgements about their QoL changes once they are better informed. Therefore, if adaptation is in some way responsible for these findings, this is not necessarily a problem. Indeed, adaptation arguably has a persistent influence in cost-utility analyses. Whilst health state utilities from the general population are recommended by agencies such as NICE, the questionnaires themselves are still filled out by patients that might have adapted to their conditions (Versteegh & Brouwer, 2016).

4.2.2. Lifestyle improvements

It could also be the case that individuals that have been diagnosed with cancer subsequently adjust their lifestyles in a positive manner as a result of their health shock (Williams et al., 2013). Therefore, the high levels of QoL that are observed in LTCS might be explained, to some extent, by lifestyle improvements rather than cancer treatment itself. Some of the included studies explored whether the LTCS had undergone a lifestyle improvement.

Three studies provided evidence to suggest that LTCS had undergone at least one lifestyle change. Study 3 found that 77% of the LTCS that had been diagnosed with lung cancer were former smokers, having quit after diagnosis. Study 11 found that women were more likely to engage in "alternative medicine" (which includes "megavitamins", massage, relaxation therapy and spiritual therapy) use after diagnosis of breast cancer than before the diagnosis. Finally, study 12 reported that two patients continued to use a dietician and two patients continued to use alcohol cessation counselling services (previously provided as part of treatment) after treatment. In contrast, it should also be noted that other studies found that the majority of LTCS did not report any lifestyle improvements. Studies 5a-5c found that most LTCS reported no changes: 67% of colorectal cancer survivors; 77% of prostate cancer survivors and; 75% of breast cancer survivors. Interestingly, these studies all reported that QoL was better in LTCS relative to the control groups. In addition, study 17 noted that most survivors (n=1681; 87%) reported having a sedentary lifestyle at both time points in the study.

None of the studies provided sufficient evidence to explore the potential effect of lifestyle changes of LTCS on QoL. It could be argued that this is an important area for further research as cancer survivors may be more likely to change their lifestyle (to a healthier one) than the general population, and this lifestyle change could have a direct impact on the QoL of the LTCS. However, the evidence from the included studies is clearly mixed and no clear conclusions can be drawn. It may be helpful for pharmaceutical companies to collect more data on patients' changes in lifestyle alongside their reported QoL outcomes in order to better understand how the former affects the latter (acknowledging that some individuals may engage in lifestyle changes for reasons unrelated to their cancer).

4.2.3. Selection and publication bias

Alternatively, a common source of bias may have influenced these findings. "Selection bias" occurs when a sample is recruited in such a way that a reasonably representative sample of patients is not achieved. In this context, it might be the case that younger or healthier LTCS are more likely to have responded to the studies relative to older or unhealthier LTCS. However, this is a common, and largely unavoidable, source of bias in

health-related observational studies. In addition, selection bias is cited in several of the papers and some attempted to explore the issue. For example:

> *"Participants and nonparticipants did not differ in age, years since diagnosis, and presence of metastases at first diagnosis."* (Hartung et al., 2016, p.60e2)

> *"Among the patients who did not respond to the questionnaire, tumor stage at diagnosis and treatment types were similar to those observed in participants. Non-participants also experienced the same kinds of adverse effects for each tumor, though less frequently."* (Ferro et al., 2014, p.130)

Another form of bias, publication bias, occurs when the outcome of a research study influences the likelihood of publishing the results. It is possible that cancer specialist researchers might be less committed to publishing results showing poor QoL outcomes for LTCS.

4.2.4. Data limitations

Finally, another possible explanation for the finding might be a lack of power. In many of the studies that make comparisons there are relatively few observations (minimum n=11; mean n=142; median n=134). It might be the case that there is not enough evidence to observe a difference between the LTCS sample and the general population sample, hence why so many studies reported similar QoL between the two samples. However, the two largest studies that made such comparisons (both n=344) reported that LTCS had better or similar QoL compared to the general population. Therefore, this may not be a suitable explanation for the observed results.

If a minimum sample size rule excluding studies with samples of n=50 or fewer (for example) had been applied, this would have reduced the number of studies reporting comparisons with a control group from 11 to eight. It would not have affected the overall conclusions of the review – of the three studies with very small samples, only one was categorised as reporting that the QoL of LTCS was better than or similar to the QoL of the comparator group (study 18). Indeed, the only study categorised as reporting *worse* QoL for LTCS (study 15) would have been excluded had such a rule been applied.

4.3. Limitations and gaps in the evidence base

This review has highlighted a range of limitations and gaps in the evidence base. First, whilst some of the included studies did use well-known PROs such as SF-36 and EORTC QLQ-C30, none of the included studies reported data on utilities. Ultimately, the research question could be better addressed with such data.

Another gap in the evidence base relates to the lack of subgroup analyses. The vast majority of studies pooled their data on LTCS when comparing this group to that of the general population. It would be useful to explore whether the stage of the disease experienced by the LTCS group would have affected the comparison with the general population. For example, it might be the case that LTCS that had only experienced the early stages of cancer had better QoL than the general population, whereas LTCS that had experienced advanced cancer did not. This would be even more important if the patients who had experienced early stage disease had been treated using curative therapies, in which case they may be less likely to be considered similar to LTCS with advanced disease.

There was also considerable variation between the included studies and therefore further research is required to improve confidence in the main finding of this review. For

example, the status of the patient samples at baseline and the definition of long-term survival differed. One study included individuals 10 years after diagnosis whereas others included individuals 3 years after diagnosis with metastatic disease. In addition, some studies collected pre-treatment QoL data whereas others only collected post-treatment data. Therefore, the comparisons being made in the included studies cannot be considered to be 'like-for-like'. The issue of the definition of long-term survival extends beyond the studies included in this review. There is little consistency between published studies regarding the number of years that constitutes long-term survival, which makes it particularly difficult to compare results and assess the overall trends in the literature.

Fewer than half of the studies included in the review reported comparisons between LTCS and a group of controls. Further, of the studies that *did* report such comparisons, some did not attempt (or report attempting) to adjust for key observable characteristics, and some of those that *did* adjust provided limited information about how the adjustment was undertaken. This lack of adjustment and reporting rigour limits our ability to make strong conclusions about the QoL of LTCS relative to the QoL of the general population.

Given that several studies reported that the QoL of LTCS with advanced disease was better than or similar to that of healthy controls, it seems plausible that their QoL would also be better than or similar to pre-progression baseline levels. However, none of the studies included in the review reported comparisons with pre-progression QoL. Hence, there is a lack of specific evidence to support an assumption in economic models that the QoL of LTCS with advanced disease should be modelled at their pre-progression levels.

Finally, another gap in the literature is that, whilst the results in this review are generally positive with respect to QoL in LTCS, we did not identify relevant data for individuals that have been treated with an immunotherapy. A large number of studies were identified during the review process that examined immunotherapies, a small number of which had follow-up periods that were in excess of five years for melanoma and lung cancer. However, none of the immunotherapy studies that we identified had collected and/or reported QoL data over the long-term. Immunotherapies have been used to treat several types of cancer and shown very encouraging results in recent years, with some treatments leading to impressive long-term survival rates (McDermott et al., 2016; Brahmer et al., 2017). Therefore, it is important to collect long-term QoL data for individuals that have been treated with immunotherapy. If immunotherapy is associated with fewer side effects relative to radiotherapy or chemotherapy (the most common treatments in the included studies), one might expect that the results would be more favourable in individuals treated with immunotherapy. It should be noted, however, that immunotherapy tends to involve relatively long-term treatment for patients with advanced disease, whereas many of the studies in this review examined the QoL of patients with predominantly early stage disease who received a shorter course of treatment.

It is important to generate longer-term QoL data in clinical development programmes, and for these data to be made available in the public domain (i.e. via journal articles and conference presentations) in order to provide robust evidence to support assumptions about the value of immunotherapies and other novel therapies. It would be helpful, for example, to include QoL measures such as the EQ-5D (or other preference-based measures for which utilities can be obtained via established value sets or mapping algorithms)in clinical trials and to report relevant outcomes alongside data relating to primary endpoints. We are aware that this has been done is some of the more recent

trials (not yet mature enough to be considered 'long-term') and anticipate an increase in the number of published studies reporting QoL outcomes for LTCS following immunotherapy treatment in the coming year or two. It would be particularly valuable to be able to understand how QoL outcomes compare across patients in different stages (i.e. III vs. IV) and sub-stages (i.e. IIIa vs. IIIb vs. IIIc).

4.4. Limitations of the review

This review has several limitations. First, whilst the initial intention was to identify studies that explored the QoL of LTCS that were treated with immunotherapy, no suitable records were identified. However, as this was considered as a possibility before the searches were conducted, it was possible to relax this focus by searching the records identified from the other, non-immunotherapy-related searches. Second, every effort was made to take a systematic and transparent approach during the review by producing and utilising clear inclusion and exclusion criteria. However, it should be acknowledged that both OHE Consulting and members of the project steering group played a role in the screening of the records at different stages. Finally, due to the nature of the project, the number of studies included in the review was limited and therefore this review cannot be considered an exhaustive review on the topic area of interest.

5. CONCLUSION

Notwithstanding the very limited evidence base, of the studies reporting comparisons with population norms, the majority provided evidence that the QoL of LTCS is similar to that of the general population. This general result also holds for the small number of studies in which most or all of the patients had late stage cancer. However, it is unclear how generalisable the results are, due to the different study settings. Therefore, care is required in asserting that general population utilities could or should be used for LTCS in economic models. Further, more directed research is required in a broad range of cancer and treatment types to provide a stronger foundation for such an assertion.

REFERENCES

Brahmer, J., Horn, L., Jackman, D., et al., 2017. *Five-year follow-up from the CA209-003 study of nivolumab in previously treated advanced non-small cell lung cancer: Clinical characteristics of long-term survivors*. Presentation at the 2017 AACR Annual Meeting. Abstract CT077. 3 April.

Brazier, J., Akehurst, R., Brennan, A., Dolan, P., Claxton, K., McCabe, C., Sculpher, M. & Tsuchiya, A., 2005. Should patients have a greater role in valuing health states? *Applied Health Economics and Health Policy*. 4 (4), pp.201–208.

Caravati-Jouvenceaux, A., Launoy, G., Klein, D., Henry-Amar, M., Abeilard, E., Danzon, A., Pozet, A., Velten, M. and Mercier, M., 2011. Health-related quality of life among long-term survivors of colorectal cancer: a population-based study. *The Oncologist*, 16(11), pp.1626-1636.

Carmichael, C., Yuh, B.E., Sun, V., Lau, C., Hsu, J., Saikia, J., Liu, X., Wilson, T., Ferrell, B. and Pal, S.K., 2013. Quality of life in patients with metastatic renal cell carcinoma: assessment of long-term survivors. *Clinical Genitourinary Cancer*, 11(2), pp.149-154.

Clark, M.M., Novotny, P.J., Patten, C.A., Rausch, S.M., Garces, Y.I., Jatoi, A., Sloan, J.A. and Yang, P., 2008. Motivational readiness for physical activity and quality of life in long-term lung cancer survivors. *Lung Cancer*, 61(1), pp.117-122.

Cubí-Mollá, P., Jofre-Bonet, M. & Serra-Sastre, V., 2017. Adaptation to health states: Sick yet better off? *Health E*conomics. doi:10.1002/hec.3509

Duke, R.L., Campbell, B.H., Indresano, A.T., Eaton, D.J., Marbella, A.M., Myers, K.B. and Layde, P.M., 2005. Dental Status and Quality of Life in Long-Term Head and Neck Cancer Survivors. *The Laryngoscope*, 115(4), pp.678-683.

Ferro, T., Aliste, L., Valverde, M., Fernández, M.P., Ballano, C. and Borràs, J.M., 2014. Health status and health resource use among long-term survivors of breast, colorectal and prostate cancer. *Gaceta Sanitaria*, 28(2), pp.129-136.

Greimel, E., Daghofer, F. and Petru, E., 2011. Prospective assessment of quality of life in long-term ovarian cancer survivors. *International Journal of Cancer*, 128(12), pp.3005-3011.

Hammerlid, E. and Taft, C., 2001. Health-related quality of life in long-term head and neck cancer survivors: a comparison with general population norms. *British Journal of Cancer*, 84(2), pp.149-156.

Hartung, T.J., Mehnert, A., Friedrich, M., Hartmann, M., Vehling, S., Bokemeyer, C. and Oechsle, K., 2016, February. Age-related variation and predictors of long-term quality of life in germ cell tumor survivors. *Urologic Oncology: Seminars and Original Investigations*, 34(2), pp. 60-e1.

Jansen, L., Hoffmeister, M., Chang-Claude, J., Brenner, H. and Arndt, V., 2011. Benefit finding and post-traumatic growth in long-term colorectal cancer survivors: prevalence, determinants, and associations with quality of life. *British Journal of Cancer*, 105(8), pp.1158-1165.

Lutgendorf, S.K., Shinn, E., Carter, J., Leighton, S., Baggerly, K., Guindani, M., Fellman, B., Matzo, M., Slavich, G.M., Goodman, M.T. and Tew, W., 2017. Quality of life among

long-term survivors of advanced stage ovarian cancer: A cross-sectional approach. *Gynecologic Oncology*, doi:10.1016/j.ygyno.2017.05.008

Meisel, J.L., Domchek, S.M., Vonderheide, R.H., Giobbie-Hurder, A., Lin, N.U., Winer, E.P. and Partridge, A.H., 2012. Quality of life in long-term survivors of metastatic breast cancer. *Clinical Breast Cancer*, 12(2), pp.119-126.

McDermott, D.F., Motzer, R.J., Atkins, M.B., et al., 2016. Long-term overall survival (OS) with nivolumab in previously treated patients with advanced renal cell carcinoma (aRCC) from phase I and II studies. *Journal of Clinical Oncology*, 34(15), suppl. (May 2016), p.4507.

National Institute for Health and Clinical Excellence (NICE), 2015. *Obinutuzumab in combination with chlorambucil for untreated chronic lymphocytic leukaemia*. NICE Technology Appraisal Guidance 343. Available from: https://www.nice.org.uk/guidance/ta343 [Accessed 19th December 2017]

National Institute for Health and Clinical Excellence (NICE), 2017. *Avelumab for merkel cell carcinoma.* NICE Technology Appraisal Guidance (in development) ID1102. Available from: https://www.nice.org.uk/guidance/indevelopment/gid-ta10136 [Accessed 19th December 2017]

Nes, L.S., Liu, H., Patten, C.A., Rausch, S.M., Sloan, J.A., Garces, Y.I., Cheville, A.L., Yang, P. and Clark, M.M., 2012. Physical activity level and quality of life in long term lung cancer survivors. *Lung Cancer*, 77(3), pp.611-616.

Ogorevc, M., Murovec, N., Fernandez, N.B. & Rupel, V.P., 2017. Questioning the differences between general public vs. patient based preferences towards EQ-5D-5L defined hypothetical health states. *Health Policy*. doi:10.1016/j.healthpol.2017.03.011

Oskam, I.M., Verdonck-de Leeuw, I.M., Aaronson, N.K., Witte, B.I., de Bree, R., Doornaert, P., Langendijk, J.A. and Leemans, C.R., 2013. Prospective evaluation of health-related quality of life in long-term oral and oropharyngeal cancer survivors and the perceived need for supportive care. *Oral Oncology*, 49(5), pp.443-448

Osoba, D., Rodrigues, G., Myles, J., Zee, B. and Pater, J., 1998. Interpreting the significance of changes in health-related quality-of-life scores. Journal of Clinical Oncology, 16(1), pp.139-144

Osthus, A.A., Aarstad, A.K.H., Olofsson, J. and Aarstad, H.J., 2011. Health-Related Quality of Life Scores in long-term head and neck cancer survivors predict subsequent survival: a prospective cohort study. *Clinical Otolaryngology*, 36(4), pp.361-368

Paterson, K., 2017. Value and cancer medicines—A personal view. *Journal of Cancer Policy*, 11, pp.26-31.

Payakachat, N., Ounpraseuth, S. and Suen, J.Y., 2013. Late complications and long-term quality of life for survivors (> 5 years) with history of head and neck cancer. *Head & Neck*, 35(6), pp.819-825.

Phipps, E., Braitman, L.E., Stites, S. and Leighton, J.C., 2008. Quality of life and symptom attribution in long-term colon cancer survivors. *Journal of Evaluation in Clinical Practice*, 14(2), pp.254-258.

Rausch, S.M., Gonzalez, B.D., Clark, M.M., Patten, C., Felten, S., Liu, H., Li, Y., Sloan, J. and Yang, P., 2012. SNPs in PTGS2 and LTA predict pain and quality of life in long term lung cancer survivors. *Lung Cancer*, 77(1), pp.217-223.

van der Schroeff, M.P., Derks, W., Hordijk, G.J. and de Leeuw, R.J., 2007. The effect of age on survival and quality of life in elderly head and neck cancer patients: a long-term prospective study. *European Archives of Otorhinolaryngology*, 264(4), pp.415-422.

Sprangers, M.A. & Schwartz, C.E., 1999. Integrating response shift into health-related quality of life research: a theoretical model. *Social Science & Medicine*. 48 (11), pp.1507–1515.

Stiggelbout, A.M. & de Vogel-Voogt, E., 2008. Health state utilities: a framework for studying the gap between the imagined and the real. *Value in Health*. 11 (1), pp.76–87.
Ubel, P.A., Loewenstein, G. & Jepson, C., 2003. Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. *Quality of Life Research*. 12 (6), pp.599–607.

Vainshtein, J.M., Moon, D.H., Feng, F.Y., Chepeha, D.B., Eisbruch, A. and Stenmark, M.H., 2015. Long-Term Quality of Life After Swallowing and Salivary-Sparing Chemo–Intensity Modulated Radiation Therapy in Survivors of Human Papillomavirus–Related Oropharyngeal Cancer. *International Journal of Radiation Oncology* Biology* Physics*, 91(5), pp.925-933.

van der Schroeff, M.P., Derks, W., Hordijk, G.J. and de Leeuw, R.J., 2007. The effect of age on survival and quality of life in elderly head and neck cancer patients: a long-term prospective study. *European Archives of Otorhinolaryngology*, 264(4), pp.415-422.

Vartanian, J.G. and Kowalski, L.P., 2009. Acceptance of major surgical procedures and quality of life among long-term survivors of advanced head and neck cancer. *Archives of Otolaryngology–Head & Neck Surgery*, 135(4), pp.376-379.

Versteegh, M.M. and Brouwer, W.B.F., 2016. Patient and general public preferences for health states: a call to reconsider current guidelines. *Social Science & Medicine*, 165, pp.66-74.

Williams, K., Steptoe, A. & Wardle, J., 2013. Is a cancer diagnosis a trigger for health behaviour change? Findings from a prospective, population-based study. *British Journal of Cancer*. 108 (11), pp.2407–2412.

Appendix A. Data Extraction Form

	Extracted Information
Full Citation [from Google Scholar, Harvard style]	
Type of Publication [journal article / conference poster]	
Country	
Study Objective(s) [include stated secondary objectives; use authors' words if possible]	
Study Type	
Sample Size [report size of final sample and any key subgroups]	
Population Characteristics	
Cancer Type	
Cancer Stage	

	Extracted Information
[stage at time of study; if reported, stage at time of diagnosis]	
Intervention(s) [intervention being assessed in study; any prior interventions reported]	
Any reference to immunotherapy?	
Length of Survival	
Outcome Measure(s)	
Length of Study Follow-Up	
Main Results [i.e. PRO/utility data] [if utility values are reported, the actual values should be extracted]	
Comparison with General Population [if one is made]	

	Extracted Information
Comparison with patients at Baseline [if one is made]	
Direct Quotes [that relate to the research question(s)]	
Key Papers Referenced [that aren't in our list]	
Comments	