

CRITIQUE OF A SYSTEMATIC REVIEW OF STUDIES REPORTING HEALTH RELATED QUALITY OF LIFE EFFECTS FOR CAREGIVERS AND FAMILY MEMBERS

REPORT BY THE DECISION SUPPORT UNIT

13 February 2020

Alison Scope¹, Ruth Wong¹ Becky Pennington¹

¹School of Health and Related Research, University of Sheffield

Decision Support Unit, ScHARR, University of Sheffield, Regent Court, 30 Regent
Street
Sheffield, S1 4DA

Tel (+44) (0)114 222 0734
E-mail dsuadmin@sheffield.ac.uk
Website www.nicedsu.org.uk
Twitter [@NICE_DSU](https://twitter.com/NICE_DSU)

ABOUT THE DECISION SUPPORT UNIT

The Decision Support Unit (DSU) External Assessment Centre is based at the University of Sheffield with members at York, Bristol, Leicester and the London School of Hygiene and Tropical Medicine. The DSU is commissioned by The National Institute for Health and Care Excellence (NICE) to provide a research and training resource to support the Institute's Centre for Health Technology Evaluation Programmes. Please see our website for further information www.nicedsu.org.uk.

The production of this document was funded by the National Institute for Health and Care Excellence (NICE) through its Decision Support Unit. The views, and any errors or omissions, expressed in this document are of the authors only. NICE may take account of part or all of this document if it considers it appropriate, but it is not bound to do so.

Acknowledgements

The authors would like to thank Dr Tessa Peasgood (School of Health and Related Research, University of Sheffield), Dr Clara Mukuria (School of Health and Related Research, University of Sheffield) and Ian Watson (National Institute for Health and Care Excellence) for reviewing this report and providing insightful comments.

EXECUTIVE SUMMARY

The health burden of family caregiving has been shown to be an important aspect in the analysis of the cost effectiveness of an intervention, with wide ranging impacts of illness on the primary caregiver, and wider family, including physical, psychological, and emotional effects. These effects have been termed spillover effects. NICE's Reference Case for the methods of Technology Appraisal (TA); also used in Highly Specialised Technologies (HSTs); states that "all direct health effects, whether for patients, or when relevant, carers" should be considered. NICE's Reference Case states that health effects should be measured in quality-adjusted life years (QALYs).

A recent systematic review has sought to examine these spillover effects, and to catalogue spillover-related utilities to facilitate their consideration in cost effectiveness analyses. The authors identified articles that reported preference-based measures of spillover effects, including 80 full text articles in the review. The authors of the systematic review present a catalogue of utilities to facilitate the calculation of QALYs and inform CEAs. The objective of the current paper was to summarise the findings and provide a critical appraisal of this recent review.

Data were extracted from the systematic review and a quality assessment of the review using a tool specifically designed for assessing systematic reviews, the AMSTAR-2 (A MeaSurement Tool to Assess systematic Reviews), was performed. The quality assessment, specifically examined the following areas; the research question and inclusion criteria for studies, the literature search, data extraction and checking, assessment of study quality, and the description of the included studies.

This recent review highlights similar issues to our previous reports, the April 2019 review of NICE appraisals (Stage 1), and the stage 2 update review of cost-utility analyses that have included carer and family member health-related quality of life. For example, as in our reports, Alzheimer's disease and other types of dementia were the most frequently studied conditions (15 articles). The majority of the studies focused on caregivers/family members of ill adults (59%), but again, as in the stage 1 and stage 2 reports, 14 studies focused on ill children (18%). A minority of studies focused on rare conditions. Again, like our reports the EQ-5D was the most common

instrument used to measure caregiver/family member utility, with some articles reporting multiple measurement methods, and most of the studies were conducted in Europe (66%), the US and Canada (25%).

The results of the critique demonstrated that, overall the quality of the review was acceptable, considering the objectives of the review, with adequate/good searches, and transparent and detailed criteria for study eligibility and selection methods. Selection of studies for inclusion was completed independently by two researchers. However, it does not appear that double data extraction or checking of data extractions was completed and quality assessment of the included studies was not conducted. The remit of the systematic review was to document clearly and comprehensively the range of studies reporting preference-based family member utility values. The review did not seek to provide further analysis of these data, other than narrative synthesis, and tabulation of the studies, and did not intend to infer anything further from these results.

The authors concluded that in general, utilities indicated a loss in quality of life associated with being a caregiver or family member of an ill relative. The authors suggest that additional research is needed on methods of measuring and incorporating spillover QALYs. Most studies reported caregiver/family member utility without any comparator, limiting the ability to infer spillover effects. They also highlight the spectrum of diseases and conditions for which caregiver and family members' spillover effects have been measured, and the variation in measurement methods used. The authors of the review acknowledge that as articles including utilities for spillover effect are being published regularly, their catalogue is likely to become incomplete quickly, but they are developing an online repository that will be updated regularly as a public resource. The critique of the review showed that quality assessment had not been conducted and further the authors of the review acknowledged that to facilitate accessibility of information some important detail may have been missed from the tables in their repository, this emphasises that users of the repository should be advised to revisit individual original papers if they intend to use the data from them.

With careful consideration of quality of the individual studies, there may be generalisations made in terms of the methods and measures used in future CUAs,

and this may have implications for NICE and HTA bodies, pharmaceutical companies and researches. For example, across all three reports (stage 1, 2 and 3) Alzheimer's disease and other types of dementia were the most frequently studied conditions and the EQ-5D was the most common instrument used to measure caregiver/family member utility. Furthermore, most of the studies were conducted in Europe, the US and Canada, making generalisations to these settings most comparable. The review critiqued here highlighted the importance of using studies with a comparator, and that a standard definition of caregiving is required when incorporating carer and family quality of life into cost effectiveness analyses. The review is useful as it documents the entire literature at present. As the body of literature grows a review including an assessment of quality, and further analysis of the data reported in each study will be required.

CONTENTS

1. INTRODUCTION	8
1.1. BACKGROUND.....	8
2. METHODS	8
2.1. DATA EXTRACTION.....	8
2.2. QUALITY ASSESSMENT.....	9
3. RESULTS	10
3.1. DATA EXTRACTION.....	10
3.2. SUMMARY OF RESULTS AS REPORTED BY WITTENBERG ET AL. (2019)10	
3.2.1. <i>Author conclusions</i>	11
3.2.2. <i>Limitations and uncertainties</i>	12
3.3. QUALITY ASSESSMENT OF THE SYSTEMATIC REVIEW.....	13
3.3.1. <i>Research question and inclusion criteria</i>	13
3.3.2. <i>Literature Search</i>	14
3.3.3. <i>Data extraction / checking</i>	17
3.3.4. <i>Quality assessment of the included studies conducted by the systematic review authors</i>	17
3.3.5. <i>Description of included studies</i>	18
4. CONCLUSION	19
5. REFERENCES	22
APPENDIX	23
A.1 DATA EXTRACTION FROM THE SYSTEMATIC REVIEW.....	23

TABLES

Table 1: Summary of the methods and results of Wittenberg et al. (2019).....	10
Table 2: Update of the Wittenberg, James & Prosser (2019) MEDLINE search in August 2019.....	17

ABBREVIATIONS AND DEFINITIONS

Abbreviation	Definition
AMSTAR	A MeaSurement Tool to Assess systematic Reviews
CarerQoL	Care-related Quality of Life instrument
CEA	Cost effectiveness analysis
CES	Carer Experience Scale
CHU-9D	Child Health Utility-Nine Dimension
CUA	Cost Utility Analysis
EG	Example
EQ-5D	EuroQol-Five Dimension
EQ-VAS	EuroQol-visual analogue scale
HRQL	Health-related quality of life
HUI	Health Utilities Index
ICER	Incremental cost-effectiveness ratio
NICE	The National Institute for Health and Care Excellence
PRESS	Peer Review of Electronic Search Strategies checklist
QALY	Quality-adjusted life year
QWB	Quality of wellbeing scale
TA	Technology Appraisal
SD	Standard deviation
SE	Standard error
SF-36	Short Form 36
STA	Single Technology Appraisal

1. INTRODUCTION

1.1. BACKGROUND

NICE's Reference Case for the methods of Technology Appraisal (TA) (1); also used in Highly Specialised Technologies (HSTs); states that "all direct health effects, whether for patients, or when relevant, carers" should be considered. These quality of life impacts to carers and family members are known as spillover effects. Taking into account the burden of family caregiving has been shown to be an important aspect in the analysis of the cost effectiveness of an interventions, with wide ranging impacts on the primary caregiver, and wider family, including physical, psychological, and emotional effects (2).

A recent systematic review has sought to examine these spillover effects (3). The objective was specifically, to catalogue spillover-related utilities to facilitate their consideration in cost effectiveness analyses. The authors identified articles that reported preference-based measures of spillover effects, identifying 3695 citations and ultimately including 80 full text articles in the review. The authors have stated that they are developing an online repository, which will be updated regularly with newly published articles reporting preference-based measure of spillover effects.

This review

The objective of this report is to present a critical appraisal of this recently published paper, Wittenberg, Lyndon and Prosser (2019) Spillover Effects on Caregivers' and Family Members' Utility: A Systematic Review of the Literature (3).

2. METHODS

2.1. DATA EXTRACTION

We extracted data from the review and created a summary table of methods and results. The data extracted included details of authors, searching methods, study selection methods, data extraction and quality assessment methods, number and types of included studies, study characteristics of included studies, results of the included studies, author conclusions, and limitations and uncertainties.

2.2. QUALITY ASSESSMENT

We undertook quality assessment of the systematic review using some of the questions in AMSTAR-2 (A Measurement Tool to Assess systematic Reviews)(4). AMSTAR-2 is a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions. Although the Wittenberg et al review is not an intervention review, several of the items detailed in AMSTAR 2 have relevance to the review. The tool comprises 16 items covering the stages of the systematic review process, five of these were relevant to the assessment of this systematic review and were used to assess if the Wittenberg et al. (2019)(3) review was deemed to be good quality:

- Did the review authors present a clear research question and inclusion criteria for studies?
- Did the review authors use a comprehensive literature search strategy? This includes searching at least 2 databases, providing keywords and/or search strategies, and justifying publication restrictions.
- Did the review authors perform data extraction/checking in duplicate?
- Did the review authors use a satisfactory technique for assessing study quality?
- Did the review authors describe the included studies in adequate detail?

To fully assess the literature search strategy, the reported search methods and strategies included in the supplementary material were reviewed using the Peer Review of Electronic Search Strategies (PRESS) checklist (5). The PRESS checklist comprises seven elements (translation of the research question; Boolean and proximity operators; subject headings; text word search; spelling, syntax and line numbers; and limits and filters). In addition, the contribution of sources searched and sensitive of the MEDLINE were assessed.

3. RESULTS

3.1. DATA EXTRACTION

The summary table of methods and results are presented in Table 1. A full data extraction is presented in Appendix A.1 DATA EXTRACTION FROM THE SYSTEMATIC REVIEW.

Table 1: Summary of the methods and results of Wittenberg et al. (2019)

Author (date)	Methods	Number and types of included studies,	Study characteristics of included studies
Wittenberg et al. (2019)(3)	Three Electronic databases were searched. Study selection completed in duplicate. Data extraction not completed in duplicate. Quality assessment not completed.	N= 80 8 reported spillover utility/disutility: 4 reported spillover disutility as the difference between population utility and the observed family caregiver utility, 1 reported disutilities only, 1 reported both the difference between the observed caregiver utility and the population utility, as well as a utility for a hypothetical scenario, and 2 reported spillover utilities only. 25 reported a comparison group. 50 reported caregiver/ family member utilities only.	53 were conducted in Europe, 20 in the US and Canada. 15 reported on Alzheimer's disease/dementia. 6 reported on cancer. 46 focused on caregivers/family members of ill adults, 14 on ill children, the remainder focused on adults and children combined. 58 reported EQ-5D 13 reported SF-6D 3 reported HUI 2 reported QWB 7 reported Caregiver-focused instruments (the CarerQoI and CES). 6 reported caregiver/family member utility in the context of a patient and/or caregiver intervention trial.

3.2. SUMMARY OF RESULTS AS REPORTED BY WITTENBERG ET AL. (2019)

The search conducted by Wittenberg et al. (2019) yielded 5205 records, with 3695 studies screened by title and abstract after removing duplicates. 177 full-text articles were screened for eligibility, and 80 articles were subsequently included in the review, and will therefore be included in the online repository.

Table 1 above, summarises the methods and results of the studies included in the Wittenberg et al. (2019) review. The authors describe how the study reported results, including, how spillover utility or disutility was reported across the included studies and that twenty-five (30%) reported a comparison group, either general population norms (n = 9) or matched, non-caregiver/family members or hypothetical scenarios' utilities (n = 16). Fifty (63%) reported caregiver/family member utilities only, some articles reported utilities for multiple conditions or using multiple measurement methods, or for multiple strata of caregivers/family members.

Similarly to the stage 1 and stage 2 reports, Alzheimer's disease and other types of dementia were the most frequently studied conditions (15 articles) in the Wittenberg et al. (2019) review. The majority of the studies focused on caregivers/family members of ill adults (59%), but again, as in the stage 1 and stage 2 reports, 14 studies focused on ill children (18%). A minority of studies focused on rare conditions with three on Duchenne Muscular Dystrophy, two on Fragile X syndrome and one on Dravet syndrome.

Again, like our reports the EQ-5D was the most common instrument used to measure caregiver/family member utility , with some articles reporting multiple measurement methods. The authors also reported that 95% of the studies used generic (i.e., indirect) measurement instruments, with the the SF-6D used in 13 instances (16%), and the HUI and QWB used three and two times, respectively. The caregiver-focused instruments (the CarerQoI and CES) were used in seven instances (six uses and one use, respectively; 9%). Six articles (8%) reported caregiver/family member utility in the context of a patient and/or caregiver intervention trial. Again, as in our reports, most of the studies were conducted in in Europe (66%), the US and Canada (25%).

3.2.1. Author conclusions

The authors of the systematic review present a catalogue of utilities to facilitate the calculation of QALYs and inform CEAs. The authors concluded that in general, utilities indicated a loss in quality of life associated with being a caregiver or family

member of an ill relative. The authors suggest that additional research is needed on methods of measuring and incorporating spillover QALYs, to ascertain which are the most appropriate methods. They highlight the inconsistency of the terminology and the definitions of caregiving and family involvement and how this may causes variability in the estimation of spillover QALYs, and how they are interpreted. Most studies reported caregiver/family member utility without any comparator, limiting the ability to infer spillover effects. The authors raise the issue that They also highlight the spectrum of diseases and conditions for which caregiver and family members' spillover effects have been measured, and the variation in measurement methods used, again suggesting that further research needs to be undertaken to strengthen the evidence base within disease areas and using the same measures.

Nevertheless, these values provide a starting point for considering spillover effects in the context of CEA, opening the door for more comprehensive analyses. They confirm their goal is to advance the inclusion of spillover in CEA by providing this accessible overview of the spillover effects of health-related quality of life (HRQOL) literature, and to expand the knowledge base of spillover-based CEAs, to answer remaining questions.

3.2.2. Limitations and uncertainties

The authors of the review acknowledge that as articles including utilities for spillover effect are being published regularly, their catalogue is likely to become incomplete quickly, but they are developing an online repository that will be updated regularly as a public resource. Other reported limitations include the fact that grey literature and other unpublished sources are excluded from the catalogue. The authors acknowledge that the data included in the tables were limited to facilitate accessibility for readers, therefore it should be considered that some detail may have been omitted. A final limitation that was noted by the authors, was that they had 'made subjective judgments about the relative salience of utilities in articles reporting multiples, but describe others in the 'notes' section of the tables'.p.495. This emphasises that users of the repository should be advised to revisit individual original papers if they intend to use the data from them.

3.3. QUALITY ASSESSMENT OF THE SYSTEMATIC REVIEW

3.3.1. Research question and inclusion criteria

The authors reported that the objective of the review was to catalogue spillover-related health utilities to facilitate their consideration in CEAs. The populations of patients included across studies was not limited and depended on the other elements of the eligibility criteria, namely the outcome measures applied in the studies. As such, articles reporting on any patient diseases and conditions, including those that specified no disease were included, as were studies reporting on patient health states, defined as a distinct phase of a disease or condition. The primary inclusion / exclusion criterion was that the study must report a 'preference based measure of caregiver or family member utility or disutility, including caregiver-focused measures for which population tariffs exist (i.e., CarerQoL and CES)'p.476-477 (3). The authors also documented exclusions for studies that only reported a visual analog scale unless it had been transformed into utilities using a known algorithm. The eligibility criteria also allowed for the inclusion of studies reporting on multiple patient conditions and/or using multiple preference based measures or instruments, and this is clearly documented in the paper in the appropriate section. Implicit in the eligibility criteria was that data was from a family member, and it was clearly documented how the family member should be defined, namely, anyone identified as having a familial relationship to the patient, and included studies reporting on informal caregivers unless they were specifically described as non-familial. Paid caregivers were excluded. No age limits were applied on either patients or caregivers for inclusion in the review. The study types included were those specifically designed to measure spillover utility, those measuring caregiver/family member utility among other outcomes, and caregiver or patient interventions that included utility as an outcome. The intervention and comparator involved in the studies were not defined, as inclusion was primarily based on the appropriate outcome measures being present in the study. The authors' eligibility criteria only allowed articles written in English to be included. This is a commonly used exclusion criterion in systematic reviews and is generally viewed as an acceptable limit, it should be noted that it can result in the exclusion of relevant data. No justification for this was provided.

3.3.2. Literature Search

The article authors searched three electronic databases in April 2018 to retrieve studies reporting preference-based measure of caregiver or family member spillover effects in MEDLINE, Embase and EconLit. The authors reported that the search strategy was developed and refined from a known set of relevant articles from the previous review (2). The previous review searches only covered literature from 2000 until 2012 and the MEDLINE strategy was not reviewed in this report.

Overall search strategy and reporting: The search strategies were systematic in the approach taken, concept combinations and text words included were reasonably comprehensive. There was variation in the search strategy syntax used (field searching, nesting, limits applied) which suggests that MEDLINE, Embase and EconLit were searched on different platforms. The database host platforms, date coverage of the database since inception were not reported. Having repeated the author's search in MEDLINE, the numbers retrieved were consistent with those reported in the article.

Sources searched: The value of searching Embase in addition to MEDLINE is unclear given that the conference abstracts indexed in Embase were excluded. Whilst EconLit covers economic literature, economic evaluation searches should also include the NHS Economic Evaluations Database via the NIHR Centre for Reviews and Dissemination (6).

In section 4.3. study limitations, the authors acknowledged that search of grey literature and unpublished articles including utilities for spillover should have been searched. We conducted precise searches for unpublished studies, in International Society for Pharmacoeconomics and Outcomes Research Presentations Database (7) and International Society for Quality of Life Research Supplements (8). These searches yielded 70 records to screen for inclusion.

As mentioned previously, the following paragraphs relate to the 7 elements of the PRESS checklist (5)

Translation: The MEDLINE search strategy comprises text words for utility, specifically named measures (EQ-5D, SF-6D, HUI, QWB, CarerQoL, CES and CHU-9D), methods for measuring health state, preference weights and economic evaluations combined with Medical Subject Headings (MeSH) heading/Emtree and text words for caregiver and family members (partner, spouse, parent, child, sibling, grandparent, and next of kin). The translation of the utility terms across all three databases are consistent. However, there are minor inconsistencies in the translation of the caregiver combined with the outcome terms (i.e. effect, burden, impact and consequence) between databases. There was no justification reported for the inclusion of named measures, and exclusions of others.

Boolean and Proximity Operators: The use of Boolean and proximity operators are correctly applied across the databases. The sensitivity of the caregiver concept could be increased by searching for carer or carers or caring or caregiver* or caregiving without limiting with terms for 'burden' using the proximity operator. However, taking this broad approach would lead to additional records retrieved, although the number of relevant studies are unclear.

MeSH heading or Emtree terms: There were no subject terms included in the utility search concept and only free-text words. According to the Cochrane handbook, both free-text and appropriate subject headings should be searched (9). The sensitivity of the search could be increased by the inclusion of terms from the health state utility values (HSUV) search filter such as Quality-Adjusted Life Years/(10). For caregivers, appropriate subject terms were included but could be broadened e.g by adding exp family/. The inclusion of these subject terms in MEDLINE would give about 50 percent increase in the number of records retrieved for screening.

Text words: For HSUVs, see Arber et al., (2017)(10) for potentially relevant text words to increase sensitivity of the search by including the term variants for the 7 named measures . For the caregiver concept, additional text words could include care giv* or families or relative* or wives or son* or daughter* or offspring* or sibling* or dependents or kinship. However, the impact of including these terms on the number of eligible records retrieved are unclear.

Spelling, syntax and line numbers: There were no spelling or line number errors in the search strategies.

Limits and filters: No date or language limits were applied to the search. The application of the conference abstract limit to the Embase search strategy would exclude unpublished but potentially relevant studies. Search filters exist for health state utility values as well as economic evaluations. However, there are no search filters available for caregivers.

Search sensitivity: Assessment of the sensitivity of the MEDLINE search strategy revealed that 72 (90%) of the 80 records were retrieved in the MEDLINE search alone. The remainder eight records were indexed in MEDLINE but not retrieved by the search strategy. Seven out of eight included studies were not retrieved by the strategy due to the caregiver concept (Brisson, 2010; Chevreul, 2016; Chevreul, 2015; Chevreul, 2015; Mohide, 1988; Sjolander, 2012; Vroomen, 2016), whereas only one study was due to the utility search concept (van Exel, 2005). It is possible that these studies were identified via other sources searched by the authors.

To give an indication of the growth and size of the literature since the publication searches in April 2018 (of 3695 records), an update of the original MEDLINE search alone in August 2019 resulted in a further 533 records to screen since the publication (**Error! Reference source not found.**). This would be retrieved by the author's online repository.

Table 2: Update of the Wittenberg, James & Prosser (2019) MEDLINE search in August 2019

#	Searches	Results
1	(utilit* or disutilit* or preference-based or quality adjusted life year* or qaly or qalys or sf 6d or sf6d or eq-5d* or eq5d* or euroqol* or carerqol* or ces or carer experience scale or hui or health utilities index or quality of wellbeing scale or quality of well-being scale or qwib or standard gamble or time tradeoff or time trade-off or visual analog or visual analogue or vas or chu-9d or (quality of life and economic*) or preference weight* or eqvas or preference score* or economic evaluation*).ab,ti.	302355
2	exp caregivers/ or (spillover* or spill over*).ab,ti. or ((carer or carers or caring or caregiver* or caregiving or family or partner* or husband or wife or spouse or spouses or spousal or child or children or parent* or mother* or maternal or father* or paternal or sibling* or brother* or sister* or grandparent* or grandmother* or grandfather* or close person* or next of kin) adj4 (effect* or burden* or impact* or consequence*)).ab,ti.	135675
3	1 and 2	3032
4	(201804* or 201805* or 201806* or 201807* or 201808* or 201809* or 20181* or 2019*).dt,ed.	2547530
5	3 and 4	533

3.3.3. Data extraction / checking

The authors report that two reviewers independently screened titles, abstracts and full text articles, and resolved any conflicts by consensus, with reasons for exclusions documented. The items to be extracted were clearly documented in the review, together with how multiple conditions, and multiple utilities reported in a study would be dealt with in data extraction. Although the data to be extracted was clearly documented, no detail on the method of data extraction, including examples and data extraction forms, or whether double data extraction or checking had taken place, was provided. Previous work has highlighted the importance of double data extraction, demonstrating that errors can occur if this process is not completed (11).

3.3.4. Quality assessment of the included studies conducted by the systematic review authors

The review authors do not report completing any assessment of study quality for the included studies. Given that the objective of the review was to catalogue studies identified in a specified topic area, quality of the studies would not be a criterion on which studies could be excluded from the review. However, analysts using the results of the paper and decision-makers critiquing analyses may determine that

studies identified in this review were not of sufficient quality for use in decision-making.

The authors comment that publication bias is not a concern as the results are descriptive and not intended for inference. However, we note that publication bias is possible, for example if research aimed to demonstrate that a specific disease affected carer HRQL but found no difference in carer utility.

The authors further acknowledge that they made some subjective judgements about some of the results reported in the included studies. Therefore, there is the potential for readers of the review to overlook relevant detail, and it will be important for users of the repository to retrieve the individual papers to make their own judgements.

Further detail of study quality could be a useful data item for readers of the review. Items such as selection bias, bias due to missing data, bias in outcome measurement, and selective reporting bias, as included in critical appraisal tools for non-randomised intervention studies (e.g. ROBINS-I)(12) may have been applicable to several studies included in the review.

3.3.5. Description of included studies

Included studies were described in a narrative synthesis, and in more detail in a number of tables. The studies were grouped into those reporting 1) spillover utilities and disutilities (the independent utility loss as a result of being a carer); 2) caregiver and/or family member utility reported with a matched or population comparison group; and 3) caregiver and/or family member utility reported alone. The detail contained in the review appeared to be adequate to meet the authors' objectives of providing a catalogue of studies which reported preference-based measures of spillover effects. The authors comment in the limitations section that they limited the data included in the table to ensure accessibility for readers, and acknowledge that that some detail may have been omitted.

4. CONCLUSION

The study authors document a catalogue of studies that have measured caregiver and family members' utility, demonstrating the variation in measurement methods used for these utilities, and the wide range of conditions and diseases in which spillover effects have been measured. The remit of the systematic review was to document the included studies clearly and comprehensively. The review did not seek to provide further analysis of these data, other than narrative synthesis, and tabulation of the studies, and did not intend to infer anything further from these results.

The authors concluded that in general, utilities indicated a loss in quality of life associated with being a caregiver or family member of an ill relative. The authors suggest that additional research is needed on methods of measuring and incorporating spillover QALYs, to ascertain which are the most appropriate methods. They make the point that the inconsistent descriptions of caregiving and family involvement in the literature causes variability for both the estimation of spillover QALYs and the interpretation of them. In this systematic review studies were excluded if they specifically referred to carers as non-familial, again this highlights how the reporting in the primary studies may have biased which studies were included. Further, the nature of intensity of caregiving and how to deal with this for the wider family is still unclear when the existing literature is still focused on the primary caregiver. Standard definitions for what is to be included as family or caregiver spillover utility are required for future research.

The review authors note that as most studies reported caregiver/family member utility using conventionally defined health utility scores. These scores may include the impact of spillover but also the underlying health of the individual, often these studies are presented without any comparator, limiting the ability to infer spillover effects. They also raise the issue to that an assumption in the literature is that spillover effects are additive, but that interaction effects have been hypothesised, and they highlight that CUAs that incorporate family and caregiver utility still have to make a number of assumptions in these analyses. They also highlight the spectrum of diseases and conditions for which caregiver and family members' spillover effects

have been measured, and the variation in measurement methods used, again suggesting that further research needs to be undertaken to strengthen the evidence base within disease areas and using the same measures.

Whilst the review authors feel the primary studies in the review provide data to inform family-based CUAs, they also raise methodological questions around how to appropriately incorporate spillover effects into CUAs which are not addressed by their review, and would need further work to inform those questions. They also state that further work is needed to look at equity issues, for example, interventions that affect isolated patients such as homeless people may be undervalued compared to those that affect children who may have more carers associated with the analysis. A final methodological issue that the authors have not fully addressed in this review concerns situations where double counting occur, particularly where it is difficult to disentangle caregiver or family health from that of the patient.

Nevertheless, these values provide a starting point for considering spillover effects in the context of CEA, opening the door for more comprehensive analyses. They confirm their goal is to advance the inclusion of spillover in CEA by providing this accessible overview of the spillover effects of health-related quality of life (HRQOL) literature, and to expand the knowledge base of spillover-based CEAs, to answer remaining questions.

The critique of the review demonstrated that overall the quality of the review was acceptable considering the objectives of the review, with adequate/good searches, transparent and detailed criteria for study eligibility and selection methods. Selection of studies for inclusion was completed independently by two researchers. However, number of steps in the systematic review process have been omitted, limiting the review to a systematic search and identification of studies, rather than a systematic review, which together with a systematic search and identification of evidence, is defined as providing a high level of evidence. Although, it should be noted that existing definitions of systematic reviews still largely relate to intervention reviews, which may not be applicable in this topic area.

Of note the lack of any quality assessment of the included studies may have implications for users of the repository. Users of the repository should consider conducting their own quality assessment if using included studies, and analysts using the results of the paper and decision-makers critiquing analyses may determine that studies identified in this review were not of sufficient quality for use in decision-making. The authors further acknowledge that they made some subjective judgements about some of the results reported in the included studies. Therefore, there is the potential for readers of the review to overlook relevant detail, and it will be important for users of the repository to retrieve the individual papers to make their own judgements.

Some similar findings were reported in the Stage 1 and stage 2 reports, and may have implications for NICE and HTA bodies, pharmaceutical companies and researches, in that with careful consideration of quality of the individual studies, there may be generalisations made in terms of the methods and measures used in future CUAs in these disease areas. Across all three reports Alzheimer's disease and other types of dementia were the most frequently studied conditions and the EQ-5D was the most common instrument used to measure caregiver/family member utility. Although, as discussed in the stage 2 report, whether the EQ-5D is the most appropriate instrument to measure carer QoL is still to be debated. Furthermore, most of the studies were conducted in in Europe, the US and Canada, making generalisations to these settings most comparable. The review critiqued here also highlighted the importance of using studies with a comparator, and that further work is needed to generate standard definitions of caregiving when incorporating carer and family quality of life into cost effectiveness analyses, and to address equity issues and potential double counting in analyses.

This is an emerging field, the review is useful to as it documents the entire literature at present. As the body of literature grows a review including an assessment of quality, and further analysis of the data reported in each study will be required.

5. REFERENCES

1. National Institute of Health and Care Excellence . Guide to the methods of technology appraisal 2013.
2. Wittenberg E, Prosser LA. Disutility of illness for caregivers and families: a systematic review of the literature. *Pharmacoeconomics*. 2013;31(6):489-500.
3. Wittenberg E, James LP, Prosser LA. Spillover Effects on Caregivers' and Family Members' Utility: A Systematic Review of the Literature. *Pharmacoeconomics*. 2019;37(4):475-99.
4. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017;358:j4008.
5. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *J Clin Epidemiol*. 2016;75:40-6.
6. NIHR Centre for Reviews and Dissemination. CRD Database 2019 [Available from: <https://www.crd.york.ac.uk/CRDWeb/>]
7. International Society for Pharmacoeconomics and Outcomes Research. ISPOR Presentations Database 2019 [Available from: <http://www.ispor.org/heor-resources/presentations-database/search>].
8. International Society for Quality of Life Research. International Society for Quality of Life Research Abstracts Presented at the Annual Conference of the International Society for Quality of Life Research 2019 [Available from: <https://link.springer.com/journal/volumesAndIssues/11136>]
9. Lefebvre C, Glanville J, Briscoe S, Littlewood A, Marshall C, Metzendorf M-I, Noel-Storr A, Rader T, Shokraneh F, Thomas J, Wieland LS. Chapter 4: Searching for and selecting studies. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.0 (updated July 2019). Cochrane, 2019. Available from www.training.cochrane.org/handbook.
10. Arber M, Garcia S, Veale T, Edwards M, Shaw A, Glanville JM. PERFORMANCE OF OVID MEDLINE SEARCH FILTERS TO IDENTIFY HEALTH STATE UTILITY STUDIES. *Int J Technol Assess Health Care*. 2017;33(4):472-80.
11. Carroll C, Scope A, Kaltenthaler E. A case study of binary outcome data extraction across three systematic reviews of hip arthroplasty: errors and differences of selection. *BMC Res Notes*. 2013;6:539.
12. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:i4919.

APPENDIX

A.1 DATA EXTRACTION FROM THE SYSTEMATIC REVIEW

<p>Authors (date)</p>	<p>Eve Wittenberg, Lyndon P. James, Lisa A. Prosser (2019)</p>
<p>Searching methods</p>	<p>Three electronic databases: MEDLINE, EMBASE, and EconLit were searched from the inception of each database through 3 April 2018. Search terms were refined by testing them against a set of ‘known-to-us’ papers to ensure capture of relevant articles. The final search strategy combined terms describing utility measures with terms describing caregivers, family members, and burden: utility, disutility, preference weight, QALY, standard gamble, time trade-off, EuroQoL (EQ-5D), Short-Form 6-Dimension (SF-6D), Health Utilities Index (HUI), Quality of Wellbeing Scale (QWB), CarerQoL, Carer Experience Scale (CES), Child Health Utility-9 dimensions (CHU-9D), and variants thereof; spillover, caregiver, family, partner, spouse, child, sibling, parent, grandparent, next of kin, burden, consequence, and associated variants.</p>
<p>Study selection methods</p>	<p>Inclusion criteria and exclusion:</p> <ul style="list-style-type: none"> - Included peer-reviewed articles published in English that reported a preference-based measure of caregiver or family member utility or disutility, including caregiver-focused measures for which population tariffs exist (i.e., CarerQoL and CES). - Included articles that reported on multiple patient diseases and/or using multiple preference-based methods/instruments, - Excluded articles that reported only the EQ-VAS or a visual analog scale measure unless the scores were transformed into utilities using a known algorithm. - Included articles on family members - defined as anyone

identified as having a familial relationship to the patient regardless of distance (e.g., cousins would meet our inclusion criterion). Assumed that all family members classified in articles as caregivers were such.

- Included articles reporting on 'informal caregivers' unless they were described as exclusively non-familial, e.g. neighbors, church members.
- Excluded paid caregivers.
- Included articles reporting on all patient diseases and conditions, including those that specified no disease, meaning they included caregivers regardless of the patient's disease.
- Included articles reporting on patient health states, defined as a distinct phase of a disease or condition (such as chemotherapy or hospitalization); - a disease was defined as a diagnosed condition.
- Excluded disease transmission among family members from our definition of spillover effects.
- Included death as a health state when it was directly related to a disease or condition, such as maternal mortality, but did not specifically search for bereavement.
- Included articles reporting on studies specifically designed to measure spillover utility, those measuring caregiver/family member utility among other outcomes, and caregiver or patient interventions that included utility as an outcome.
- Excluded reviews, reports, study protocols, commentaries, editorials, and conference papers, as well as articles that reported what appeared to be invalid utilities, such as scores >1.0 or those described as 'WTP utilities'.

Limits

- No age limit on patients or caregivers/family members.

After excluding duplicates, two authors (EW and LJ) independently

	<p>screened titles, abstracts and full text articles; conflicts were resolved by consensus.</p> <p>Reasons for exclusions were recorded using Covidence systematic review software (Veritas Health Innovation, Melbourne, VIC, Australia).</p>
<p>Data extraction and quality assessment methods</p>	<ul style="list-style-type: none"> - Data was extracted that would allow a reader to identify potentially useful values for an analysis: patients' disease/condition; patients' age (adult/child/either); valuation measure used (EQ-5D, standard gamble, etc.); sample source (e.g., medical centers, patient association, population); country of sample; affected person's role (i.e., family member/family caregiver/informal caregiver); caregiver/family member age (mean or other summary measure); sample size; utility (mean or median); if relevant to study design: comparison group source, sample size, and utility (mean or median); and if relevant, the reporting of utilities by strata, and other notes. - A table entry was created for each patient disease/condition for which a relevant utility was reported in an article; articles that reported utility for more than one patient disease/condition were included in an entry for each. Multiple utilities measured using different methods (e.g., HUI2 and HUI3) or applying different valuation weights for the same measure (e.g., Canadian and US weights for SF-6D) were included in one entry. - If both caregiver/family member utilities and spillover disutilities were reported, we included each. - If utilities were reported for the same condition/disease for multiple countries, we reported the one with the largest country-specific sample size. - For all other instances of multiple utilities reported for the same disease/condition, we included those we deemed most salient to most readers and noted the availability of others in

	<p>the 'notes' comment.</p> <ul style="list-style-type: none"> - We included the scores/values as reported by authors, but performed no manipulations or calculations on reported data. - Entries were grouped into three categories: (1) spillover utilities or disutilities; (2) caregiver and/or family member utility reported with a matched or population comparison group; and (3) caregiver and/or family member utility reported alone.
<p>Number and type of studies included</p>	<p>The search yielded 5205 records. After duplicates, 3695 remained and were screened by title and abstract. 177 full-text articles were assessed for eligibility; 80 articles remained for inclusion. Of these;</p> <p>8 (10%) reported spillover utility/disutility: 4 reported spillover disutility as the difference between population utility and the observed family caregiver utility, 1 reported disutilities only, 1 reported both the difference between the observed caregiver utility and the population utility, as well as a utility for a hypothetical scenario in which the ill relative did not need caregiving, and 2 reported spillover utilities only, elicited using a direct method to isolate the spillover effect per se.</p> <p>Twenty-five (30%) reported a comparison group, either general population norms (n=9; 3 of which also reported disutility) (Table 1) or matched, non-caregiver/family members or hypothetical scenarios' utilities (n=16) (Table 2).</p> <p>Fifty (63%) reported caregiver/ family member utilities only (Table 3).³ Some articles reported utilities for multiple conditions or using multiple measurement methods, or for multiple strata of caregivers/family members.</p>
<p>Characteristics of the</p>	<p>Across all 80 articles,</p> <ul style="list-style-type: none"> - Alzheimer's disease and other types of dementia were the

<p>included studies</p>	<p>most frequent focus (15 articles), followed by cancer (6 articles) (Tables 1, 2, 3).</p> <ul style="list-style-type: none"> - Over half of the studies focused on caregivers/family members of ill adults (47, or 59%), 14 on ill children (18%), and the remainder focused on adults and children combined. - The EQ-5D was the most common instrument used to measure caregiver/family member utility (58, or 69%, of uses among 84 in total; some articles reported multiple measurement methods). Indeed, 95% of articles used generic (i.e., indirect) measurement instruments: the SF-6D was used in 13 instances (16%), and the HUI and QWB were used three and two times, respectively. The caregiver-focused instruments (the CarerQoI and CES) were used in seven instances (six uses and one use, respectively; 9%). Six articles (8%) reported caregiver/family member utility in the context of a patient and/or caregiver intervention trial. - Most spillover effects research has been conducted in Europe (53 articles, 66%), followed by the US and Canada (20 articles, 25%). - The earliest article reporting on this topic was published in 1988; nearly half (49%) were published between 2015 and 2018 (Tables 1, 2, 3).
<p>Author conclusions</p>	<p>The scope of CEAs is expanding from patient-based analyses to caregiver/patient dyadic and family-based analyses. While this expansion is consistent with theoretical principles of maximizing health benefits, prevailing methodological consensus, and demographic and health system changes, it raises practical challenges for CEA and highlights data gaps. It is likely, at least for the time being, that QALYs are here to stay. Caregiver and family member spillover effects will therefore be primarily measured in QALYs and will consequently require utilities. This review provides a catalogue of utilities to facilitate the calculation of QALYs and inform</p>

	<p>CEAs. Additional research is needed on methods of measuring and incorporating spillover QALYs to promote, among other things, an accurate reflection of societal preferences for caregiver/family effects relative to patients' effects. It is our goal to advance the inclusion of spillover in CEA by providing this accessible overview of the spillover effects of HRQOL literature. We also aspire to expand the knowledge base of spillover-based CEAs, from which we will answer these remaining questions.</p>
<p>Limitations and uncertainties</p>	<p>While our search is comprehensive as of our end date, articles including utilities for spillover effects are being published with increasing frequency, and will soon render our catalogue incomplete. We are in the process of developing an online, open-access repository of spillover effect utilities, which will be updated regularly as a public resource. Moreover, our catalogue excludes the gray literature or unpublished sources. Unpublished utilities that are in the pipeline, via conference presentations and abstracts, will likely find their way into the published literature in the future and will be incorporated into successive versions of the catalogue.</p> <p>Publication bias is not a concern for this review because our results are descriptive and not intended for inference. We limited the data included in our tables to ensure accessibility for readers—essentially a size that was viewable on a standard size page or computer screen—therefore details that are important to some may have been omitted. Finally, we made subjective judgments about the relative salience of utilities in articles reporting multiples, but describe others in the 'notes' section of the tables.</p>

A.2 WITTENBERG, JAMES & PROSSER (2019) SEARCH STRATEGIES

Concept	MEDLINE (2,615 records)	Embase (2,254 records)	EconLit (321 records)
Utility	utilit* or disutilit* or preference-based or quality	(utilit*:ab,ti OR disutilit*:ab,ti OR 'preference-based':ab,ti	ti(((utilit*) or (disutilit*) or ("preference-based") or

	<p>adjusted life year* or qaly or qalys or sf 6d or sf6d or eq-5d* or eq5d* or euroqol* or carerqol* or ces or carer experience scale or hui or health utilities index or quality of wellbeing scale or quality of well-being scale or qwb or standard gamble or time tradeoff or time trade-off or visual analog or visual analogue or vas or chu-9d or (quality of life and economic*) or preference weight* or EQVAS or preference score* or economic evaluation*).ab,ti.</p>	<p>OR 'quality adjusted life year*':ab,ti OR qaly:ab,ti OR qalys:ab,ti OR 'sf 6d':ab,ti OR 'sf6d':ab,ti OR 'eq-5d*':ab,ti OR eq5d*:ab,ti OR euroqol*:ab,ti OR carerqol*:ab,ti OR ces:ab,ti OR 'carer experience scale':ab,ti OR hui:ab,ti OR 'health utilities index':ab,ti OR 'quality of wellbeing scale':ab,ti OR 'quality of well-being scale':ab,ti OR qwb:ab,ti OR 'standard gamble':ab,ti OR 'time tradeoff':ab,ti OR 'time trade-off':ab,ti OR 'visual analog':ab,ti OR 'visual analogue':ab,ti OR vas:ab,ti OR eqvas:ab,ti OR 'chu-9d':ab,ti OR 'preference weight*':ab,ti OR 'preference score*':ab,ti OR 'economic evaluation*':ab,ti OR ('quality of life':ab,ti AND economic*:ab,ti) AND [embase]/lim</p>	<p>("quality adjusted life year*") or (qaly) or (qalys) or ("sf 6d") or (sf6d) or ("eq-5d*") or (eq5d*) or (euroqol*) or (carerqol*) or (ces) or ("carer experience scale") or (hui) or ("health utilities index") or ("quality of wellbeing scale") or ("quality of well-being scale") or (qwb) or ("standard gamble") or ("time tradeoff") or ("time trade-off") or ("visual analog") or ("visual analogue") or (vas) or (eqvas) or ("chu-9d") or ("preference weight*") or ("preference score*") or ("economic evaluation*")) or (("quality of life") and (economic*)) or ab(((utilit*) or (disutilit*) or ("preference-based") or ("quality adjusted life year*") or (qaly) or (qalys) or ("sf 6d") or (sf6d) or ("eq-5d*") or (eq5d*) or (euroqol*) or (carerqol*) or (ces) or ("carer experience scale") or (hui) or ("health utilities index") or ("quality of wellbeing scale") or ("quality of well-being scale") or (qwb) or ("standard gamble") or ("time tradeoff") or ("time trade-off") or ("visual analog") or ("visual analogue") or (vas) or</p>
--	--	--	---

			(eqvas) or ("chu-9d") or ("preference weight*") or ("preference score*") or ("economic evaluation*")) or (("quality of life") and (economic*))
Spillover	exp caregivers/ or (spillover* or spill over*).ab,ti. or ((carer or carers or caring or caregiver* or caregiving or family or partner* or husband or wife or spouse or spouses or spousal or child or children or parent* or mother* or maternal or father* or paternal or sibling* or brother* or sister* or grandparent* or grandmother* or grandfather* or close person* or next of kin) adj4 (effect* or burden* or impact* or consequence*)).ab,ti.	('caregivers'/exp OR spillover*:ab,ti OR 'spill over*':ab,ti OR (((carer OR carers OR caring OR caregiver* OR caregiving OR family OR partner* OR husband OR wife OR spouse OR spouses OR spousal OR child OR children OR parent* OR mother* OR maternal OR father* OR paternal OR sibling OR brother* OR sister* OR grandparent* OR grandmother* OR grandfather* OR 'close person' OR 'next of kin') NEAR/4 (effect* OR burden* OR impact*)):ab,ti)) AND [embase]/lim	su((caregivers) or (spillover)) OR ti((spillover*) OR ("spill over*") OR (((carer) OR (carers) OR (caring) OR (caregiver*) OR (caregiving) OR (family) OR (partner*) OR (husband) OR (wife) OR (spouse) OR (spouses) OR (spousal) OR (child) OR (children) OR (parent*) OR (mother*) OR (maternal) OR (father*) OR (paternal) OR (sibling) OR (brother*) OR (sister*) OR (grandparent*) OR (grandmother*) OR (grandfather*) OR ("close person") OR ("next of kin")) NEAR/4 ((effect*) OR (burden*) OR (impact*))) or ab((spillover*) OR ("spill over*") OR (((carer) OR (carers) OR (caring) OR (caregiver*) OR (caregiving) OR (family) OR (partner*) OR (husband) OR (wife) OR (spouse) OR (spouses) OR (spousal) OR (child) OR (children) OR (parent*) OR (mother*) OR (maternal) OR (father*) OR (paternal) OR (sibling) OR (brother*) OR (sister*) OR

			(grandparent*) OR (grandmother*) OR (grandfather*) OR ("close person") OR ("next of kin") NEAR/4 ((effect*) OR (burden*) OR (impact*)))
Limits		NOT ('conference abstract'/it OR 'conference paper'/it OR 'conference review'/it)	