Draft Methods Research Report

Number XX

Progression-Free Survival:

What Does it Mean for Psychological Well-being or Quality of Life?

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. xxx-xx-xxxx

Prepared by:
<Name> Evidence-based Practice Center
<City, State>

Investigators:
First and Last Names, X.X.
First and Last Names, X.X.

AHRQ Publication No. xx-EHCxxx
<Month Year>

This report is based on research conducted by an Evidence-based Practice Center under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. xxx-xxxx-xxxxx-I). The findings and conclusions in this document are those of the author(s), who are responsible for its content, and do not necessarily represent the views of AHRQ. No statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help clinicians, employers, policymakers, and others make informed decisions about the provision of health care services. This report is intended as a reference and not as a substitute for clinical judgment.

This report may be used, in whole or in part, as the basis for the development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products or actions may not be stated or implied.

This document is in the public domain and may be used and reprinted without special permission. Citation of the source is appreciated.

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

Suggested citation: Author I, Author I. <Title>. Methods Paper No. xx. (Prepared by the <Name> Evidence-based Practice Center under Contract No. xxx-xx-xxxx.) AHRQ Publication No. xx-EHCxxx. Rockville, MD: Agency for Healthcare Research and Quality. <Month Year>. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-Based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To improve the scientific rigor of these evidence reports, AHRQ supports empiric research by the EPCs to help understand or improve complex methodological issues in systematic reviews. These methods research projects are intended to contribute to the research base in and be used to improve the science of systematic reviews. They are not intended to be guidance to the EPC program, although may be considered by EPCs along with other scientific research when determining EPC program methods guidance.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers, as well as the health care system as a whole, by providing important information to help improve health care quality. The reports undergo peer review prior to their release as a final report.

We welcome comments on this Methods Research Project. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by e-mail to epc@ahrq.hhs.gov.

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director
Evidence-based Practice Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Elise Berliner, Ph.D.
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Acknowledgments

The authors gratefully acknowledge the following individuals for their contributions to this project:

Key Informants

<Name>
<Place>
<City>, <ST>

Technical Expert Panel

<Name>
<Place>
<City>, <ST>

<Name>
<Place>
<City>, <ST>

Peer Reviewers

<Name>
<Place>
<City>, <ST>

<Name>
<Place>
<City>, <ST>

Progression-Free Survival: What Does it Mean for Psychological Well-being or Quality of Life?

Structured Abstract

Background: Progression-free survival (PFS), defined as the time from random assignment in a clinical trial to disease progression or death from any cause, has recently become an endpoint of considerable interest in the study of new oncology drugs. In comparison to overall survival (OS), the gold standard for cancer drug evaluation, PFS can be evaluated using shorter, smaller and less costly studies. Its use as a primary endpoint, however, can be challenging, as it is subject to a wide range of potential biases, and its use as a surrogate for OS has been demonstrated only for certain disease and treatment scenarios. The objective of this methods project is to address whether PFS is an outcome related to psychological well-being or QoL.

Methods: Two Key Questions (KQ) were posed: (1) when PFS is used as a primary clinical endpoint in treating patients with advanced cancer, is there direct evidence that knowing PFS impacts patient anxiety, depression, or psychological well-being, and (2) for agents where PFS is the primary outcome measure being used to establish the performance (efficacy and safety) of a new drug, what evidence exists on the association of PFS with QoL and related outcomes, such as disease symptoms? A Technical Expert Panel (TEP) was convened to refine the KQs, comment on the methodological approach, and identify publications. The literature search for KQ1 sought to identify studies that showed a causal relationship between PFS and improvement on measures of psychological well-being. The search was not limited by tumor type or study design. KQ2, which addressed the association between PFS and QoL was designed to indirectly answer KQ1, as a low yield of relevant articles was expected from the initial search. The literature search for KQ2 included terms for drugs approved by one or more regulatory agencies on the basis of PFS outcomes for treatment of solid tumor disease between 2005 and 2010, including Avastin, Ixempra, Tykerb, Vectibix, Doxil, Gemzar, Yondelis, Nexavar, Votrient, Sutent, Tarceva, Taxotere (Appendix B). Both KQ searches were conducted in MEDLINE®, PubMed, EMBASE®, and the Cochrane Central Register of Controlled Trials (CENTRAL). From relevant articles, information was abstracted and summarized in data tables on study characteristics, treatment efficacy and safety, patient-reported outcome measure descriptions, and the PFS/QoL association. A quality assessment of the included individual studies was conducted to identify potential biases in the measurement of either PFS and/or QoL.

Results: No studies were identified that addressed KQ1. There was no direct evidence demonstrating that knowing PFS status impacts patient anxiety, depression, or psychological well-being. KQ2 sought to determine an association between PFS and QoL or related outcomes, such as disease symptoms, and four studies were identified that provided such evidence. The four studies reported an improvement in QoL or disease symptoms among patients who remained progression-free compared with those who had disease progression. Study quality limitations resulted in a poor quality rating for all four studies, and the strength of the evidence was insufficient. Common study limitations included significant data missing not at random, failure to evaluate patient-reported outcomes beyond the window of PFS, and lack of patients and investigator blinding to treatment.

Discussion: The goal of this methods project was to determine what evidence exists to support PFS as a primary clinical endpoint, based on its ability to impact psychological well-being or based on its association with QoL or related outcomes, such as disease symptoms. It focused on the relationship between PFS, an *outcome*, and other *outcomes* of importance to patients. In contrast to a traditional comparison of *interventions*, the variable of interest, PFS, is somewhat problematic, because the presence or absence of progression can only be observed, precluding the ability to design a prospective randomized clinical trial. Defining the relationship between PFS and PRO involves considering data as if obtained from observational study. A hypothetical framework is provided, including potentially important causal relationships, for considering further study of the role of PFS as an outcome of interest to patients.

Evidence Gaps: There is a need for prospective research to evaluate: 1) the causal relationship between patient knowledge of PFS status with QoL; 2) patient impressions of the meaningfulness of PFS as an outcome in the absence of association with OS; and 3) the extent to which improvement in QoL measures associated with improvement in PFS is related to a common underlying mechanism.

Conclusion: There is insufficient evidence to make any conclusion about the association between PFS and improvement in QoL or disease symptoms. In cases when measurement of OS is impractical, the direct measurement of both PFS and QoL may be an alternative strategy.

Contents

Background **1**
Overall Survival 1
The Growing Interest in PFS 1
Relationship of PFS to OS 2
The Issue of Surrogacy 2
PFS as a Surrogate for OS 2
PFS as an OS Surrogate for Specific Cancers 2
Issues in the Measurement and Reporting of PFS 3
Assessment Bias 3
Evaluation Bias 4
Performance Bias 4
Attrition Bias 5
Detection Error 5
PFS as a Health Outcome 6
Methods **7**
Technical Expert Panel 7
Key Questions 7
KQ1, PFS as an Outcome of Importance to Patients 7
KQ2, Associations of PFS with Outcomes of Importance to Patients 7
Eligibility Criteria 8
Literature Search 9
Published Literature 9
Gray Literature 10
Citation Screening 10
Data Extraction 11
Data Synthesis 11
Quality Assessment of Individual Studies 11
Strength of Evidence 12
Results **13**
Searches of Published Literature 13
Searches of Gray Literature 14
KEY QUESTIONS 14
Key Question 1 14
Key Question 2 14
Discussion **23**
Limitations 26
Evidence Gaps 26
Conclusion 26
References **27**

Tables

Table 1. Progression-Free Survival, Inclusion/Exclusion Criteria for KQ1 8
Table 2. Progression-Free Survival, Inclusion/Exclusion Criteria for KQ2 9
Table 3. Progression-Free Survival, Definitions of Overall Study Quality Ratings 12
Table 4. Progression-Free Survival, Study Characteristics 17
Table 5. Progression-Free Survival, Treatment Efficacy and Safety 18
Table 6. Progression-Free Survival, QoL Measurement Description 19
Table 7. Progression-Free Survival, QoL Measurement – Missing Data 20
Table 8. Progression-Free Survival, PFS/QoL Association 21
Table 9. Progression-Free Survival, Summary of Individual Article Quality Assessments 22

Figures

Figure 1. Progression-Free Survival, PRISMA Flow Diagram of Study Selection 13
Figure 2. Basic causal model for treatments, progression, and outcomes 25

Appendices

Appendix A. Definitions of Additional Oncology Drug Endpoints
Appendix B. PFS Search Strategy
Appendix C. List of Drugs Approved Using Studies with a Primary Endpoint of PFS
Appendix D. Screening Questions
Appendix E. Abstraction Form Items
Appendix F. Relevance of Quality Assessment Items
Appendix G. Articles Excluded at Full-Text Screening

Background

Progression-free survival (PFS) is defined as the time from random assignment in a clinical trial to disease progression or death from any cause. PFS as an outcome is of interest to a variety of disciplines, most especially, for purposes of this project, to oncologists, pharmacologists, trialists, social scientists, and other scientists with interest in designing or interpreting clinical trials. The objective of this methods project was to address whether PFS is an outcome related to psychological well-being or quality of life (QoL). The use of PFS in the study of new oncology drugs, with a particular focus on advanced solid epithelial tumors was examined. This background section addresses how PFS is used, its role as a surrogate for overall survival (OS), the challenges it presents in obtaining accurate and reproducible measurements, and finally its role as a health outcome. The literature on PFS as a predictor of overall survival is far more extensive than that on use of PFS as an indicator of QoL or related patient-reported outcomes. Thus an overview of the uses and limitations of PFS as a predictor of OS informs the topic of this paper: the relationship of PFS to psychological well-being, QoL, or other outcomes of interest to patients, such as disease symptoms.

Overall Survival

This section briefly reviews the use of OS as a standard outcome, and the reasons why other survival outcomes, such as PFS, have garnered interest from the clinical research community. OS has long been considered by the Food and Drug Administration (FDA) and the European Medicines Agency (EMEA) as the gold standard for the evaluation of new oncologic therapies.[1] It is defined as the time from random assignment to the date of death due to any cause, or to the date of censoring at the last time the subject was known to be alive in intention-to-treat populations. OS is "an unambiguous endpoint measure because it is evaluated on a continuous time scale, which gives precise accuracy for the time of the event."[2]

However, the use of OS can be challenging. For example, if survival is only incrementally improved by a new treatment, the demonstration of increased OS may require large patient populations, several years of accrual and follow-up, and higher costs.[3, 4] This is especially true if the natural history of the disease course is lengthy.

The Growing Interest in PFS

Over the past 10 years there has been increasing interest in the use of outcomes other than OS to study new drugs, including PFS. The interest in PFS stems in part from the challenges associated with OS as an endpoint, but it has also been fueled by the fact that many new drugs are targeted toward molecular mechanisms of action that are cytostatic rather than cytotoxic. These drugs are not expected to provide the same objective response rates of earlier drugs, and instead act to prevent progression rather than cause tumors to shrink, disappear, and thereby impact mortality. Interest in PFS has also been sparked by the increasingly common use of treatment paradigms that allow for multiple rounds of drug treatment (first-, second-, third-, and even fourth-stage therapies), each producing incremental changes difficult to capture in the context of a single study using OS as the primary endpoint. In contrast, PFS can be studied in the short-term context of each treatment, without the confounding influence of the next.

While this methods project focuses specifically on PFS, it is recognized that there is widespread interest in a number of alternative endpoints, including disease-free survival, relapse-

free survival, time to progression, and objective response rate. For informational interest, definitions of these can be found in Appendix A.

Relationship of PFS to OS

The Issue of Surrogacy

The term "clinical endpoint" has been defined by the Biomarkers Definition Working Group as an outcome which measures how a patient feels, functions, or survives. The term "surrogate endpoint" is an outcome measure that has been validated as an adequate substitute for the clinical endpoint. Ideally, identification of a surrogate endpoint in a drug study provides a reliable signal that the clinical endpoint of the study has been met. Surrogate endpoints may be laboratory variables, single measures of disease activity (recurrence, progression, etc.), or composite measures of disease activity. It is important to note that the validation of a surrogate endpoint requires evidence that goes beyond merely showing a correlation between the surrogate and clinical endpoints. As noted by Shi and Sargent,[5] as a guiding principle, "the treatment effect observed on a valid surrogate endpoint (substitute) should reliably and precisely predict the treatment effect on the clinical endpoint (entity being replaced)."

PFS as a Surrogate for OS

Considerable interest has been focused on the use of PFS as a surrogate endpoint for predicting OS. A number of methodologies have been used to evaluate the relationship between these two parameters, including use of hypothesis testing,[6] estimation and prediction,[7, 8] and meta-analytical approaches.[5] It is now recognized that the correlation between PFS and OS is both variable and unpredictable and depends on tumor type and tumor stage, as well as the particular drug being investigated.[5] That PFS is not always a reliable surrogate for OS is not entirely surprising, given that the tumor pathways affected by new drugs and the nature of drug and tumor interaction, as well as drug toxicity, are often incompletely known.

Broglio and Berry[9] have recently performed simulation studies partitioning OS into two parts, the first PFS, and the second what they call survival post-progression (SPP). They defined SPP as OS minus PFS. Using preset 6- and 9-month medians for PFS in each arm of a hypothetical two-arm study, they concluded that a statistically significant increase in OS was detected with 90 percent probability if median SPP was 2 months, but less than 20 percent if median SPP was 24 months. They recommended OS be used as a primary endpoint only when median SPP is short. These conclusions are confirmed by Amir et al.,[10] who evaluated 26 studies of chemotherapy for solid tumors in which a hazard ratio was reported for both OS and PFS (or time to progression, related to PFS, see definition, Appendix A). They also found a higher correlation between OS and PFS when SPP is short than when SPP is long. However, even in instances in which SPP was less than 12 months, they identified only a moderate correlation coefficient of 0.64 between PFS and OS.

PFS as an OS Surrogate for Specific Cancers

Efforts to establish PFS as a surrogate for OS in oncology trials have had variable results depending on the specific cancer. For example, several studies have shown that PFS is a valid surrogate for OS in colorectal cancer,[11-14] and it has been argued that PFS is a reasonable primary endpoint for the disease on its own merit.[15, 16] Similar conclusions have been reached about PFS

as a surrogate for OS in first-line therapy for ovarian cancer.[17-19] Expert panelists at two major workshops agreed, however, that the PFS to OS relationship with regard to ovarian cancer may be different for different patient groups or for first-line compared with second- or third-line therapy.[17, 19]

Contrary to the success of PFS as a surrogate endpoint in first-line treatment of colorectal and ovarian cancer, a strong relationship between PFS and OS has not been demonstrated in studies of metastatic breast cancer.[4, 11, 13, 20]

Depending on the toxicity of a new drug that has been found to increase PFS, it is possible to postulate scenarios in which treatment accelerates both psychological and physical morbidity, resulting in decreased patient QoL. In the worst case scenario, use of a drug to increase PFS may have the unanticipated downside of actually compromising the balance between tumor and patient resistance to tumor, causing a shorter rather than a longer duration in OS. As witnessed in the recent FDA decision to remove the indication for use of bevacizumab (Avastin®, Genentech/Roche) in breast cancer, there are strong feelings about both the use and interpretation of PFS, as well as differing opinions on the risk-to-benefit value of drug efficacy and toxicity.

Issues in the Measurement and Reporting of PFS

There are several potential sources of measurement bias or variability in studies using PFS as a primary endpoint. Potential for bias should be addressed prospectively in trial designs to ensure the validity of any differences in PFS found between treatment arms.[3, 21-24] This section describes the main sources of potential bias, as well as suggested mechanisms for controlling their impact. In addition to discussing four major sources of bias, including assessment, evaluation, performance, and attrition, the role of detection error is also described. Like bias, this measurement issue can lead to incorrect conclusions about the performance of drugs.

Assessment Bias

The exact date of progression cannot be known, since it is determined based on the types and timing of assessments. At the point in time progression is identified, it is only known that this event occurred at some point between the last negative evaluation and the one at which this reclassification of disease status occurs. In general, the date of first progression is taken as the date of the evaluation at which progression was first evident, which is likely to be an overestimate of PFS. As Panageas et al.,[25] have recently noted, in a trial this overestimation of median PFS can lead to erroneous conclusions about new treatments, suggesting benefits that in fact may not actually exist.

Use of the last date the patient was identified to be progression-free or an intermediate interval, such as the midpoint between the two dates, has been considered as possible alternative or additional mechanisms for reporting PFS. The former may underestimate PFS and the latter, like the date of first identification, likely overestimates PFS. According to Panageas et al.,[25] what is important in capturing an accurate PFS measurement is the timing of the measurement interval in relationship to the true median PFS. They suggested PFS be characterized using interval reporting in which estimates of this event are characterized by the time interval in which they occur. Zhuang et al.,[24] recommended that the assessment interval not exceed the expected improvement in median PFS in the experimental versus control arm.

Freidlin et al.[26] recommended the use of two preselected scan time points with strictly chosen schedule limits, instead of multiple regular testing intervals. For optimal evaluation of performance, they suggested the selected time points represent the median PFS and twice the

median PFS expected in the control arm of a study, and that a significance test of the difference in PFS rates at the two scan time points be assessed on the grouped data.

Evaluation Bias

Another timing bias has to do with unevenness in the timing of tumor assessment between the two treatment arms. This can result in progression being identified earlier in one arm than the other, even when there is no actual difference in efficacy.[3, 24] Asymmetry can result when assessments are scheduled around the treatment cycle and one arm has more cycle delays than the other, or when there is disparity in unscheduled or missed visits between the two treatment arms. Small treatment-related differences in measurement time (as short as 2 days) have been reported to result in false study conclusions.

For a reliable measurement, patients must be evaluated on a regular and balanced basis across treatment arms. While most well-designed comparative studies address this issue, they should still be monitored for asymmetry. It is important to report and analyze progression events confirmed at preplanned time points and at unscheduled visits. One statistical technique for assessing this form of bias is to perform sensitivity analyses to examine the strength of a positive result in a clinical trial relative to the sources of bias.[27]

Performance Bias

The patient's response to treatment or progression status may be influenced by knowledge of the treatment arm.[3, 28, 29] Physician or patient bias that the experimental drug is more beneficial than the control therapy may result in earlier identification of progression in the control arm than in the experimental arm in an effort to assure patients who have progressed receive the perceived best therapeutic choice. Alternatively, this bias may result in a tendency to underestimate progression in the experimental arm in an effort to maximize the opportunity for patients to benefit from a new treatment perceived as being superior to control. Both scenarios could lead to overestimation of the PFS value for the drug under study.

The ideal mechanism for addressing performance bias is to perform a double-blinded study. Unfortunately, because of differences in drug administration or in the toxic profiles of treatments, blinding is not always possible. One mechanism for addressing bias in local evaluations, particularly in studies using standardized radiologic endpoints, is use of blinded independent central review (BICR).[3, 23, 28, 29] Radiological images being evaluated are blindly and independently reviewed by an outside centralized, often expert, group of readers. Conditions for reading are standardized as much as possible (e.g., images are evaluated serially to assure changes from baseline are carefully tracked).

Although recommended in regulatory guidance,[30] BICR has generated some controversy because of its complexity and cost.[23, 28, 31] In addition, depending on the timing of central review, discrepancies in evaluation of progression between local and central evaluations can lead to informative censoring (i.e., removal from the study of patients with early local evaluator calls who are not confirmed by central review), leading to potential bias. Suggestions to address this problem include performing BICR in real time and feeding back results to local study sites or designing studies to allow for continued evaluation of progression for at least one scan after local progression is called.

Tang et al.[32] studied eight trials using PFS as an endpoint comparing results of local evaluations with BICR. They concluded that although benefits of treatment could be quite variable (-2 to 2.4 months), there was no evidence of systemic bias. Amit et al.,[33] in a meta-

analysis of 27 blinded studies with independent central review of progression performed by the Pharmaceutical Research and Manufacturers Association (PhRMA), reported a strong correlation between local evaluation and BICR. They concluded that when studies are blinded, and/or when large study size effects are anticipated, a BICR might not be warranted. They described a sample-based approach to BICR that defines early and late discrepancy rates between local and central review. Although they did not define a threshold for differential discordance, they suggested this discordance be used in decision-making about whether a full BICR is necessary. Unfortunately, it appears that small, but potentially significant differences (up to 15 percent discordance) are not detected by their approach.[31]

Attrition Bias

When too many patients withdraw from a study or are lost to follow-up, and when losses are not at random, remaining results may be biased. This is especially problematic when attrition is greater in one arm than the other.[3, 24] PFS data are censored at the time of last available assessment, so the proportion of censored patients should be reported for both treatment arms, along with the reasons for censorship. Short assessment intervals and ongoing physician and patient education regarding the goal of treatment have been found to help minimize patient withdrawal and loss to follow-up.[24] Sensitivity analysis can be performed to look at various subgroups of patients subject to attrition bias, as well as the effect of attrition bias in total, to assess the impact of this bias on study conclusions.[27]

Detection Error

PFS is most commonly a composite endpoint including radiologic progression, death, and in some cases, nonradiologic criteria, such as symptomatic progression. Dancey et al.,[3] have identified four criteria to establish progression: the appearance of new radiologic lesions; an increase in the size of measurable target lesions; a clear, unequivocal increase in nontarget disease; and/or worsening of nonradiologic signs and symptoms.

There are numerous caveats associated with these criteria. New radiologic lesions, for example, must be unequivocal and significant enough in size to avoid the measurement error in the methodology being used. Oxnard et al.,[34] recently studied 30 patients with non-small cell carcinoma of the lung (one cm or larger in size), undergoing two computed tomography (CT) scan evaluations within a 15-minute interval. All scans were read side-by-side by three radiologists. Measurement changes were within ± 10 percent for 84 percent of measurements. Changes of 20 percent or more were observed in 3 percent of measurements. They concluded that CT scan measurement of lung lesions has clinically meaningful variability and suggest caution in the interpretation of small changes in lesion size in the care of individual patients and in the interpretation of clinical trial results.

Identification of new lesions should unequivocally demonstrate a metastatic deposit. In order to be certain this is the case, baseline anatomic scanning is required to detect the presence of disease in all areas likely to be the site of metastases based on what is known about the tumor being evaluated.

Most problematic is establishing progression in disease that is detectable, but not measurable. Assessing a worsening of disease burden, such as disease-related symptoms, from a nonmeasurable baseline, is a largely subjective determination. Efforts should be directed at creating an operational definition of progression of non-measurable disease. If changes indicating disease progression are equivocal, Dancey et al.,[3] recommend, when medically

possible, that the patient remain on study until progression is unequivocal. At that time, a decision would need to be made as to whether the progression date is backdated to the first equivocal finding or recorded as the date of the unequivocal determination.

Of note, measurement variability will generally not lead to study bias since it occurs in all treatment arms of the study. It is, however, likely to lead to failure to identify changes in disease status that result from use of a new drug.

PFS as a Health Outcome

Many advocates for the use of PFS as an endpoint contend that delaying tumor progression independently confers clinical benefit, since improvement in PFS is considered an indication of disease control and stabilization. A direct result should be stability, and perhaps even reduction, in disease symptoms, thus improving QoL for patients. While in PFS, patients are spared the symptoms of progressive disease, from undergoing further treatment with additional therapies and their attendant toxicities, and from the psychological burden and uncertainty associated with disease progression.[35] In this scenario, the main impact of PFS is expected to be in QoL, which may or may not represent a causal relationship.

As noted by Fallowfield and Fleissig,[36] "New treatments that increase PFS may not be of sufficient value to patients with advanced-stage cancer unless accompanied by tangible quantity or QoL advantages. Any symptom relief that patients gain from treatment resulting in tumor shrinkage or stabilization must be balanced against the toxic effects that drug therapy itself creates."[36]

Recent studies have used conjoint analysis to try to understand and quantify the value of PFS to patients versus avoidance of risk of toxicities. Mohamed et al.,[37] evaluated the benefit-risk preferences of patients with renal cell carcinoma, using a series of 12 trade-off questions to determine what magnitude of PFS improvement was worth significant treatment-related risks. Patients were willing to accept significant treatment-related risks of two to three percent for liver failure and blood clot to increase PFS by 11 months.

Bridges et al.,[38] in a conjoint analysis of patients with advanced non-small cell lung cancer, examined the trade-offs patients were willing to make between increased PFS and the risk of experiencing disease symptoms, such as fatigue, diarrhea, nausea and vomiting, fever, infection and rash. They concluded the value patients attribute to improvements in PFS was conditional upon the severity of disease symptoms experienced.

These studies suggest that QoL, in relation to PFS, is important to patients. However, because PFS itself is an outcome, it is not possible to study it in the same manner applied to a drug treatment. Normally, randomization of patients to either intervention or control arms in a clinical trial ensures the equal distribution of confounding variables. However, because PFS is an outcome, it cannot be predicted in advance, and patients cannot be randomized according to its improvement or lack thereof. Thus, an important aspect to the investigation of PFS, and its impact on other outcomes of importance to patients, is an examination of how the relationship is studied and whether it is feasible to define the relationship exclusive of other factors clearly. . The purpose of this project is to address whether PFS is an outcome related to psychological well-being or QoL.

Methods

The Methods section describes the purpose and composition of the Technical Expert Panel (TEP), the Key Question (KQ) development and rationale, eligibility criteria, the search strategies used for published and gray literature, and the processes for article screening and selection, data abstraction, quality assessment, and data synthesis.

Technical Expert Panel

The TEP, a group of six individuals with particular expertise in a variety of areas directly pertinent to this methods paper, was assembled to provide input regarding the scope and execution of the project. This group included representatives of the Food and Drug Administration (FDA) and the Centers for Medicare and Medicaid, an academic oncologist, a professional advocate for cancer treatment and research (representing the American Cancer Society), a patient advocate, a social scientist, and a Task Order Officer from the Agency for Healthcare Research and Quality (AHRQ). The Evidence-Based Practice Center (EPC) held teleconferences with the TEP, which provided input on the scope and Key Questions of this project and helped to define parameters for the methodology of the critical appraisal performed.

Key Questions

After discussion with the TEP, two Key Questions were developed in an effort to evaluate evidence demonstrating progression-free survival (PFS) to be an outcome of importance to patients, or showing an association between PFS and other outcomes of importance to patients.

KQ1, PFS as an Outcome of Importance to Patients

When PFS is used as a primary clinical endpoint in treating patients with advanced cancers, is there direct evidence that knowing PFS impacts patient anxiety, depression, or psychological well-being?

- If yes, does the manner of communicating PFS affect patient anxiety, depression or psychological well-being?

KQ1 is focused on the direct causal relationship between knowledge of PFS status and the psychological impact on patients. This is an important question because PFS, when measured using imaging techniques, may occur as an asymptomatic event and may or may not be associated with changes in overall survival (OS). In these cases, PFS can be a stand-alone finding, and its greatest impact may be in promoting patient psychological well-being, depending on the information provided and a patient's understanding of that information.

KQ2, Associations of PFS with Quality of Life Outcomes

For agents where PFS is the primary outcome measure being used to establish the performance (efficacy and safety) of a new drug, what evidence exists regarding the association of PFS with QoL, and related outcomes such as disease symptoms?

KQ2 is focused on the association between PFS and QoL-related outcomes of importance to patients that might be observed as part of the results of a drug treatment trial. To answer KQ2 we identified a targeted sample of drugs. First, we chose the category of drugs used for treating

patients with locally advanced or metastatic solid epithelial tumors (breast, ovary, colon, kidney, lung and head/neck). These high profile tumors represent an attractive study set because they exhibit variable speeds of progression, variable effect sizes, and differences in manifestations of symptoms. We identified drugs approved between 2005 and 2010, by one or more regulatory agencies, on the basis of trials using PFS as the primary endpoint for treatment of solid epithelial tumors. This timeframe was selected because it represented a unique period in time in which a relatively large number of drugs for treatment of advanced epithelial tumors were approved for use in the US, Europe and/or Canada based on studies using PFS as an endpoint. These drugs included Avastin, Ixempra, Tykerb, Vectibix, Doxil, Gemzar, Yondelis, Nexavar, Votrient, Sutent, Tarceva, Taxotere (Appendix B). The applicability of findings regarding PFS and its association with outcomes of importance to patients may be restricted to this select group of drugs and the patients who received them. However, because these represent all drugs within this unique category, bias was expected to be minimal. Based on the importance of the decision-making process in patients with advanced cancers, an understanding of the data would clearly move the field forward or identify research gaps in need of attention.

Eligibility Criteria

Study inclusion/exclusion criteria for KQ1 and KQ2 are outlined in Table 1 and Table 2, respectively. This methods project focuses on the causal relationship (KQ1) or association (KQ2) between PFS and other outcomes of importance to patients.

In contrast to a traditional comparison of interventions, the primary variable of interest, PFS, is an outcome. As such it cannot be pre-selected or randomized as part of a prospective study. Thus, a new parameter appears in the inclusion/exclusion criteria, that of "comparator outcomes," which is a modification of the traditional Population, Intervention, Comparator, Timing, and Setting (PICOTS) framework. The comparator outcomes of interest are those of importance to patients, such as QoL and psychological well-being.

Table 1. Progression-Free Survival, Inclusion/Exclusion Criteria for KQ1

Category	Inclusion/Exclusion Criteria
Population	Inclusion: Patients with cancer
Interventions	Inclusion: Chemotherapy
Primary outcomes	Inclusion: PFS Exclusion: Studies without PFS as an endpoint
Comparator outcomes	Inclusion: Outcomes of importance to patients, including psychological well-being (or anxiety, depression) and quality of life
Time period	From 1999 to 2012
Setting	Oncology care settings for patients with neoplasms
Publication language	English or English language translations when available
Study designs	Clinical studies
Follow-up duration	All
Sample size	Studies of any size.

PFS, progression-free survival

Table 2. Progression-Free Survival, Inclusion/Exclusion Criteria for KQ2

Category	Inclusion/Exclusion Criteria
Population	Inclusion: Patients with advanced solid epithelial cancers studied as part of a randomized clinical trial for the following targeted drugs: Avastin, Ixempra, Tykerb, Vectibix, Doxil, Gemzar, Yondelis, Nexavar, Votrient, Sutent, Tarceva, Taxotere) (see Appendix B)
Interventions	Inclusion: Chemotherapy using a drug approved for use by a U.S. Canadian, or European regulatory body based on PFS outcomes
Primary outcomes	Inclusion: PFS Exclusion: Studies not including PFS Note: Studies in which a drug also improves OS will not be excluded
Comparator outcomes	Inclusion: Outcomes of importance to patients, including psychological well-being (or anxiety, depression) and quality of life
Time period	2004 to 2012
Setting	Oncology care settings for patients with advanced solid epithelial tumors
Publication language	English or English language translations when available
Study designs	Prospective randomized clinical trials
Follow-up duration	All
Sample size	Studies of any size

PFS, progression-free survival; OS, overall survival

Literature Search

Published Literature

Literature searches were conducted in the following databases:

- MEDLINE® (via PubMed)
- EMBASE®
- Cochrane Central Register of Controlled Trials (CENTRAL)

The research librarian, in collaboration with the project team, developed and implemented search strategies designed to identify evidence relevant to each Key Question. Appendix C provides the actual terms employed in the search strategy.

To address KQ1, a comprehensive search was developed that would find any article looking for a causal relationship between PFS and outcomes of interest to patients, in particular those that affected psychological states (see KQ1, PFS as an Outcome of Importance to Patients). Databases were queried using search terms such as PFS and psychological, anxiety, and depression, in combination with terms for cancer. Accepted study designs included randomized controlled trials, observational studies, case studies, or other clinical studies. The timeframe (1999 to 2012) was selected because it represents a transition period in drug approvals, in which the FDA moved away from use of response rate or time to progression toward an increased use of OS or PFS.[2, 39] Because input from the TEP suggested there would be few, if any, publications addressing KQ1, the search criteria were deliberately broad and include all tumors and most types of published studies for use as evidence, including randomized prospective clinical trials, observational studies, and case series. Editorials, commentaries, and reviews were not included as evidence, but may have contributed additional references or to the background information.

For KQ2, a targeted search strategy included the identified drugs of interest -- drugs approved for treatment of advanced solid epithelial tumors based on PFS as a primary endpoint (see KQ2, Associations of PFS with Outcomes of Importance to Patients) and QoL. The search

encompassed the years 2004 to 2012 and as noted above was aimed at identifying studies that included drugs approved between 2005 and 2010 (Appendix B), a period of time in which a number of drugs were approved using PFS as the primary endpoint in the newly formed Office of Oncology Drug Products at the FDA and/or by European and Canadian regulators.

The literature searches for KQs 1 and 2 were restricted to the study of humans in the English language, with the exception of published articles in other languages for which English translations were available. Previous studies have shown that excluding non-English-language studies has little impact on conclusions relative to the resources required for translation.[40, 41]

Gray Literature

The gray literature was searched for relevant studies that met the inclusion/exclusion criteria. For the purposes of this methods project, the gray literature comprises information that is not controlled by commercial publishing, including abstracts presented at major oncology meetings (e.g. the National Comprehensive Cancer Network Outcomes Database Abstracts, the Annual Congress or the American Society of Clinical Oncology Annual Meeting); U.S. regulatory documents (e.g. guidance documents on cancer endpoints and patient reported outcomes, public summaries of FDA meetings on cancer endpoints or FDA new drug approval (NDA) briefings packages), meeting transcripts and materials, appeals, and clinical reviews; clinical trial registries (ClinicalTrials.gov), and documents developed and posted by European and Canadian drug regulating agencies. Because unpublished literature was not used as evidence in this project, the gray literature was used to identify potential publications in progress and for background information.

Citation Screening

Articles obtained from the searches were uploaded to an EndNote® (Thomson Reuters Corporation, New York, United States) reference manager database and then transferred to DistillerSR® (Evidence Partners Inc., Ottawa, Canada) software for study eligibility screening. Using six preset questions that reflected the inclusion criteria, two reviewers independently screened titles and abstracts, marking each of the questions as: 1) yes (eligible for full-text screening); 2) no (ineligible for inclusion); or 3) uncertain (include in full-text screening to resolve eligibility). DistillerSR® provided a report on reviewer discrepancies, which were resolved by discussion and consensus opinion; a third reviewer was consulted as needed.

Using a second set of three selection questions, two reviewers independently screened full-text articles using the same approach as above to determine eligibility for data abstraction. The title/abstract and full-text screening questions are presented in Appendix D. Although the title/abstract screening and full-text screening questions used “PRO or QoL” as the pivotal selection terms, for KQ1, reviewers included articles with reference to changes in any psychological state, such as anxiety, mood, or depression, as outcomes of importance to compare to PFS. The logic for considering QoL and, in particular, psychological outcomes is described above (see Background, PFS as a Health Outcome). For KQ2, reviewers included articles with any reference to changes in the above psychological states, plus QoL, disease symptoms, emotional or social, or any similar type of outcome of importance to patients.

Data Extraction

Articles were distributed among the team members for abstraction. A Microsoft Excel® spreadsheet was created and used for abstracting information from the eligible studies. Each article was abstracted by a single reviewer, with 100 percent fact-checking being completed by a second reviewer. Discrepancies were resolved by discussion and, if necessary, a third opinion was sought. The abstraction form items are presented in Appendix E.

Data Synthesis

The analysis of study data is qualitative only. No effort was made to pool results, and no quantitative assessment was performed. The unique features of the included studies were captured in separate tables describing study characteristics, treatment safety and efficacy, description of the QoL measurement, the extent of missing data and how the issue was addressed, the statistical analysis, and the description of the association between PFS and QoL.

A hypothetical causal model, represented as a directed acyclic graph, was posited based on the associations expressed as plausible or likely by the TEP and also informed by the literature reviewed. This model is based on techniques described in Supplement 2 of the recently published draft AHRQ report.[42]

Quality Assessment of Individual Studies

To assess the quality of included studies, existing quality assessment tools were reviewed for their applicability to the association being evaluated. In this methods project, the relationship of interest was not a comparison of treatment effects, but instead, compared two outcomes: PFS (considered the primary outcome) and other outcomes of interest to patients (considered the comparator outcomes). Owing to the nature of this focus, none of the existing quality assessment tools was judged entirely appropriate to capture the potential biases associated with this type of comparison. Accordingly, a more general approach developed and applied by Turner et al.,[43] was adopted for evaluation of internal bias (i.e., study quality). This approach entails "envisioning" an "idealized version" of a study that would be free of bias, and then compares the included study to the idealized one to identify potential sources of bias. This approach allows considering any unique aspects of questions and studies, and includes well-established sources of potential bias, including selection, performance, attrition, detection, and other sources.[44-46] Taking this approach was further justified based on its applicability to observational studies.[43] Using this approach, and with consensus among the study investigators, four additional items that specifically addressed the potential for bias in studies evaluating an association between PFS and QoL were added to the quality assessment criteria (see Appendix F).

For the quality assessment, the general qualitative principles described in the AHRQ methods guide, *Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions* were applied.[42] Narrative descriptions were developed to correspond to the quality assessment of excellent, fair, and poor quality studies (Table 3). The most critical sources of potential bias were deemed to be failure to blind, significant patient drop out, and failure to provide follow-through observations after progression of disease.[47-49]

Table 3. Progression-Free Survival, Definitions of Overall Study Quality Ratings

Study Quality	Quality Attributes
Excellent	Adequate blinding, minimal patient attrition (20% or less), no censoring at progression, and conformance with 12 or more additional positive quality factors (Appendix F)
Fair	Adequate blinding, minimal patient attrition (20% or less), no censoring at progression, and conformance with 6 or more additional positive quality factors (Appendix F)
Poor	No blinding and/or more than minimal patient attrition (more than 20%) and censoring at progression (Appendix F)

Strength of Evidence

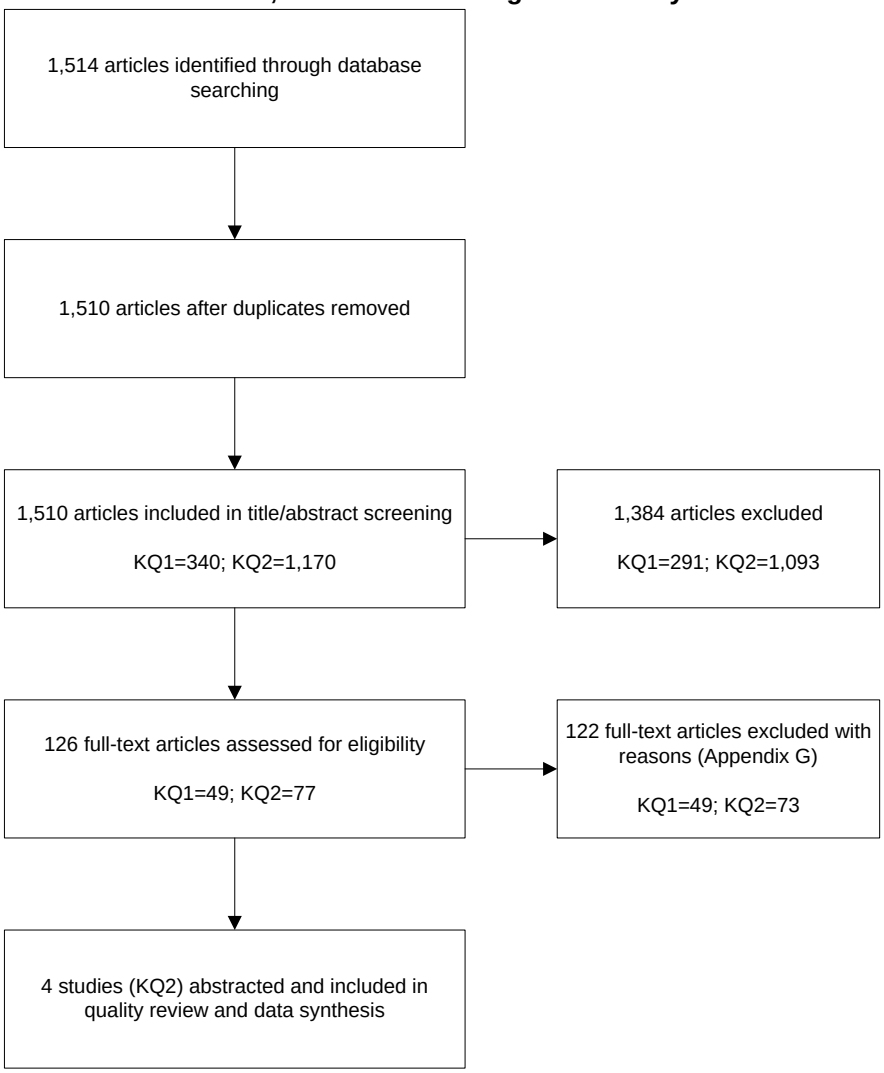
The system published by Owens et al.[50] for grading the strength of evidence was used to systematically describe the evidence in this methods project. The system considers four standard domains: risk of bias, consistency, directness, and precision. These allow for four possible grades: high, moderate, low, or insufficient.

Results

Searches of Published Literature

The searches for published literature conducted in Medline, PubMed, EMBASE, and Cochrane Database of Systematic Reviews identified 1,510 citations. Figure 1 presents a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of the study selection process. During title and abstract screening, 1,384 articles were excluded, and the remaining 126 articles were retrieved and evaluated at the full-text screening level. Of these, no articles were identified which addressed Key Question (KQ) 1, and only four articles addressed KQ2. The four articles were eligible for abstraction, because they included at least one of the targeted disease/drug combinations, assessed QoL, and reported an association between progression-free survival (PFS) and QoL. Appendix G provides a list, by reason, of the 122 articles excluded during the full-text screening. After abstraction, all four studies remained eligible for quality review and data synthesis.

Figure 1. Progression-Free Survival, PRISMA Flow Diagram of Study Selection



KQ: Key Question

Searches of Gray Literature

The gray literature search yielded the following results:

- Regulatory Information: The search yielded a Food and Drug Administration (FDA) guidance document[51] that provided background information on cancer endpoints, including detailed information on the use of PFS, but did not specifically elucidate relationships between PFS and other endpoints of interest to patients with cancer. Similar guidelines were identified in publications by the European Medicines Agency (EMEA)[52] and Health Canada.[53] In addition, four FDA new drug application briefing documents, two meeting transcripts, one approval package, one clinical review, an appeals document with review response, and 13 public presentations at workshops on endpoints were reviewed. Finally, nine European regulatory scientific summaries and seven Health Canada drug Summary Basis of Decisions were reviewed. None of these specifically addressed the relationship between PFS and other outcomes of importance to patients; therefore, these documents provided only background information.
- Clinical trial registries: ClinicalTrials.gov was searched using combined search terms: "progression-free survival" AND "primary endpoint" AND "quality of life" and by combined search terms: "progression-free survival" AND "cancer" AND either "anxiety," "depression" or "psychological." Thirty-seven trials were identified, but review of the study descriptions revealed none that were relevant.
- The National Comprehensive Cancer Network (NCCN) oncology outcomes database and the ASCO (American Society of Clinical Oncology) annual meeting abstract database (past 3 years) were searched using the same terms as in the literature searches. Only one[54] of 35 identified ASCO abstracts was relevant (an update of Zhou[55]), but no full-text article has been published. No relevant NCCN abstracts were identified.

Key Questions

Key Question 1

Based on the search criteria established for KQ1, we identified no studies that directly addressed the question. No information was obtained on patients' psychological response to knowing their PFS status, or whether the manner of communicating PFS results affects patient anxiety or psychological well-being.

Key Question 2

Four studies[55-58] were identified that statistically evaluated the association between PFS and a QoL-related measure, with each study a *post hoc* analysis following a published primary study of drug efficacy and safety (see Study Design, Table 4). The study characteristics are presented in Table 4. Sample sizes ranged from 219 to 463. Overall, 1,515 patients were enrolled in these studies. The types of cancers and (treatment drug) included breast (Tykerb/lapatinib),[55, 57] colorectal (Vectibix/panitumumab),[58] and renal cell (Votrient/pazopanib).[56]

The four drug studies all reported statistically significant improvement in median PFS for the treatment arm compared with the control arm (Table 5). Hazard ratios for PFS ranged from 0.46 to 0.71. In response to therapy, PFS showed increases in median duration from as low as 0.7 weeks (8 vs. 7.3 weeks)[58] to 5 months (9.2 vs. 4.2 months).[57] In all but one study,[55] incidence of Grade 3 and 4 adverse events was higher for patients in the experimental treatment arm than in

controls.

The four studies used validated assessment tools to collect QoL information (Table 6). They included at least one or more multidimensional tools, such as the European Organization for Research and Treatment of Cancer – Quality of Life Questionnaire C-30 (EORTC-QLQ-C30), the functional assessment of cancer therapy: general (FACT-G), or the EuroQoL 5 dimensions (EQ-5D) index. All four studies established a minimally important difference threshold before analysis, but none provided detailed information on administration of the QoL measure (e.g., whether the QoL measure was administered in a standardized, reproducible manner each time, the conditions of administration, whether patients were informed of their PFS status prior to administration, etc.). QoL testing intervals ranged from 2- to 12-week intervals.

A problem common to all four studies was significant patient dropout (Table 7). In two studies,[57, 58] the minimum requirement for inclusion in the analyses was QoL assessment at baseline and at least one follow-up. Even using this relatively liberal inclusion threshold, overall dropout rates ranged from 66 percent to 73 percent in experimental arms and from 77 percent to 96 percent in control arms. Disease progression was the major cause of dropout. All four studies addressed data missing not at random in the statistical plan.[55-58] The most common technique used by these studies for addressing data missing not at random was last observation carried forward (LOCF).[55, 57, 58]

The most important finding relating to this methods project was the relationship between PFS and other outcomes of importance to patients found in all four studies (Table 8). More specifically, being progression-free had a statistically significant positive association with patients' improved QoL and/or decreased disease symptoms. Of the four studies, three[55, 57, 58] reported an improvement in QoL or disease symptoms among patients who remained progression-free compared with those who had progressed. The remaining study[56] showed a similar improvement in patients exhibiting complete or partial responses to treatment. In this last study, the data was available to compare patients in PFS versus those with disease progression, but the analysis was not performed. Statistical methods to analyze the association between PFS and the QoL ranged from correlational techniques to more robust analysis of covariance (ANCOVA) and regression analyses. All four of the studies adjusted for QoL baseline scores in their analyses.

In the four included studies there was an improvement in QoL or disease symptoms among patients who remained progression-free; however, the ability to draw conclusions from this important finding was compromised by the significant problems observed in the quality of these studies. As shown in Table 9, each of the four included studies met approximately 50 percent of identified quality items. However, all four failed to address the four items specifically developed to address the potential for bias in studies evaluating an association between PFS and QoL (items in italics). In addition, in all four studies there was substantial data missing not at random, and in two studies,[55, 58] investigators and patients were not blinded to treatment. Based on the assessment of risk of bias and the overall study quality items (Table 3), all four studies received a poor quality rating.

The strength of evidence assessment goes beyond looking at study design alone, and takes into account other facets of the evidence, including the presence or absence of bias, directness, consistency and precision.[50] This allows clinicians, policy-makers, and patients to make well-informed decisions based on a more comprehensive evaluation of the evidence. Based on the poor quality of the four included articles, risk of bias was considered high. All four studies produced evidence of a positive association between PFS and outcomes of importance to

patients, so although the body of evidence was small, it exhibited high consistency. However, this observation was tempered by the fact that the studies were of different designs, used different measurement tools, and assessed different outcomes. Consequently, it was not possible to measure the range of effect sizes. All associations were indirect. A quantitative analysis of pooled performance could not be performed, so no estimate of precision was possible. Based on the small body of literature identified and the evaluation described above, the strength of evidence is insufficient to demonstrate an association between PFS and QoL. Moreover, this association is only supportive of, but does not directly address Key Question 1, whether knowing PFS impacts patient anxiety, depression, or psychological well-being, for which no direct evidence was found at all. Thus, the overall body of evidence is judged insufficient.

Table 4. Progression-Free Survival, Study Characteristics

Author, Year	Cancer type	Cancer Stage	Experimental Drug vs. Control	Study Design	1° / 2° Outcomes	Total N (n Exp/Control)	Age, Mean (SD); Median (range)
Cella, 2012[56]	Renal Cell Cancer	advanced, Stage IV, untreated or cytokine pretreated	pazopanib (Votrient) vs. placebo	post-hoc HRQL analysis of Sternberg 2010, phase III, randomized, double-blind, placebo-controlled, multicenter	PFS / OS, ORR, QoL, safety	434 (289/145) ITT	pazopanib 59 (28-85); placebo 60 (25-81)
Sherrill, 2010[57]	Breast	metastatic, receptor-pos HER-2+, Stage IIIB/IIIC or IV, untreated	letrozole + lapatinib (Tykerb) vs. letrozole + placebo	post-hoc subset analysis (HER-2+ pts) of Johnston 2009, phase III, randomized, double-blind, placebo-controlled, parallel-group, multicenter	PFS / OS, ORR, QoL, clinical benefit rate	HER-2+ subset: 219 (111/108)	NR
Siena, 2007[58]	Colorectal Cancer	metastatic, progressed on prior Tx	panitumumab (Vectibix) + BSC vs. BSC alone	PFS-QoL association analysis of Van Cutsem 2007, phase III, randomized-controlled, open-label, multicenter	PFS / OS, ORR, QoL, safety	463 (231/232) ITT	pan + BSC 62 (27-82); BSC alone 63 (27-83)
Zhou, 2009[55]	Breast	advanced, metastatic, HER2+, progressed on prior treatment	lapatinib (Tykerb) + capecitabine vs. capecitabine alone	HRQL analysis of Cameron 2008, phase III, randomized, open-label, multicenter	TTP / PFS, OS, ORR, QoL, clinical benefit rate, safety	399 (198/201) ITT	lap+ cap 54 (26-80); cap alone 51 (28-83)

BSC, best supportive care; HER2, human epidermal growth factor receptor type 2; HRQL, health-related quality of life; ITT, intention-to-treat; NSCLC, non-small-cell lung cancer; ORR, object response rate; OS, overall survival; PFS, progression-free survival; QoL, quality of life; TTP, time to progression

Table 5. Progression-Free Survival, Treatment Efficacy and Safety

Author, Year	Experimental Drug vs. Control	Median PFS (Experimental vs. Control)	PFS, Experimental vs. Control HR (95% CI)	Median OS (Experimental vs. Control)	OS, Experimental vs. Control HR (95% CI)	ORR, Experimental vs. Control % (95%CI)	Any Grade 3/4 Adverse Events in Experimental Group N (%)	Any Grade 3/4 Adverse Events in Control Group N (%)
Cella, 2012[56]	pazopanib vs. placebo	9.2 vs. 4.2 mos	0.46 (0.34-0.62), p< 0.0001	NR	NR (data not mature)	30 vs. 3, p< 0.001	40%	20%
Sherrill, 2010[57]	letrozole + lapatinib vs. letrozole + placebo	8.2 vs. 3 mos	0.71 (0.53-0.96), p= 0.019	33.3 vs. 32.3 mos	0.74 (0.5-1.1), p= 0.113	28 vs. 15, p= 0.021	diarrhea 10%; rash 1%	diarrhea 1%; rash 0%
Siena, 2007[58]	panitumumab + BSC vs. BSC alone	8 wks vs. 7.3 (mean 13.8 wks vs. 8.5)	0.54 (0.44-0.66), p<0.0001	NR	1.0 (82-1.22), p= 0.81	10 vs. 0 (at 1 yr)	of 229, 79 (35)	of 234, 45 (20)
Zhou, 2009[55]	lapatinib + capecitabine vs. capecitabine alone	NR	0.55 (0.4-0.74), p< 0.001	15.6 vs. 15.3 mos	0.78 (0.55-1.12), p= 0.177	24 vs. 14, p= 0.017	35%	33%

BSC, best supportive care; CI, confidence interval; HR, hazard ratio; Mo(s), months; NR, not reported; ORR, objective response rate; P, probability; PFS, progression-free survival; OS, overall survival; Wks, weeks

Table 6. Progression-Free Survival, QoL Measurement Description

Author, Year	QoL Measure	Significant Change Threshold	QoL Measure Administration	QoL Testing Intervals & Discontinuation Criteria	Timing of Drug Treatment	Variables Measured or Compared (Experimental vs. Control)
Cella, 2012[56]	EORTC QLQ-C30, EQ-5D Index, EQ-5D VAS	QLQ-C30 5 to 10 pts; EQ-5D 0.08 pts; EQ-5D VAS 7 pts	blinded patient self-report	baseline, 6, 12, 18, 24, 48 weeks	daily until disease progression, unacceptable toxicity, death, or withdrawn consent	baseline scores; time to HRQL deterioration; correlation between response and HRQL
Sherrill, 2010[57]	FACT-B total score, FACT-G, TOI	FACT-B 7-8 pts; BCS 2-3 pts; FACT-G / TOI 5-6 pts	self-administered	Day 1 (baseline), every 12 weeks, at study withdrawal	daily until disease progression or withdrawal from study due to toxicity or other reasons	baseline scores; change scores; correlation between QoL and tumor response
Siena, 2007[58]	FACT-FCSI, EQ-5D Index, EQ-5D VAS, QLQ-C30 GHS	FCSI - 4-pt; EQ-5D - 0.08-pt.; EQ-5D VAS - 5.48-pt; GHS - 7.07-pt.	NR	baseline, every 2 weeks (FCSI) or monthly (all others) during treatment, at 30-day safety follow-up visit	2-week cycles	association between mean changes in HRQL/disease symptoms and PFS
Zhou, 2009[55]	FACT-B total score, FACT-G, TOI	FACT-B 7-8 pts; BCS 2-3 pts; FACT-G / TOI 5-6 pts	self-administered	at screening (baseline) visit, every 6 wks for 24 wks, every 12 wks, study discontinuation, until disease progression or withdrawal due to toxicity or other reasons	21-day cycles until PD or withdrawal because of toxicity or other reasons	Baseline scores; change scores; QoL changes from baseline based on tumor response

BCM20, Brain Cancer Module20; BSC, best supportive care; EORTC-QLQ, European Organization for Research and Treatment of Cancer Quality of Life Questionnaire; FACT-B, Functional Assessment of Cancer Therapy-Breast; FACT-G, FACT-general; FCSI, FACT Colorectal Symptom Index; GHS, Global Health Status; HRQL, health-related quality of life; PD, progressive disease; PF, physical functioning; PFS, progression-free survival; QoL, quality of life; TOI, Trial Outcome Index; VAS, Visual Analog Scale

Table 7. Progression-Free Survival, QoL Measurement – Missing Data

Author, Year	QoL Analysis-Requirement for Inclusion	Baseline Completion Rate N (%) (Exp/Ctrl)	Experimental Group Dropout N (%)	Control Group Dropout N (%)	Primary Reasons for Dropout	Statistical Method for Missing Data
Cella, 2012[56]	all available HRQL assessments during periods of best response for each pt were used	QLQ-C30: 288 (99.7)/142 (97.9) EQ-5D: 287 (99.3)/143 (98.6) EQ-5D VAS: 283 (97.9)/141 (97)	wk 48, QLQ-C30 194 (67), EQ-5D 192 (67), EQ-5D VAS 188 (66)	wk 48, QLQ-C30 121 (85), EQ-5D 121 (85), EQ-5D VAS 118 (84)	disease progression	sensitivity analyses of HRQL deterioration using a composite end-point, where PD and HRQL deterioration were considered as an event
Sherrill, 2010[57]	baseline & at least one follow-up QoL assessment	110 (100)/101 (98)	wk 48, 79 (72)	wk 48, 78 (77)	disease progression or withdrawal from Tx	LOCF - missing scores imputed from the last nonmissing score at a previous visit
Siena, 2007[58]	baseline & at least one follow-up QoL assessment	207 (90)/184 (79)	wk 16, 145 (70)	wk 16, 177 (96)	disease progression	two methods, last value carried forward and slope method, used to impute missing values
Zhou, 2009[55]	all data collected up to close of study enrollment (no data after crossover or alt tx)	171 (86)/168 (84)	wk 24, 124 (73)	wk 24, 138 (82)	disease progression or withdrawal for other reasons	LOCF; exploratory random pattern effects model to deal with missing data showed QoL comparable bet Tx arms (data not shown)

BSC, best supportive care; HRQL, health-related quality of life; LOCF, last observation carried forward; NR, not reported; PD, progressive disease; PF, physical functioning; QoL, quality of life; QLQ, Quality of Life Questionnaire; VAS, Visual Analog Scale; Tx, treatment; Wk, week

Table 8. Progression-Free Survival, PFS/QoL Association

Author, Year	Cancer type	Exp Drug vs. Control	PFS-QoL Association	PFS-QoL Statistical Method
Cella, 2012[56]	Renal Cell Carcinoma	pazopanib vs. placebo	progression-free (CR/PR) patients experienced significantly less HRQL deterioration than patients with SD and PD for GHS, EQ-5D utility index, EQ-5D VAS	univariate and multivariate regression models, adjusted for baseline HRQL scores, were developed to determine the association between changes in HRQL and tumor response
Sherrill, 2010[57]	Breast	letrozole + lapatinib vs. letrozole + placebo	at wks 24 and 36, average FACT-B change scores from baseline were statistically and clinically significantly improved for patients with no PD versus those with PD	least squared means from an ANCOVA, adjusted for baseline value, was used to compared QoL score changes from baseline between progressors and nonprogressors
Siena, 2007[58]	Colorectal cancer	panitumumab + BSC vs. BSC alone	at week 8, being progression-free was associated with significantly and clinically meaningful lower CRC symptomatology for both treatment arms and higher HRQL for panitumumab patients only	t-tests and least-squares estimates were calculated for differences in QoL measures at weeks 4, 8, 12, 16 controlling for baseline score by progression status as of wk 8 (PD vs. no PD) within each treatment arm
Zhou, 2009[55]	Breast	lapatinib + capecitabine vs. capecitabine alone	at 12 wks, patients with SD had significantly positive changes from baseline, whereas patients with PD had negative changes in FACT-B, FACT-G, TOI, EQ-5D utility index, EQ-5D VAS scores	ANCOVA examined the relationship between tumor response status (SD vs. PD) adjusted for baseline scores (treatment arms pooled for this analysis)

ANCOVA, analysis of covariance; BSC, best supportive care; CR, complete response; FACT-B, Functional Assessment of Cancer Therapy-Breast; FACT-G, Functional Assessment of Cancer Therapy-General; GHS, Global Health Status; HRQL, health-related quality of life; PD, progressive disease; PF, physical functioning; PR, partial response; QoL, quality of life; SD, stable disease; SF, social functioning; VAS, Visual Analog Scale

Table 9. Progression-Free Survival, Summary of Individual Article Quality Assessments

	Quality Criteria	Cella, 2012[56]	Sherrill, 2010[57]	Siena, 2007[58]	Zhou, 2009[55]	# of Studies Meeting Criterion
Selection Bias	Subjects in different groups recruited from same population	Yes	Yes	Yes	Yes	4
	Subjects in different groups recruited over same time period	Yes	Yes	Yes	Yes	4
	Inclusion/exclusion criteria clearly stated	Yes	Yes	Yes	Yes	4
	Randomization used	Yes	Yes	Yes	Yes	4
	Comparable baseline characteristics between groups	Yes	Yes	Yes	Yes	4
	Apart from treatment under investigation, groups treated equally	Yes	Yes	Yes	Yes	4
	*Similarity between groups in potentially important confounders (e.g., those with and without PFS) or adequate adjustment for confounding**	No	No	No	No	0
	Patients not censored at progression	No	No	No	No	0
Performance Bias	PFS assessed in valid and reliable manner (central radiology review, RECIST criteria used, inter- and intra-observer variability reported)	Yes	Yes	Yes	Yes	4
	Subjects and investigators blind to treatment group	Yes	Yes	No	No	2
	Subjects awareness of PFS status is described	No	No	No	No	0
Attrition Bias	Missing data addressed by description and analysis	Yes	Yes	Yes	Yes	4
	Results unlikely to be affected by losses to follow-up	No	No	No	No	0
	Missing data likely to be at random	No	No	No	No	0
Detection Bias	Outcome (progression) assessors blind to treatment group	Yes	No	Yes	Yes	3
	PFS and other endpoints defined	Yes	Yes	Yes	Yes	4
	QoL administration process was described	No	No	No	No	0
	A priori hypothesis regarding the relationship between PFS and QoL	No	No	No	No	0
	# Yes (of 18 items)	**12**	**11**	**11**	**11**	
	Quality Assessment Rating (definitions in Table 3)	**Poor**	**Poor**	**Poor**	**Poor**	

PFS, progression-free survival; QoL, quality of life; RECIST, Response Evaluation Criteria in Solid Tumors
*Italicized criteria are those that are uniquely important to trials comparing PFS and QoL (see Appendix F)

Discussion

The objective of this methods project is to address whether PFS is an outcome related to psychological well-being or QoL. As described in the Background section of this document, the use of PFS can be challenging, especially in terms of potential bias (assessment, evaluation, performance, detection, and attrition) and its utility as a surrogate for other outcomes. There is a growing literature describing the best practices to address sources of bias, including careful prospective definition of progression,[51] blinding of patients and physicians to treatment,[3, 28] use of blinded external radiological review,[3, 23, 28, 29] and attention to patient attrition and data censoring.[3, 24, 27] However, accurate measurement of PFS can still be problematic. In addition, the endpoint itself is dependent on the context of use, including the specific disease, disease state, the types of drugs and their toxicities, and patient response to the risks and benefits of treatment.[5] Regardless of these issues, PFS has been an attractive choice as an endpoint because, compared with OS, studies may be conducted using fewer subjects, at lower costs, and with a shorter timeline.

This methods project focused specifically on the causal or associational relationship between an *outcome*, PFS, and other *outcomes* of importance to patients. In contrast to a traditional comparison of *interventions*, the variable of interest, PFS, is somewhat problematic, because the presence or absence of progression is observed and not manipulated. This has implications for study design, interpreting results, quality evaluation, and the potential for a study informing the questions posed here. The questions must be addressed from the vantage of observational data and their inherent potential for bias.

We first performed a broad-based review of the literature to identify studies examining the direct impact of PFS status on patient anxiety, depression or psychological well-being. The search revealed no studies that addressed this question. This is a notable gap in the evidence, because one can speculate that knowing they are progression-free might make patients feel relief,[22] while news about disease progression might make them feel worried, but this has not been systematically studied. There is literature on the delivery of bad news to patients, including reports specifically addressing the direct psychological impact of telling patients they have been diagnosed with cancer.[59-61] How applicable this is to the setting of PFS remains unexamined. The psychological response to information on PFS may depend on the way it is delivered by a physician and on the patient's understanding of what PFS actually means. It is not known whether patients would view PFS differently if news of measurable progression or lack thereof was delivered with an explanation that progression has no demonstrated significant relationship with OS.

We also conducted a focused review using drugs recently approved by one or more regulatory authorities using PFS as a primary study endpoint in order to explore what evidence exists regarding the statistical association of PFS with QoL, disease symptoms, or other QoL related outcomes. Four studies[55-58] were identified. These four studies reported an improvement in QoL or disease symptoms among patients who remained progression-free compared with those who had progressed.

Evidence for a causal relationship in these studies remains unclear. It could be that a common underlying mechanism improves both PFS and QoL. For example: tumor shrinkage leads to improved pain which is detected via a QoL questionnaire and also is seen as disease regression on imaging. Alternatively, it could be that patient knowledge of progression status itself affects subjective impressions of overall QoL (or affects a specific symptom that impacts overall QoL measure scores, such as anxiety).The issue of causality is important in terms of considering

future use of the PFS endpoint in clinical research. If the underlying mechanism affecting PFS and QoL are the same, then the clinically meaningful endpoint is actually QoL, with PFS as a supportive outcome. In the methods section of this report, several sources of potential bias associated with PFS were identified and discussed, and unfortunately, all four of the included studies[55-58] had significant sources of bias, as well as other quality issues, which weakens confidence in the reported PFS/QoL association. All were limited by a failure to develop an *a priori* hypothesis regarding the association of PFS and QoL, by censoring at the time of progression, and by substantial rates of missing data not at random. While all four studies did describe analytical approaches to address missing data not at random, the ability to account for the extreme magnitude of dropout observed is doubtful. In addition, two[55, 58] of the four studies failed to provide patient-investigator blinding to treatment.

All four of the studies adjusted for QoL baseline scores in their analyses. This is particularly important because patients with better baseline scores may be more likely to have longer PFS. Although the use of QoL as a prognostic factor for PFS is not the focus of this methods project, there is evidence that QoL itself is a prognostic indicator of survival.[62-65] The basic mechanism involved appears unrelated to the association of PFS with QoL; however, if QoL is prognostic for survival, there is the potential to confound studies evaluating the impact of PFS on QoL.

Improvements in survival times for both PFS and OS were relatively small in these studies (i.e., weeks to months). If treatments are making only small improvements in survival outcomes, it becomes that much more important to develop reliable estimates of their impact on disease symptoms and QoL, so that patients and physicians can make informed choices.

It is informative to consider an explicit framework or understanding of how treatment, disease progression (and knowledge of it), and other outcomes are related. Such a framework is hypothetical, but could potentially be instructive for a number of reasons.[66] First, it would delineate the causal connections between variables of interest. Second, it would be explicit concerning confounding and more clearly direct study design and analysis. Finally, and possibly most importantly, it could be expanded or adapted to different agents and diseases.

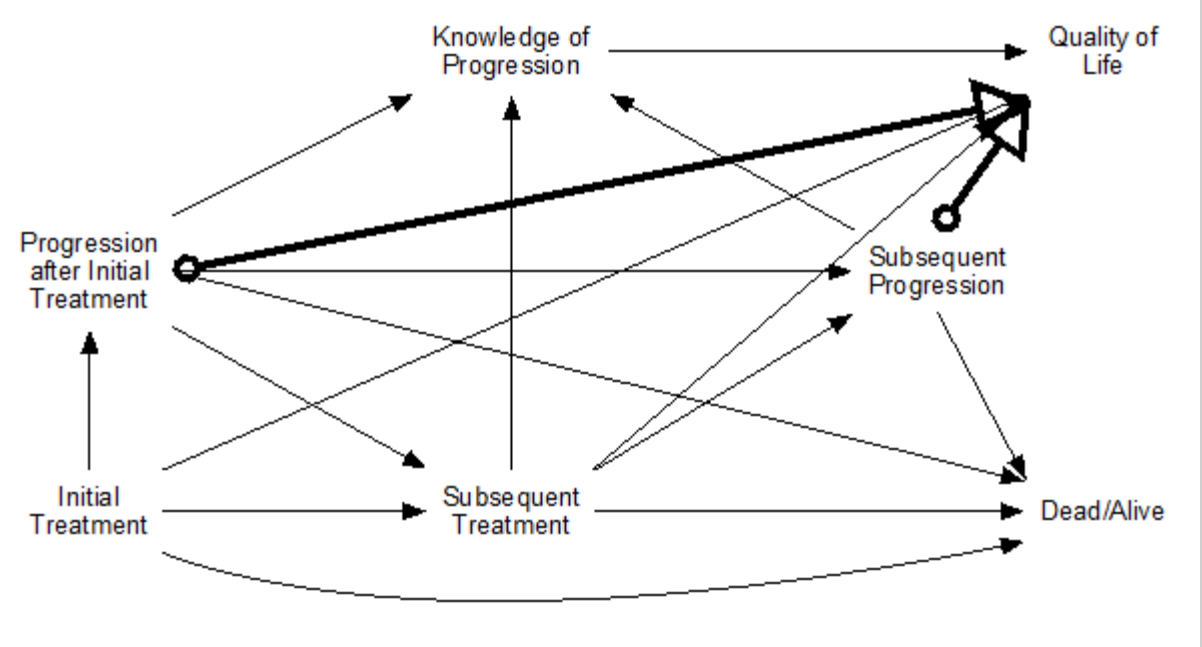
Studies can then be considered, or even designed, with such a framework in mind. Included would be the required *a priori* hypothesis, appropriate measurements of both progression and the QoL measure of interest of interest, and an inclusive approach to evaluating study populations, including, when possible, all therapies received and all relevant subsequent outcomes (progression, quality of life, survival, treatment toxicity). Studies would include an evaluation of the association of PFS to QoL and other clinical outcomes and an assessment of post-progression survival.

A hypothetical study applying this framework to the use of PFS would entail two treatments with similar survival rates and toxicities, so as to not confound QoL outcomes. Next, a patient failing either treatment due to disease progression would not receive further therapy, to eliminate time-dependent confounding. One treatment would result in longer PFS; the study arms are then similar in all aspects *except* for PFS. Under those circumstances, comparison of other clinically important outcomes and their causal or associational relationship with disease progression could be assessed without confounding. A second randomization to being informed about or blinded to PFS status could be conducted to determine if knowledge about PFS by itself impacts QoL outcomes. Patients would be informed that progression is not associated with survival in this case. An analysis could determine if knowledge of PFS itself impacts QoL, specific symptoms, and/or QoL subdomains. Moreover, qualitative debriefing of respondents could inquire about how patients feel about progression, whether they consider it to have inherent value, what PFS

means to them, and whether they associate progression with other endpoints such as survival despite the explanation to the contrary. Because such a study is hypothetical, one is left with largely observational efforts that attempt to examine causal or associational relationships of QoL outcomes with disease progression. Disease progression cannot be manipulated like a treatment to evaluate its effect.

Leaving the notion of a hypothetical study behind, what would be realistic and practical to address the question? First, one must assume or show that disease progression is sufficiently measurable; no design or analysis can obviate large measurement error. Second, a therapeutic failure evidenced by disease progression would be typically followed by subsequent treatment (we assume avoiding second therapy is a potential benefit). If one posited that knowing disease has not progressed improves QoL, a basic causal model might resemble Figure 2.

Figure 2. Basic causal model for treatments, progression, and outcomes



The graph depicts relationships among the variables of interest—causal and confounding. Beginning on the left lower corner, initial treatment may impact disease progression and influence survival either through progression or other mechanisms (arrow from treatment to dead/alive). Treatment can influence QoL through toxicities. The arrow from initial to subsequent treatment reflects that a second line agent, if used, will not be chosen independently. The potential second treatment also requires considering time-dependent confounding. The remainder of the graph can be interpreted in a likewise manner.

The bolded arrows are of primary interest here; that is, the relationship between progression and QoL (e.g., patient-reported outcomes). To define those effects, the others must somehow be accounted for—a potentially difficult task. It is necessary to take into account initial and subsequent treatments, whether patients are aware of progression (and meaning attached), treatment efficacy, and toxicities. While accounting for confounding and other potential biases in this setting might not be insurmountable, the challenges are significant. Even if the many shortcomings noted in the studies reviewed here were absent, the data would be limited to inform the relevant associations.

In summary, there is evidence that PFS provides information on the direct effect of treatment on tumor burden,[22, 67] although there is a lack of consensus regarding its use. This is not surprising given the fact that there is a significant gap in information on the direct and indirect relationship between PFS and QoL or other related outcomes.

Limitations

In this methods project, what was being studied was not a comparison of treatment interventions, but rather the relationship between two outcomes of treatment; therefore the standard approaches used for randomized clinical trials were not applicable. We developed a causal model for addressing this issue, but recognize that this is a hypothetical construct that requires further study and validation.

Existing validated quality assessment tools were not entirely applicable for our purposes. This resulted in the adaptation of the quality assessment tool to include items addressing potential biases unique to the study of the association between PFS and QoL. However the modified quality assessment is unvalidated.

Evidence Gaps

There is a need for prospective research to evaluate: 1) the causal relationship between patient knowledge of progression status and other outcomes of importance to patients, including specific symptoms and overall QoL; 2) patient impressions of the meaningfulness of progression as a health outcome in the absence of association with overall survival; and 3) the extent to which improvement in patient-reported outcomes associated with improvement in PFS is related to a common underlying mechanism, such as tumor shrinkage positively impacting symptomatology and therefore QoL.

Pending further study of the association between PFS and outcomes of importance to patients, there is insufficient evidence to support the use of PFS alone to demonstrate QoL or related outcomes. It is acknowledged that in some situations, measurement of OS is impractical or unfeasible.[9] In such cases, the direct measurement of PFS and QoL is a practical alternative.

Conclusion

The objective of this methods project was to address whether PFS is an outcome related to psychological well-being or QoL in patients with advanced cancer, and in particular, for KQ2, advance solid epithelial cancers. There were no studies that directly addressed the question of the relationship between PFS and patient anxiety, depression, or psychological well-being. Due to limitations in their design, the four studies demonstrating an association between PFS and improvement in QoL or disease symptoms were all of poor quality. There is insufficient evidence to make any conclusions about the association between PFS and improvement in QoL or disease symptoms. It is acknowledged that in some situations, measurement of OS is impractical or unfeasible.[9] In such cases, the direct measurement of PFS and QoL is a practical alternative.

References

1. Pazdur R. Endpoints for assessing drug activity in clinical trials. The Oncologist. 2008;13(Suppl 2):19-21. PMID: 18434634.

2. Sridhara R, Johnson JR, Justice R, et al. Review of oncology and hematology drug product approvals at the US Food and Drug Administration between July 2005 and December 2007. J Natl Cancer Inst. 2010 Feb 24;102(4):230-43. PMID: 20118413.

3. Dancey JE, Dodd LE, Ford R, et al. Recommendations for the assessment of progression in randomised cancer treatment trials. Eur J Cancer. 2009 Jan;45(2):281-9. PMID: 19097775.

4. Hotte S, Bjarnason G, Heng D, et al. Progression-free survival as a clinical trial endpoint in advanced renal cell carcinoma. Curr Oncol. 2011;18(Suppl 2):S11-S9.

5. Shi Q, Sargent DJ. Meta-analysis for the evaluation of surrogate endpoints in cancer clinical trials. Int J Clin Oncol. 2009 Apr;14(2):102-11. PMID: 19390940.

6. Prentice RL. Surrogate endpoints in clinical trials: Definition and operational criteria. Stat Med. 1989;8(4):431-40.

7. Buyse M, Molenberghs G. Criteria for the validation of surrogate endpoints in randomized experiments. Biometrics. 1998;54(3):1014-29.

8. Freedman L, Graubard B. Statistical validation of intermediate endpoints for chronic diseases. Stat Med. 1992;11(2):167-78.

9. Broglio KR, Berry DA. Detecting an overall survival benefit that is derived from progression-free survival. J Natl Cancer Inst. 2009 Dec 2;101(23):1642-9. PMID: 19903805.

10. Amir E, Seruga B, Kwong R, et al. Poor correlation between progression-free and overall survival in modern clinical trials: Are composite endpoints the answer? Eur J Cancer. 2011 Nov 22;in pressPMID: 22115991.

11. Buyse M. Use of meta-analysis for the validation of surrogate endpoints and biomarkers in cancer trials. Cancer J. 2009;15(5):421-5.

12. Buyse M, Burzykowski T, Carroll K, et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. J Clin Oncol. 2007 Nov 20;25(33):5218-24. PMID: 18024867.

13. Saad ED, Katz A, Hoff PM, et al. Progression-free survival as surrogate and as true end point: insights from the breast and colorectal cancer literature. Ann Oncol. 2010 Jan;21(1):7-12. PMID: 19901012.

14. Tang P, Bentzen S, Chen E, et al. Surrogate end points for median overall survival in metastatic colorectal cancer: Literature-based analysis from 39 randomized controlled trials of first-line chemotherapy. J Clin Oncol. 2007 October 10, 2007;25(29):4562-8.

15. Gill S, Berry S, Biagi J, et al. Progression-free survival as a primary endpoint in clinical trials of metastatic colorectal cancer. Curr Oncol. 2011;18(Suppl 2):S5-S10.

16. Yothers G. Toward progression-free survival as a primary end point in advanced colorectal cancer. J Clin Oncol. 2007 Nov 20;25(33):5153-4. PMID: 18024861.

17. Bast RC, Thigpen JT, Arbuck SG, et al. Clinical trial endpoints in ovarian cancer: report of an FDA/ASCO/AACR Public Workshop. Gynecol Oncol. 2007 Nov;107(2):173-6. PMID: 17950384.

18. Markman M. Why overall survival should not be the sole valid primary endpoint of phase 3 ovarian cancer chemotherapy trials. Gynecol Oncol. 2007 Aug;106(2):279-81. PMID: 17662376.

19. Oza A, Castonguay V, Tsoref D, et al. Progression-free survival in advanced ovarian cancer: a Canadian review and expert panel perspective. Curr Oncol. 2011;18(Suppl 2):S20-S7.

20. Burzykowski T, Buyse M, Piccart-Gebhart MJ, et al. Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer. J Clin Oncol. 2008 April 20, 2008;26(12):1987-92.

21. Bergmann L, Hirschfeld S, Morris C, et al. Progression-free survival as an end-point in clinical trials of biotherapeutic agents. Eur J Cancer Suppl. 2007;5(9):23-8.

22. Fleming TR, Rothmann MD, Lu HL. Issues in using progression-free survival when evaluating oncology products. J Clin Oncol. 2009 Jun 10;27(17):2874-80. PMID: 19414672.

23. Stone AM, Bushnell W, Denne J, et al. Research outcomes and recommendations for the assessment of progression in cancer clinical trials from a PhRMA working group. Eur J Cancer. 2011 Aug;47(12):1763-71. PMID: 21435858.

24. Zhuang S, Xiu L, Elsayed Y. Overall survival: A gold standard in search of a surrogate: The value of progression-free survival and time to progression as

end points of drug efficacy. Cancer J. 2009;15(5):395-400.

25. Panageas KS, Ben-Porat L, Dickler MN, et al. When you look matters: the effect of assessment schedule on progression-free survival. J Natl Cancer Inst. 2007 Mar 21;99(6):428-32. PMID: 17374832.

26. Freidlin B, Korn EL, Hunsberger S, et al. Proposal for the use of progression-free survival in unblinded randomized trials. J Clin Oncol. 2007 May 20;25(15):2122-6. PMID: 17513819.

27. Bhattacharya S, Fyfe G, Gray RJ, et al. Role of sensitivity analyses in assessing progression-free survival in late-stage oncology trials. J Clin Oncol. 2009 Dec 10;27(35):5958-64. PMID: 19826121.

28. Amit O, Bushnell W, Dodd L. Blinded central review of the progression-free survival endpoint. The Oncologist. 2010;15(5):492-5.

29. Dodd LE, Korn EL, Freidlin B, et al. Blinded independent central review of progression-free survival in phase III clinical trials: important design element or unnecessary expense? J Clin Oncol. 2008 Aug 1;26(22):3791-6. PMID: 18669467.

30. Ford R, Schwartz L, Dancey J, et al. Lessons learned from independent central review. Eur J Cancer. 2009 Jan;45(2):268-74. PMID: 19101138.

31. Pignatti F, Hemmings R, Jonsson B. Is it time to abandon complete blinded independent central radiological evaluation of progression in registration trials? Eur J Cancer. 2011 Aug;47(12):1759-62. PMID: 21641204.

32. Tang PA, Pond GR, Chen EX. Influence of an independent review committee on assessment of response rate and progression-free survival in phase III clinical trials. Ann Oncol. 2010 Jan;21(1):19-26. PMID: 19875758.

33. Amit O, Mannino F, Stone AM, et al. Blinded independent central review of progression in cancer clinical trials: results from a meta-analysis. Eur J Cancer. 2011 Aug;47(12):1772-8. PMID: 21429737.

34. Oxnard GR, Zhao B, Sima CS, et al. Variability of lung tumor measurements on repeat computed tomography scans taken within 15 minutes. J Clin Oncol. 2011 Aug 10;29(23):3114-9. PMID: 21730273.

35. Axelrod RC. The significance of progression-free survival as a clinical endpoint in oncology Drug Benefit Trends. 201022(2):39-45.

36. Fallowfield LJ, Fleissig A. The value of progression-free survival to patients with advanced-stage cancer. Nat Rev Clin Oncol. 2011;9(1):41-7. PMID: 22009075.

37. Mohamed AF, Hauber AB, Neary MP. Patient benefit-risk preferences for targeted agents in the treatment of renal cell carcinoma. Pharmacoeconomics. 2011;29(11):977-88.

38. Bridges JFP, Mohamed AF, Finnern HW, et al. Patients' preferences for treatment outcomes for advanced non-small cell lung cancer: A conjoint analysis. Lung Cancer. (2012);in press.

39. Johnson JR. End points and United States Food and Drug Administration Approval of oncology drugs. J Clin Oncol. 2003;21(7):1404-11.

40. Juni P, Holenstein F, Sterne J, et al. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. Int J Epidemiol. 2002 Feb;31(1):115-23. PMID: 11914306.

41. Moher D, Pham B, Klassen TP, et al. What contributions do languages other than English make on the results of meta-analyses? J Clin Epidemiol. 2000 Sep;53(9):964-72. PMID: 11004423.

42. Viswanathan M, Ansari MT, Berkman ND, et al. Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions. Agency for Healthcare Research and Quality Methods Guide for Comparative Effectiveness Reviews. AHRQ Publication No. 12-EHC047-EF; March 2012.

43. Turner R, Spiegelhalter D, Smith G, et al. Bias modelling in evidence synthesis. J R Stat Soc 2009;172(1):21-47.

44. Deeks JJ, Dinnes J, D'Amico R, et al. Evaluating nonrandomized intervention studies. Health Technol Assess. 2003;7(27):1-173.

45. Higgins JPT, Green S. The Cochrane handbook for systematic reviews of interventions. Cochrane Handbook, Issue 3. Chichester, West Sussex, UK: John Wiley & Sons, Ltd; 2005.

46. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality borth of randomised and nonrandomised studies of health care interventions. J Epidemiol Community Health. 1998;52(6):377-84.

47. Hao Y. Patient-reported outcomes in support of oncology product labeling claims: regulatory context and challenges. Expert Rev Pharmacoecon Outcomes Res. 2010;10(4):407-20.

48. Meyer KB, Clayton KA. Measurement and analysis of patient-reported outcomes. Methods Mol Biol. 2009;473:155-69.

49. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. Clin Trials. 2004;1(4):368-76.

50. Owens DK, Lohr KN, Atkins D, et al. AHRQ Series Paper 5: Grading the strength of evidence when comparing medical interventions-Agency for Healthcare Research and Quality and the Effective Healthcare Program. J Clin Epidemiol. 2010;63(5):513-23.

51. FDA. Guidance for Industry: Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. May, 2007.

52. EMEA. Methodological considerations for using progression-free survival (PFS) as a primary endpoint in confirmatory trials for registration. January, 2008.

53. . Health Canada Issue Analysis Summary: The use of progression-free survival as the efficacy endpoint for approval of targeted and chemotherapeutic agents for advanced cancer. August, 2007.

54. Zhou C, Wu YL, Chen G, et al. Updated efficacy and quality-of-life (QoL) analyses in OPTIMAL, a phase III, randomized, open-label study of first-line erlotinib versus gemcitabine/carboplatin in patients with EGFR-activating mutation-positive (EGFR Act Mut+) advanced non-small cell lung cancer (NSCLC). ASCO Meeting Abstracts. 2011 June 9, 2011;29(15_suppl):7520.

55. Zhou X, Cella D, Cameron D, et al. Lapatinib plus capecitabine versus capecitabine alone for HER2+ (ErbB2+) metastatic breast cancer: quality-of-life assessment. Breast Cancer Res Treat. 2009;117(3):577-89.

56. Cella D, Pickard AS, Duh MS, et al. Health-related quality of life in patients with advanced renal cell carcinoma receiving pazopanib or placebo in a randomised phase III trial. Eur J Cancer. 2012;48(3):311-23.

57. Sherrill B, Amonkar MM, Sherif B, et al. Quality of life in hormone receptor–positive HER-2+ metastatic breast cancer patients during treatment with letrozole alone or in combination with lapatinib. The Oncologist. 2010 September 1, 2010;15(9):944-53.

58. Siena S, Peeters M, Van Cutsem E, et al. Association of progression-free survival with patient-reported outcomes and survival: results from a randomised phase 3 trial of panitumumab. Br J Cancer. 0000;97(11):1469-74.

59. Lheureux M, Raherison C, Vernejoux JM, et al. Quality of life in lung cancer: Does disclosure of the diagnosis have an impact? Lung Cancer. 2004;43(2):175-82.

60. Otani H, Morita T, Esaki T, et al. Burden on oncologists when communicating the discontinuation of anticancer treatment. Jpn J Clin Oncol. 2011;41(8):999-1006.

61. Shaw J, Dunn S, Heinrich P. Managing the delivery of bad news: An in-depth analysis of doctors' delivery style. Patient Educ Counsel. 2011;Sept 13 [Epub ahead of print].

62. Beaumont JL, Butt Z, Baladi JF, et al. Patient-reported outcomes in a phase III study of everolimus versus placebo in patients with metastatic carcinoma of the kidney that has progressed on vascular endothelial growth factor receptor tyrosine kinase inhibitor therapy. The Oncologist. 2011;16:632-40.

63. Carey MS, Bacon M, Tu D, et al. The prognostic effects of performance status and quality of life scores on progression-free survival and overall survival in advanced ovarian cancer. Gynecol Oncol. 2008;108(1):100-5.

64. Cella D, Bushmakin AG, Cappelleri JC, et al. Baseline quality of life as a prognostic survival tool in patients receiving sunitinib for metastatic renal cell carcinoma. Br J Cancer. 2012;106(4):646-50.

65. Cella D, Cappelleri JC, Bushmakin A, et al. Quality of life predicts progression-free survival in patients with metastatic renal cell carcinoma treated with sunitinib versus interferon alpha. J Oncol Pract. 2009;5(2):66-70.

66. Rubin DB. On the limitations of comparative effectiveness research. Stat Med. 2010;29:1991-5.

67. Wilkerson J, Fojo T. Progression-free survival is simply a measure of a drug's effect while administered and is not a surrogate for overall survival. Cancer J. 2009;15(5):379-85.