IMPORTANCE The strength of association between surrogate end points and survival in oncology is important to understand because surrogate end points are frequently used in oncology clinical trials, supporting US Food and Drug Administration approvals and National Comprehensive Cancer Network guideline recommendations.

OBJECTIVE To identify and evaluate trial-level meta-analyses of randomized clinical trials quantifying the association between a surrogate end point and overall survival in medical oncology. Trial-level correlations test whether treatments that improve the surrogate end point also improve the final end point and are widely considered the strongest evidence to validate a surrogate end point.

EVIDENCE REVIEW Our literature search was built on earlier reported data sets and updated with Google Scholar and MEDLINE searches conducted on December 26, 2014. For MEDLINE, search terms included (“regression” or “correlation”) and “surrogate” and “end point [or endpoint]” and (“oncology” or “cancer”). For Google scholar, search terms included (“regression” or “correlation”) and “surrogate end point [or endpoint]” and “overall survival” and “trial level.” A total of 108 abstracts were retrieved, and 62 articles were read in full in addition to articles identified through prior reviews.

FINDINGS We found 36 articles in which 65 specific correlations between a surrogate end point and survival were identified. Surrogate end points were studied in the neoadjuvant, adjuvant, locally advanced, and metastatic settings. The most common sources for trials included in the 36 articles were systematic reviews of the published literature (10 of 36; 28%), and published literature and meeting abstracts (14 of 36; 39%). Four meta-analyses (11%) used a convenience sample, and only 5 studies (14%) attempted to include unpublished trials by surveying clinical trial registries. Among these 5 studies, only 352 of 684 eligible trials (51.1%) were included in the analyses. More than half of reported correlations (34 of 65; 52%) were of low strength ($r<0.7$). Approximately a quarter (16 of 65; 25%) were of medium strength ($0.7<r<0.85$), and 15 of 65 (23%) were highly correlated ($r\geq0.85$) with survival.

CONCLUSIONS AND RELEVANCE Most trial-level validation studies of surrogate end points in oncology find low correlations with survival. All validation studies use only a subset of available trials. The evidence supporting the use of surrogate end points in oncology is limited.
Although there is growing recognition that adopting new medical practices based on improvements in surrogate outcomes can lead to misleading conclusions, surrogates continue to play a prominent role in oncology. In the United States, most new cancer drugs are approved based on surrogate measures, such as response rate (RR) and progression-free survival (PFS), through the US Food and Drug Administration’s (FDA) Accelerated Approval pathway. In addition, clinical practice guidelines often expand treatment recommendations for approved drugs based on studies assessing surrogates. For example, the drug carfilzomib (Kyprolis; Onyx Pharmaceuticals) received FDA approval in 2012 based on RR for relapsed and/or refractory multiple myeloma. In 2013, the National Comprehensive Cancer Network (NCCN) expanded the drug’s indication to include untreated myeloma—based on RR in 2 small, uncontrolled phase 1/2 studies (category 2A). Notably, NCCN endorsement obliges many commercial insurers and Medicare to reimburse the therapy. Thus, surrogate end points are used both to introduce new cancer drugs and to promote new uses for the drugs we already have.

The use of surrogate end points in oncology has led to use of toxic drugs that do not improve survival. For example, bevacizumab (Avastin; Roche/Genentech) gained FDA accelerated approval for metastatic breast cancer in 2008 based on evidence that it could improve PFS. In 2011, that approval was withdrawn when multiple randomized trials confirmed that bevacizumab did not improve overall survival and that gains in PFS were more modest than those seen in the earlier trial. Nevertheless, Medicare and other insurers are still obliged to pay for bevacizumab because it remains endorsed by the NCCN for this indication.

For surrogates such as PFS or tumor RR to be reliable, it is important that they are validated. Several authors have described statistical methods to validate surrogates, but the framework that uses a hierarchy (Figure 1) is the most useful. In this model, level 3 is surrogacy based on biological plausibility alone. Level 2 occurs when a strong correlation exists between the surrogate and the final endpoint across cohorts or at the level of the individual patient. The highest level is trial-level surrogacy, meaning that across many trials treatments that improve the surrogate end point also improve the final outcome. Trial-level surrogacy directly asks the question that faces regulators and guideline writers: “If a drug improves a surrogate, will it also improve survival?” Others have noted that individual patient correlations do not directly validate surrogate measures.

Establishing trial-level (level 1) surrogacy requires a meta-analysis of all randomized trials on a question of interest, with each trial serving as a unique data point. The x coordinate is typically the improvement in the surrogate when a drug is used, and the y coordinate is the improvement in the final outcome. Regression analysis is typically performed to assess what correlation exists between the change in the surrogate end point and the change in overall survival, across the randomized studies. The correlation coefficient (r) for the analysis, ranging from 0 to 1, provides a measure of the strength of the surrogate-survival correlation.

Two general medical examples can illustrate this approach. Consider the question of hypertension (the surrogate), blood pressure control (the intervention), and adverse cardiovascular end points (the final outcome). A level 2 analysis would show that groups of patients with higher average blood pressure have worse cardiovascular outcomes or that this correlation is observed at an individual patient level. A level 1 or trial-level analysis would show that across many trials, drugs that lowered blood pressure also decreased cardiovascular events. If a drug lowered blood pressure by 10%, and another lowered it by 20%, we would “validate” hypertension as a surrogate if there were fewer cardiovascular events for the drug with greater blood pressure reduction.

Looking at many blood pressure trials, one could perform regression analysis and draw a general conclusion about the relationship between blood pressure reduction and adverse outcomes. In fact, this analysis has been performed.

The case of premature ventricular contractions (PVCs) post-myocardial infarction illustrates the importance of level 1 analysis. Although studies showed a correlation between frequent PVCs and death (level 2), and certain anti-arrhythmic drugs reduce PVCs, it did not follow that anti-arrhythmic drugs (level 1) that lowered PVCs reduced mortality. In fact, a randomized clinical trial showed that these drugs actually increased mortality. Thus, only a level 1 analysis asks the right question: “Can we trust a reduction in PVCs as a reliable surrogate for improved survival in pharmacologic trials?” Because oncology typically has dozens of trials examining both surrogate end points and survival in many clinical situations, performing level 1 analyses is more straightforward than in many other medical fields. For these reasons, we only examined level 1 studies.

We sought to examine all level 1 studies aiming to validate a surrogate in oncology. Specifically, we asked what percentage of these meta-analyses used data from published and unpublished studies. We also ascertained the median number of studies used to establish surrogacy in these studies and the reported correlation coefficients (r).

**Methods**

**Eligible Studies**

We identified meta-analyses of randomized clinical trials quantifying the association between a surrogate end point—RR, pathologic complete response, locoregional control, disease-free sur-

---

**Figure 1. Hierarchy of Evidence for Surrogate End Point and Overall Survival Association**

A hierarchy of evidence used to validate surrogate end points in oncology.
vival, event-free survival, time to progression, and/or PFS—and overall survival (OS) in medical oncology. As such, we excluded studies of radiation therapies, surgery, procedures, or supportive measures.

We sought only studies that examined trial-level correlations in meta-analyses of randomized clinical trials—ie, level 1 surrogate studies. We placed no restriction on the setting of medical therapy; thus, we include analyses in the neoadjuvant, adjuvant, locally advanced, and metastatic settings.

We excluded all analyses that did not examine trial-level (level 1) surrogacy. Specifically, we excluded analyses that only studied individual patient level correlations or those that did not examine the difference between investigational and control arms, considering each trial arm as a unique data point (level 2). We excluded conference abstracts, letters to the editor, and descriptive reviews. Finally, we excluded meta-analyses that used each treating center or nation (rather than each trial) as a unique data point. We were not convinced that such analyses accurately capture the question of our study because the individual data points in such analyses are no longer independent (comparisons are both within and across trial) and may spuriously inflate trial-level correlation. Nevertheless, including these studies would not have changed our results.

**Literature Search**

We constructed our data set by updating prior systematic reviews. An analysis by Sherrill et al. conducted in October 2010, systematically identified all studies prior to that date that studied surrogate end points in oncology. Over 1000 articles were screened in that investigation, and all articles assessing the relationship between a surrogate end point and overall survival were listed. We then evaluated this list to ensure that the articles performed trial-level correlation—all other analyses were discarded. We also drew on a prior report commissioned for the National Institutes of Clinical Excellence, which identified meta-analyses of surrogate end points in malignant solid tumors. Again, only trial-level analyses were selected from this set.

We built on this data set, updating it for the last 4 years. Specifically, Google Scholar and MEDLINE were searched with the following search terms on December 26, 2014: For MEDLINE, search terms included (“regression” or “correlation”) and “surrogate” and “end point [or endpoint]” and (“oncology” or “cancer”). For Google scholar, search terms included (“regression” or “correlation”) and “surrogate end point [or endpoint]” and “overall survival” and “trial level.” A total of 108 abstracts were retrieved, and 62 articles were read in full.

**End Points Assessed**

Descriptive statistics were ascertained for the following end points: the percentage of meta-analyses that used data from published and unpublished studies, the number of individual randomized trials included in each surrogate meta-analyses, and the reported correlation coefficients. We considered *convenience sample* to mean a set of trials that the authors were able to obtain readily. We only credited authors with performing a search of the published literature, meeting abstracts, and/or unpublished trials if they performed a systematic search. We scored strength of trial-level correlation according to a modification to surrogate criteria proposed by the Institute of Quality and Efficiency in Health Care: low correlation ($r ≤ 0.7$), medium strength correlation ($r > 0.7$ to $r < 0.85$), and high correlation ($r ≥ 0.85$). The specific cut points were adapted to function even when confidence intervals were not presented. Finally, for meta-analyses using published trials, abstracts, and registered studies, we investigated what percentage of eligible studies were ultimately included in analyses.

**Results**

We identified 36 articles that met our inclusion criteria and were considered trial-level or level 1 analyses. Claims of surrogacy were examined for 19 distinct clinical questions in the neoadjuvant, adjuvant, metastatic, and locally advanced settings (eTable in the Supplement).

The sources of randomized clinical trials included in the 36 articles most commonly were systematic reviews of the published literature (10 of 36; 28%) and published literature and...
meeting abstracts (14 of 36; 39%). Four meta-analyses (11%) used a convenience sample, and only 5 studies (14%) attempted to include unpublished trials by surveying clinical trial registries. Figure 2 summarizes search strategies.

Across the 36 studies, we noted 65 unique trial-level analyses. The median number of individual randomized clinical trials used in these analyses was 21, with as few as 9 and as many as 191 trials. For each analysis, we documented the study characteristics and the correlation coefficient (r) for all trial-level analyses (Table 1). There were 65 trial-level correlations scored according to criteria proposed by the Institute of Quality and Efficiency in Health Care (Figure 3). More than half of trial-level correlations (34 of 65; 52%) were of low strength (r = 0.7). Approximately a quarter (16 of 65; 25%) were of medium strength (r > 0.7 to r < 0.85), and 15 of 65 (23%) were highly correlated (r ≥ 0.85) with survival.

We further examined the 5 studies that analyzed both published and unpublished trials. Although these articles searched the broadest possible collection of trials, they were unable to obtain data from all relevant studies. Table 2 lists the numbers of eligible and included trials for each of these meta-analyses. Notably, only 352 of 684 eligible studies (51.1%) were ultimately included. Reasons for omission included missing or lost data, missing information on progression, or missing median PFS or median OS reports.

We also examined the 15 correlations of high strength. Six occurred in the adjuvant setting, 4 in the metastatic setting, 3 in the locally advanced setting, and 3 in the metastatic setting. The median number of individual randomized trials included in these analyses was 14. Notably, 3 high correlations occurred in the same setting—adjuvant colorectal cancer—and 4 others occurred in settings where multiple other studies had found lesser correlations, specifically: metastatic breast cancer and colon cancer (Table 1). In both these settings, the largest analyses conducted to date found low correlation.

Discussion

Most trial-level meta-analyses in oncology found low correlation between a surrogate end point and overall survival, and all were based on only a subset of clinical trial evidence. Our findings call into question the widespread use of surrogate end points in oncology as a basis for treatment decisions.

Trial-level or level 1 meta-analyses are the highest form of evidence that surrogate end points can predict efficacy regarding final outcomes, such as overall survival, for new therapies or new combinations (Figure 1). The majority of these claims resulted in correlations of low (52%) or medium strength (25%) (Figure 3), which are generally considered insufficient evidence on which to base clinical or regulatory decisions.

We were unable to identify any analyses that looked at all randomized trials on a topic. Most analyses relied on published articles (28%) or articles and meeting abstracts (39%), and only 5 authors (14%) sought unpublished trials. Even when authors were able to identify published and unpublished studies, they were unable to obtain data from all eligible trials. Only 51.5% of eligible studies were ultimately included in even the most rigorous meta-analyses (Table 2). The major barriers to including more studies that we noted were that primary investigators did not provide data, or primary study publications did not include information that the meta-analysts needed, such as the rates of surrogate improvement or improvement in overall survival.

Our findings suggest that most correlations of surrogacy in oncology are based on only a subset of potentially informative trials. Unpublished trials (and those that do not report data that meta-analysts require) may have poorer correlations than those that are published, and such discrepancies may contribute to reluctance among sponsors and authors to submit those findings to journals and conferences and to a poorer likelihood of acceptance.

The use of surrogate end points in oncology, particularly PFS, has grown in recent years and it is frequently the primary end point of cancer clinical trials and the basis for regulatory approval of novel agents. Surrogate end points are also used as the basis of many NCCN recommendations, which obliges many insurers and Medicare to cover those drugs for those indications. Our analysis confirms that the use of this end point occurs beyond the specific settings in which it has been validated. Simply because a surrogate is measurable does not make it predictive or meaningful, and our analysis builds on a prior study that concludes that the association between surrogates and survival in oncology is generally poor.

High-profile medical reversals, such as the case of bevacizumab in breast cancer, emphasize the dangers of excessive reliance on unvalidated surrogates. Although bevacizumab was approved for breast cancer on the basis of an improvement in PFS, this surrogate-survival correlation is not well supported. We identified 8 meta-analyses examining whether gains in PFS predict overall survival in metastatic breast cancer. Six reported low correlation; 1 reported medium correlation; and only 1 reported a strong correlation (Table 1). Despite this evidence, in 2012 another drug, everolimus, was approved for metastatic breast cancer based on improvement in PFS. In 2014, updated follow-up from the everolimus study did not find an overall survival benefit. In 2015, palbociclib received accelerated approval for metastatic breast cancer on the basis of improvement in PFS. Again, there is no demonstration of an overall survival advantage. Both palbociclib and everolimus remain on the market.

In 2007, liposomal doxorubicin received FDA approval for treatment of multiple myeloma based on a delay in time to progression when given in combination with bortezomib compared to bortezomib alone. In our investigation, we were unable to identify any analysis validating this surrogate and survival in multiple myeloma, and an updated analysis of liposomal doxorubicin confirms that the drug confers no survival benefit. Liposomal doxorubicin remains FDA approved.

The FDA’s use of pathologic complete response as grounds for accelerated approval in the neoadjuvant setting (the basis for the approval of pertuzumab [Perjeta; Roche/Genentech]) also appears to run counter to both meta-analyses we identified. Both meta-analyses on this topic found that improvements in pathologic complete response were poorly correlated with subsequent event-free survival and OS.

Figure 1: Meta-analyses of correlation between surrogate end points and survival outcomes. The use of surrogate end points in oncology, particularly PFS, has grown in recent years and is frequently the primary end point of cancer clinical trials and the basis for regulatory approval of novel agents. Surrogate end points are also used as the basis of many NCCN recommendations, which obliges many insurers and Medicare to cover those drugs for those indications. Our analysis confirms that the use of this end point occurs beyond the specific settings in which it has been validated.
Table 1. Summary of Trial-Level Surrogate Meta-analyses Testing the Validity of Surrogate End Points in Oncology

<table>
<thead>
<tr>
<th>Source</th>
<th>Tumor</th>
<th>Trial Set</th>
<th>Trials Examined, No.</th>
<th>Years Included</th>
<th>Trial-Level Correlations* Between Surrogate End Points and Survival, r</th>
<th>Surrogacy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Surrogacy in the Neoadjuvant Setting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cortazar et al,35 2014</td>
<td>Breast</td>
<td>Articles</td>
<td>12</td>
<td>Published 1990-2011</td>
<td>0.17⁹</td>
<td>0.49⁹</td>
</tr>
<tr>
<td>Berruti et al,36 2014</td>
<td>Breast</td>
<td>Articles and abstracts</td>
<td>29</td>
<td>Enrolled patients 1990-2009</td>
<td>0.28⁹</td>
<td>0.30⁹</td>
</tr>
<tr>
<td>Michiels et al,40 2009</td>
<td>HNSCC</td>
<td>Articles and abstracts</td>
<td>30</td>
<td>Accrual period 1965-2000</td>
<td>0.53⁹</td>
<td>0.79⁹</td>
</tr>
<tr>
<td><strong>Surrogacy in the Adjuvant Setting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ng et al,37 2008</td>
<td>Breast</td>
<td>Articles</td>
<td>126</td>
<td>1966-2006</td>
<td>NA</td>
<td>0.62⁹</td>
</tr>
<tr>
<td>Sargent et al,25 2005</td>
<td>Colon</td>
<td>Convenience sample</td>
<td>18</td>
<td>Accrual period 1977-1999</td>
<td>NA</td>
<td>0.94⁹</td>
</tr>
<tr>
<td>Buyse et al,23 2008</td>
<td>Colon</td>
<td>Not well described</td>
<td>10</td>
<td>Not reported</td>
<td>NA</td>
<td>0.94⁹</td>
</tr>
<tr>
<td>Burzykowski et al,38 2008</td>
<td>Colon</td>
<td>Convenience sample</td>
<td>21</td>
<td>Accrual period 1977-1999</td>
<td>NA</td>
<td>0.95⁹</td>
</tr>
<tr>
<td>Oba et al,39 2013</td>
<td>Gastric</td>
<td>Articles and registries</td>
<td>14</td>
<td>Closed to accrual before 2004</td>
<td>NA</td>
<td>0.98⁹</td>
</tr>
<tr>
<td>Michiels et al,40 2009</td>
<td>HNSCC</td>
<td>Articles and abstracts</td>
<td>9</td>
<td>Accrual period 1965-2000</td>
<td>0.84⁹</td>
<td>0.93⁹</td>
</tr>
<tr>
<td>Maquen et al,41 2013</td>
<td>NSCLC</td>
<td>Articles, abstracts, and registries</td>
<td>17</td>
<td>Accrual period 1988-2003</td>
<td>NA</td>
<td>0.96⁹</td>
</tr>
<tr>
<td><strong>Surrogacy in the Metastatic Setting</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bruzzi et al,42 2005</td>
<td>Breast</td>
<td>Articles</td>
<td>10</td>
<td>Published 1991-2001</td>
<td>Delta response vs delta median OS $R = 0.45g$; log of the HR, $R = 0.32g$</td>
<td>NA</td>
</tr>
<tr>
<td>Hackshaw et al,43 2005</td>
<td>Breast</td>
<td>Articles</td>
<td>46 for RR, 26 for PFS</td>
<td>Published 1978-2005, accrual period 1968-2002</td>
<td>0.58⁹</td>
<td>0.75⁹</td>
</tr>
<tr>
<td>Burzykowski et al,44 2008</td>
<td>Breast</td>
<td>Articles and abstracts</td>
<td>11</td>
<td>Published 1999-2008, accrual period 1981-1990</td>
<td>0.47⁹</td>
<td>0.48⁹</td>
</tr>
<tr>
<td>Miksad et al,45 2008</td>
<td>Breast, anthracycline only</td>
<td>Articles and abstracts</td>
<td>16</td>
<td>Published 1987-2005</td>
<td>NA</td>
<td>0.7⁹</td>
</tr>
<tr>
<td>Miksad et al,46 2008</td>
<td>Breast, taxane only</td>
<td>Articles and abstracts</td>
<td>15</td>
<td>Published 1996-2006</td>
<td>NA</td>
<td>0.59⁹</td>
</tr>
<tr>
<td>Sherrill et al,47 2008</td>
<td>Breast</td>
<td>Articles</td>
<td>67</td>
<td>Published 1994-2006</td>
<td>NA</td>
<td>0.55⁹</td>
</tr>
<tr>
<td>Wilkerson and Fojo,48 2009</td>
<td>Breast</td>
<td>Articles</td>
<td>14</td>
<td>Published 2000-2008</td>
<td>Correlation based on HR $R = 0.88¹⁰$; based on delta $R = 0.35¹⁰$</td>
<td>NA</td>
</tr>
<tr>
<td>Beuchemin et al,49 2014</td>
<td>Breast</td>
<td>Articles, abstracts, and registries</td>
<td>144</td>
<td>Databases for 1950-2010</td>
<td>NA</td>
<td>0.43⁹</td>
</tr>
<tr>
<td>Petrelli and Barni,49 2014</td>
<td>Breast, targeted therapies only</td>
<td>Articles</td>
<td>21</td>
<td>2000-2012</td>
<td>NA</td>
<td>0.7⁹</td>
</tr>
<tr>
<td>Wilkerson and Fojo,48 2009</td>
<td>Colorectal</td>
<td>Articles</td>
<td>23</td>
<td>1993-2009</td>
<td>Correlation based on HR $R = 0.72¹⁰$; based on delta, 0.80⁹</td>
<td>NA</td>
</tr>
<tr>
<td>Johnson et al,50 2006</td>
<td>Colorectal</td>
<td>Articles and abstracts</td>
<td>146</td>
<td>Before 1992-2005</td>
<td>0.32⁹</td>
<td>0.57⁹</td>
</tr>
<tr>
<td>Chirila et al,51 2012</td>
<td>Colorectal</td>
<td>Articles and abstracts</td>
<td>62 (PFS &amp; TTP), 35 (PFS)</td>
<td>1991-2009</td>
<td>NA</td>
<td>0.69³¹ and 0.77³⁹</td>
</tr>
<tr>
<td>Sidhu et al,52 2013</td>
<td>Colorectal</td>
<td>Articles and abstracts</td>
<td>24</td>
<td>2000-2011</td>
<td>0.41⁹</td>
<td>0.86⁹</td>
</tr>
<tr>
<td>Buyse et al,53 2000</td>
<td>Colorectal, first-line therapy</td>
<td>Pooled articles from 4 meta-analyses that made up convenience sample, published articles only</td>
<td>25</td>
<td>1987-1996</td>
<td>0.62⁹</td>
<td>NA</td>
</tr>
<tr>
<td>Tang et al,54 2007</td>
<td>Colorectal, first-line therapy</td>
<td>Articles and abstracts</td>
<td>39</td>
<td>1990-2005</td>
<td>0.39⁹</td>
<td>0.74⁹</td>
</tr>
<tr>
<td>Giessen et al,55 2013</td>
<td>Colorectal, first-line therapy</td>
<td>Articles and abstracts</td>
<td>50</td>
<td>2000-2012</td>
<td>NA</td>
<td>0.87³⁹</td>
</tr>
</tbody>
</table>

(continued)
Table I. Summary of Trial-Level Surrogate Meta-analyses Testing the Validity of Surrogate End Points in Oncology (continued)

<table>
<thead>
<tr>
<th>Source</th>
<th>Tumor</th>
<th>Trial Set</th>
<th>Trials Examined, No.</th>
<th>Years Included</th>
<th>Surrogate End Points and Survival, $r$</th>
<th>A1</th>
<th>A2</th>
<th>Surrogacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shi et al,54 2015</td>
<td>Colorectal, first-line therapy</td>
<td>Convenience sample</td>
<td>22</td>
<td>1997-2006</td>
<td>NA</td>
<td>0.73</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>Giessen et al,55 2013</td>
<td>Colorectal, first-line therapy, monoclonal antibodies only</td>
<td>Articles and abstracts</td>
<td>19</td>
<td>2000-2012</td>
<td>NA</td>
<td>0.47</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Giessen et al,57 2015</td>
<td>Colorectal, second-line therapy</td>
<td>Articles and abstracts</td>
<td>23</td>
<td>2000-2013</td>
<td>0.58; 0.73; Medium PFS, low RR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amir et al,59 2012</td>
<td>FDA-approved therapies for metastatic tumors with SPP $&lt;$12 mo</td>
<td>FDA-approved therapies</td>
<td>19</td>
<td>2002-2012</td>
<td>NA</td>
<td>0.64</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Amir et al,59 2012</td>
<td>FDA-approved therapies for metastatic tumors with SPP $&gt;$12 mo</td>
<td>FDA-approved therapies</td>
<td>7</td>
<td>2002-2012</td>
<td>NA</td>
<td>0.38</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Paoletti et al,60 2013</td>
<td>Gastric</td>
<td>Articles</td>
<td>20</td>
<td>1970-2010 (but closed to accrual by 2006)</td>
<td>NA</td>
<td>0.78</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>Shitara et al,61 2014</td>
<td>Gastric</td>
<td>Articles and abstracts</td>
<td>10</td>
<td>2002-2013</td>
<td>NA</td>
<td>0.36</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Flaherty et al,62 2014</td>
<td>Metastatic melanoma</td>
<td>Articles and abstracts</td>
<td>12</td>
<td>2006-2013</td>
<td>NA</td>
<td>0.89</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Lee et al,63 2011</td>
<td>NHL, aggressive lymphomas</td>
<td>Articles and abstracts</td>
<td>17</td>
<td>1978-2005</td>
<td>NA</td>
<td>0.81</td>
<td>Medium EFS-OS</td>
<td></td>
</tr>
<tr>
<td>Johnson et al,60 2006</td>
<td>NSCLC</td>
<td>Articles and abstracts</td>
<td>191</td>
<td>Not reported</td>
<td>0.40; 0.44; Low RR, low TTP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hayashi et al,64 2013</td>
<td>NSCLC</td>
<td>Articles and abstracts</td>
<td>18</td>
<td>Published 2000-2011</td>
<td>NA</td>
<td>0.29</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Hotta et al,65 2009</td>
<td>NSCLC, first-line therapy</td>
<td>Articles, abstracts and registries</td>
<td>54</td>
<td>1994-2006</td>
<td>NA</td>
<td>0.57</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Wilkerson and Fojo,66 2009</td>
<td>Ovarian</td>
<td>Articles</td>
<td>15</td>
<td>1983-2008</td>
<td>Correlation based on HR, 0.85; Medium based on delta</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delay et al,67 2012</td>
<td>RCC</td>
<td>Articles and abstracts</td>
<td>31</td>
<td>Published 1997-2010</td>
<td>0.78; Delta R = 0.54; Low delta TTP and PFS, medium RR, and medium HR PS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnson et al,82 2014</td>
<td>RCC</td>
<td>Articles and registries</td>
<td>11</td>
<td>Published before 2013</td>
<td>NA</td>
<td>0.70</td>
<td>Low targeted therapies; correlation among immunotherapies deemed too low to present</td>
<td></td>
</tr>
<tr>
<td>Foster et al,83 2011</td>
<td>SCLC</td>
<td>Articles</td>
<td>9</td>
<td>Accrual period 1987-2000</td>
<td>0.60; 0.75; Low RR, medium PFS</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Surrogacy in the Locally Advanced Setting**

| Michiels et al,80 2009         | HNSCC, locally advanced, concomitant chemotherapy | Articles and abstracts | 50 | Accrual period 1997-2001 | Duration of LRC vs OS, 0.72; $R = 0.86$; Medium duration of LRC-OS, high EFS-OS |
| Mauguen et al,81 2013          | NSCLC, concurrent             | Articles, abstracts, registries | 15 | Accrual period 1984-2001 | NA; 0.98; High |
| Mauguen et al,81 2013          | NSCLC, sequential             | Articles, abstracts, registries | 8  | Accrual period 1984-1999 | NA; 0.98; High |

Abbreviations: A1 and A2, analyses 1 and 2; CR, complete response; DFS, disease-free survival; EFS, event-free survival; FDA, US Food and Drug Administration; HNSCC, head and neck squamous cell cancer; HR, hazard ratio; LRC, locoregional control; NA, not applicable; NHL, non-Hodgkin lymphoma; NSCLC, non-small-cell lung cancer; OS, overall survival; pCR, pathologic complete response; PFS, progression-free survival; PR, partial response; RCC, renal cell cancer; RR, response rate; SCLC, small-cell lung cancer; SPP, survival post protocol; TTP, time to progression.


d Surrogate end point, pCR; survival end point, EFS.

e Surrogate end point, PR; survival end point, OS.

Strength of correlation is scored as trial level $r \leq 0.7$ indicates low correlation; $r = 0.7$ to $< 0.85$, medium correlation; and $r \geq 0.85$, high correlation. Many trials performed multiple analyses, herein categorized as A1 and A2 simply as terms of reference for the purpose of organization; no chronological sequence or clinical importance is implied by these labels.

e Surrogate end point, pCR; survival end point, EFS.

Limitations

There are several limitations to our study. First, the specific thresholds we used to grade the strength of correlation were adapted from a guidance document but have not been externally validated and are subject to disagreement. Although we believe that few would dispute that a correlation coefficient ($r$) below 0.7 is weak.

Copyright 2015 American Medical Association. All rights reserved.
support for a surrogate end point, some may be critical of the thresholds for medium and high strength. For this reason, we provide all coefficients in Table 1 to permit alternative analyses.

Second, we cannot say whether correlations will improve or weaken based on a more comprehensive assessment of clinical trials. However, prior research has found that unpublished trials are different than published studies. Just as meta-analyses on specific clinical topics may change with consideration of published and unpublished evidence, the strength of surrogacy may also change with a more comprehensive collection of data. However, this is an assumption and should be treated as such. Unfortunately, we were unable to identify even 1 study that considered the totality of the evidence.

Third, the use of crossover in cancer clinical trials is thought to complicate the role of surrogacy in oncology. Unidirectional crossover from the control or placebo arm of cancer trials to the experimental drug (but not vice versa) is thought to explain why some cancer drugs that improve PFS fail to improve OS. However, crossover is unlikely to affect our analysis because most of the individual randomized trials included in the surrogate validation studies we examined occurred prior to the widespread use of crossover.

Moreover, there are several reasons to question the common narrative regarding crossover. First, effective cancer drugs can show overall survival benefits despite crossover. Second, favorable PFS but negative OS results seen with everolimus, palbociclib, and bevacizumab in metastatic breast cancer occurred in the absence of crossover. And finally, there are several alternative interpretations for the effect of crossover, including that crossover can hide the harms of a drug or a mask a lack of clinical benefit. For these reasons, we do not believe that crossover alters our findings.

Finally, a recent study suggests that approving new drugs based on surrogates (even inaccurate ones) may provide a greater benefit for society than waiting for OS data. The authors found favorable societal benefits for approving non–small-cell lung cancer (NSCLC) drugs with PFS benefits of 3 months or more. This occurred despite the fact that the authors noted a poor correlation (as we also did in Table 1) between net mean PFS and OS in NSCLC cancer (r = 0.56). The major limitation of this analysis, like other modeling studies, is that it rests on a number of assumptions. For instance, the authors likely overestimate the delay between PFS and OS results by using date of publication rather than the data cutoff date. Also the authors draw broad conclusions from a very small set of trials, ie, the few NSCLC trials with greater than 3-month PFS gains from an overall set of 27 trials. Finally, the model does not account for the wider unintended consequences of approval based on weak surrogates, which may encourage the pharmaceutical industry to pursue even more cancer drugs with marginal to no benefits.

Conclusions

Most trial-level validation studies of surrogate end points in oncology find low or medium strength correlations with overall survival. All validation studies use only a subset of available trials. The evidence that surrogate end points predict overall survival in oncology is limited.

ARTICLE INFORMATION

Accepted for Publication: April 11, 2015.

Published Online: June 22, 2015.

Author Contributions: Dr Prasad had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Prasad.

Acquisition, analysis, or interpretation of data: Prasad, Kim, Burotto, Vandross.

Drafting of the manuscript: Prasad, Kim.

Critical revision of the manuscript for important intellectual content: Prasad, Kim, Burotto, Vandross.

Statistical analysis: Prasad, Kim.

Administrative, technical, or material support: Prasad, Burotto, Vandross.

Study supervision: Prasad.

Conflict of Interest Disclosures: None reported.

Disclaimer: The views and opinions of Drs Prasad, Kim, and Burotto do not reflect those of the National Cancer Institute.
REFERENCES
45. Willerson J, Jofo T. Progression-free survival is simply a measure of a drug’s effect while
Association Between Surrogate End Points and Survival in Oncology

Original Investigation Research

analysis from 39 randomized controlled trials of surrogate end points for median overall survival in metastatic colorectal cancer trials.

52
surrogate for overall survival in metastatic breast cancer:

48

50


52

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86
Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. JAMA. 1998;279(4):281-286.

87

88

89
Dickerson K, Min Y, Meinert CL. Factors influencing publication of research results.


Editor’s Note

Faster Drug Approvals Are Not Always Better and Can Be Worse

Rita F. Redberg, MD, MSc

A shared goal of all health professionals is to relieve suffering and prolong life. At times these goals are at odds, particularly in oncology care. Patients with severe disease and low chance of survival may be offered therapies in the hope of buying a few more weeks or even months. However, the treatments themselves are often toxic, with many unpleasant adverse effects—nausea, pain, vomiting, hair loss, and others—that detract from quality of life and result in patients spending more time in the hospital and clinic and less time at home. It is a difficult choice: extend life, or offer higher quality of life at home.

Or is it? This choice assumes that the drugs really do extend life. The analysis by Prasad et al1 shows that many oncology drugs have never been shown to increase survival, meaning that patients suffer reduced quality of life but do not necessarily gain extra time. In the effort to get more new drugs on the market more quickly, many agents are approved on the basis of surrogate end points, such as progression-free survival. Unfortunately, these surrogate end points do not necessarily translate into any mortality benefits, in which case patients experience the toxic effects but do not get any extra time.

Even when the US Food and Drug Administration takes the dramatic (and rare) step of withdrawing approval (as it should when a drug approved based on surrogate markers later turns out to not offer any mortality benefit, as was the case for bevacizumab (Avastin; Genentech/Roche) for metastatic breast cancer, women continue to be treated with this harmful drug although we know it will not extend life.

The 21st Century Cures legislation2 now being considered by Congress would make this situation much worse by allowing accelerated approval of new drugs based on preliminary results of clinical trials on surrogate markers. It even has a clause that stipulates that a drug on a restricted hospital or industry formulary list is actually approved for the indication and can be prescribed for that purpose. In fact, it introduces 2 more categories of fast drug approvals. However, in 2014 over 60% of approved drugs were already approved using expedited review or orphan drug status, which also has reduced evidence standards for safety and effectiveness.3 We must reduce drug approvals based on unreliable surrogates and change practice when clinical studies show no survival benefit. If there are no drugs available that can extend life with tolerable adverse effect, we must extend warmth and compassion and support in a continued therapeutic relationship without further chemotherapy.

In our rush to find new effective treatments, we should not harm our patients with ineffective toxic ones.

Conflict of Interest Disclosures: None reported.

