Complete Cytogenetic Response and Major Molecular Response as Surrogate Outcomes for Overall Survival in First-Line Treatment of Chronic Myelogenous Leukemia: A Case Study for Technology Appraisal on the Basis of Surrogate Outcomes Evidence

Ciani Oriana, MSc¹,*; Hoyle Martin, PhD¹; Pavey Toby, PhD²; Cooper Chris, MA¹; Garside Ruth, PhD³; Rudin Claudius, MD⁴; Taylor Rod, PhD⁵

¹PenTAG, Institute of Health Services Research, University of Exeter Medical School, Exeter, UK; ²School of Human Movement Studies, University of Queensland, Brisbane, Australia; ³European Centre for Environment and Human Health, University of Exeter Medical School, Truro, UK; ⁴Royal Devon and Exeter NHS Foundation Trust, Exeter, UK

ABSTRACT

Objectives: In 2012, the National Institute for Health and Care Excellence assessed dasatinib, nilotinib, and standard-dose imatinib as first-line treatment of chronic phase chronic myelogenous leukemia (CML). Licensing of these alternative treatments was based on randomized controlled trials assessing complete cytogenetic response (CCyR) and major molecular response (MMR) at 12 months as primary end points. We use this case study to illustrate the validation of CCyR and MMR as surrogate outcomes for overall survival in CML and how this evidence was used to inform National Institute for Health and Care Excellence’s recommendation on the public funding of these first-line treatments for CML. Methods: We undertook a systematic review and meta-analysis to quantify the association between CCyR and MMR at 12 months and overall survival in patients with chronic phase CML. We estimated life expectancy by extrapolating long-term survival from the weighted overall survival stratified according to the achievement of CCyR and MMR. Results: Five studies provided data on the observational association between CCyR or MMR and overall survival. Based on the pooled association between CCyR and MMR and overall survival, our modeling showed comparable predicted mean duration of survival (21–23 years) following first-line treatment with imatinib, dasatinib, or nilotinib. Conclusions: This case study illustrates the consideration of surrogate outcome evidence in health technology assessment. Although it is often recommended that the acceptance of surrogate outcomes be based on randomized controlled trial data demonstrating an association between the treatment effect on both the surrogate outcome and the final outcome, this case study shows that policymakers may be willing to accept a lower level of evidence (i.e., observational association).

Keywords: chronic myeloid leukemia, complete cytogenetic response, dasatinib, health technology assessment, HTA, imatinib, intermediate outcomes, major molecular response, nilotinib, surrogate end points, systematic review, technology appraisal.

Copyright © 2013, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.

Introduction

Chronic myeloid leukemia (CML) is a myeloproliferative neoplasm of hematopoietic stem cells [1]. CML used to be regarded as a progressive disease whose natural history evolves through three phases: the initial chronic phase, during which the disease is stable or at the most only slowly progressive, followed after a variable interval by transition through an accelerated phase to a rapidly fatal blast crisis [2–4]. Approximately 90% of the people affected by CML are diagnosed during the chronic phase, with a median age at diagnosis of around 65 years [2]. In the United States, about 4800 to 5200 new cases are diagnosed every year, which corresponds to an annual incidence of 1.0 to 1.3 per 100,000 population [5,6]. Similar annual age-standardized incidence rates have been published for the United Kingdom (i.e., 1.1 per 100,000 for men and 0.7 per 100,000 for women) [7].

Before the introduction of tyrosine kinase inhibitor (TKI) therapy, the median survival time after diagnosis was 6 years [6], with a predicted 5-year overall survival of 42.7% [8] and estimated prevalence of 25,000 to 30,000 cases in the United States [6] and of 4,000 to 5,000 cases in the United Kingdom [9].

The molecular pathogenesis of CML is well understood, and the disease presents the Philadelphia chromosome (Ph) as a molecular hallmark [10]. This fusion gene is the result of a reciprocal chromosomal translocation (i.e., t[9;22]), also known as breakpoint cluster region-Abelson (BCR-ABL) oncogene, that codes for BCR-ABL transcripts and fusion proteins with unusual tyrosine-kinase activity [11,12]. Diagnosis is confirmed by the identification of...
either the Ph or the BCR-ABL transcripts, in peripheral blood or bone marrow cells, through cytogenetic [11–14] analysis or reverse transcriptase polymerase chain reaction (PCR), which can be semi-quantitative (real-time PCR or quantitative PCR) [15]. Following recognition of the importance of achieving a certain depth of response at different time points for patients with newly diagnosed CML in chronic phase, the European LeukemiaNet has established guidelines on therapeutic milestones that should be achieved [15]. A complete cytogenetic response (CCyR) is defined as absence of the Ph among at least 20 cells in metaphase in a bone marrow aspirate [16], while a major molecular response (MMR) is reached if the standardized BCR-ABL:ABL ratio is less than 0.1%, which is equivalent to a 3 log reduction from the 100% baseline for untreated patients [17,18].

Knowledge of the molecular basis of this neoplastic disease has led to a new generation of drugs, the TKIs, radically changing the previous standard of care based on interferon-alpha (IFN-α) for patients with CML [13,14]. Imatinib, the first rationally developed selective TKI to be approved for the treatment of a cancer [19] by the European Medicines Agency in 2001, was rapidly adopted as first-line medical treatment for CML in chronic phase in the National Health Service in the United Kingdom [20]. The efficacy of imatinib in comparison with older treatments has been confirmed in a single randomized controlled trial (RCT), the International Randomized Study of Interferon and STI571 (IRIS) trial [24], a prospective, multicenter, open-label, phase 3 RCT comparing imatinib 400 mg/day with IFN-α plus cytarabine. In early 2012, two newer TKIs—dasatinib [17,21] and nilotinib [22–24]—initially promoted for the treatment of patients resistant or intolerant to imatinib [15,25], have been assessed by the National Institute for Health and Care Excellence (NICE) as alternative first-line treatments to imatinib in England and Wales [26]. The evidence of the relative effectiveness of these three alternative treatment options was based on two comparative RCTs, one that compared dasatinib with imatinib [21] (Dasatinib vs. Imatinib in Patients With Newly Diagnosed Chronic Phase CML, the DASISION trial) and the other comparing nilotinib with imatinib [24] (Evaluating Nilotinib Efficacy and Safety in clinical Trials—newly diagnosed patients, the ENESTnd trial). In both trials, the primary end points were biomarkers, that is, confirmed CCyR by 12 months in DASISION and MMR at 12 months in ENESTnd. Although average survival from diagnosis can reach 15 years [25] among this population, these two trials provide only immaturereg data on overall survival with a maximum follow-up of 2 years at the time of the assessment.

Central to this coverage decision, therefore, was consideration of CCyR and MMR as valid surrogate outcomes (i.e., biomarkers intended to substitute and predict for a final patient-relevant outcome [27]) for long-term overall survival in first-line TKI therapy for chronic phase CML to determine estimates of life-years gained across alternative treatments [28].

The dual aims of this study were 1) to assess the evidence base for the use of CCyR and MMR as surrogates for overall survival in patients with chronic phase CML treated with TKI (i.e., dasatinib, nilotinib, and imatinib) and 2) to describe how this evidence was used to predict long-term survival in the related cost-effectiveness model. The policy implications of the validation and use of surrogate outcomes in coverage decisions will be discussed.

Methods

This study consisted of two distinct methodological steps: 2) a systematic review and meta-analysis of the evidence base to quantify the association between CCyR and MMR as surrogates for overall survival in chronic phase patients with CML treated with first-line TKIs and 2) modeling of the observed CCyR and MMR at 12 months to predict long term (>12 months) patient survival in first- and second-generation TKI therapies.

Systematic Review and Meta-Analysis

Our systematic review and meta-analysis was conducted and reported in accord with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement [29].

Search strategy

We initially identified studies from a previous systematic review of imatinib for first-line treatment of CML in chronic phase [30]. The following bibliographic databases were searched: Medline, Embase and PsycINFO (all via OVID), The Cochrane Library, Web of Science (via ISI), and Econjournals (via CSA) from October 2002 (the end date of the previous systematic review [30]) to May 2012. The searches were limited to the English language. The Medline search strategy is reported in the Supplemental Material found at http://dx.doi.org/10.1016/j.jval.2013.07.004.

Inclusion criteria

Studies were included if they considered 1) adults (median age > 18 years) with chronic phase CML based on cytogenetic and/or fluorescence in situ hybridization and/or reverse transcriptase polymerase chain reaction results; 2) patients treated with dasatinib, nilotinib, or imatinib; 3) patients naïve to previous interferon or TKI treatment; and if they 4) reported the association between CCyR or MMR at 12 months and overall survival. We excluded studies published only as conference abstracts, narrative reviews, editorials, opinion pieces, and individual case studies, or studies whose findings were not judged to be generalizable to the CML population in the United Kingdom or Western countries. Where a study had been reported in several publications, we considered the article with the longest follow-up. Titles, abstracts, and full text of any potentially relevant studies were independently screened by two reviewers (O.C. and T.P.) with any discrepancies resolved by discussion, with the involvement of a third reviewer if necessary.

Data extraction and quality assessment

The methodological quality of included studies was assessed according to a modified list of criteria specified by the National Health Service Centre of Reviews and Dissemination [31]. Study characteristics and data were extracted by one reviewer (O.C.) by using a standardized data extraction form and independently checked by a second reviewer (T.P. or R.T.). To judge the reliability of CCyR and MMR at 12 months as surrogate measures for long-term overall survival, we referred to the following surrogate validation criteria: 1) evidence from RCTs demonstrating treatment effects on the surrogate correspond to treatment effects on the patient-relevant outcome, 2) evidence from observational studies demonstrating consistent association between surrogate outcome and final patient-relevant outcome, and 3) evidence of biological plausibility of relationship between the surrogate outcome and the final patient-relevant outcome [32].

Data analyses

For each study, overall survival was extracted at each year following trial recruitment (or randomization) up to the latest follow-up point reported, separately according to whether a CCyR or an MMR was achieved at 12 months. In all studies, overall survival data were estimated by the Kaplan-Meier method by using landmark analysis to evaluate differences in the final patient-relevant outcomes between responders and nonresponders.
The landmark method determines each patient’s response at a fixed time point, with survival calculated from that time point onward and associated statistical tests being conditional on patients’ landmark responses [33]. Data digitalization software (WinDIG Version 2.5) was used to extract data from Kaplan-Meier survival curves.

For consistency, we selected 12 months after the start of first-line therapy as the landmark for our analysis, as the DASISION [21] and ENESTnd [24] trials consider, respectively, the rate of confirmed CCyR and MMR at 12 months postrandomization as primary end points. Average overall survival rates at yearly intervals were estimated for both responders and nonresponders, weighted by the initial number of patients in the two groups for each trial. Wilson 95% confidence intervals (CIs) were derived for each point estimate assuming binomial distributed variables and no censoring of data [34]. Analyses were undertaken by using STATA v.11.2 (StataCorp, TX).

**Modeling and Data Extrapolation**

The above systematic review and meta-analysis provided a literature-based estimation of overall survival, according to whether patients with CML achieved either a CCyR and MMR response or not. A four-step analytical approach was then undertaken to estimate long-term overall survival separately for imatinib, dasatinib, and nilotinib treatment.

- **Step 1:** CCyR and MMR response rates at 12 months for dasatinib from ENESTnd [24] and for nilotinib from DASISION [21] were derived by using a WinBUGS mixed-treatment comparison analysis [35,36]. The appropriateness of the indirect comparison was assessed by checking that the baseline characteristics of the two trials were similar. A fixed-effects pairwise meta-analysis [37] was then undertaken to obtain an overall estimate of the proportion of patients achieving a CCyR and separately an MMR for each treatment [38].

- **Step 2:** Estimation of CML-related mortality from historical trial data [38]. Mortality was assumed to occur because of CML-related causes and non-CML causes. Given limited and immature historical data on CML-related death, the probability of CML-related death was assumed constant over time (as this was deemed most parsimonious), and to depend on whether a CCyR was achieved. In a separate exercise, the probability of CML-related death was assumed to depend on whether an MMR was achieved. Non-CML mortality was taken from UK life tables [39], and the age at diagnosis was estimated as the average age at diagnosis across all historical trials, weighted by the number of responders or nonresponders in each trial, as appropriate. The constant probability of CML-related death was estimated to minimize the sum of squares of differences between the actual historical overall survival and modeled overall survival at each year.

- **Step 3:** Estimation of overall survival separately for responders and nonresponders given a cohort of patients starting first-line treatment at age 57 years (i.e., the mean age at diagnosis in the United Kingdom) [7]. Overall survival was estimated by applying mortality from the general population with starting age 57 years and the appropriate estimate of CML-related mortality from step 2.

- **Step 4:** Estimation of overall survival for each treatment arm (i.e., imatinib, dasatinib, and nilotinib) by averaging the responder and nonresponder overall survival, estimated in step 3, weighted by the proportion of patients who did and did not achieve a response to first-line treatment at 12 months.

We compared our estimates of expected overall survival with the actual 24- and 36-month overall survival from the ENESTnd trial [40,41], the 24-month overall survival from the DASISION trial [42], and the longer term (i.e., 7 years) imatinib survival data from the IRIS trial [43]. In addition, two sensitivity analyses were performed on the pool of articles that contributed historical data by 1) excluding IRIS trial reports [43,44] and 2) including the unique dasatinib- and nilotinib-treated patient cohort identified in the Jabbour et al. study [45]. Given that the IRIS trial has been used for validation, the former was performed to test the influence of IRIS data on the estimated surrogate relationship, while the latter was performed to check whether the same relationship might be specific to type of TKIs.

Modeling analyses were carried out by using WinBUGS (MRC Biostatistics Unit, Cambridge, UK) and the Excel “Solver” function (Excel 2012 Microsoft Corporation, Redmond, WA).

**Results**

**Study Identification**

The process of study selection is summarized in Figure 1. Six publications met the inclusion criteria, reporting on five separate studies—two RCTs and three cohort studies (Table 1).

**Study and population characteristics**

We were able to include five studies in the quantitative analysis. One study, performed in India, was judged to be unlikely to reflect the clinical management of patients with CML and therefore excluded [49]. Only one arm in a cohort study reported patients with CML who were treated by dasatinib or nilotinib [45], with all the others considering imatinib treatment. We therefore decided to include only the imatinib treatment arm from the same study in our base-case analysis and to contrast the overall results with those reported for the dasatinib and nilotinib arm in Jabbour et al. [45]. As for the two RCTs, only the arms receiving standard-dose imatinib as first-line therapy were considered. This choice was taken because 1) the IRIS trial was inadequate to demonstrate a survival benefit for imatinib relative to IFN-α therapy in newly diagnosed Ph+ chronic-phase CML in the light of the high rate of crossover (65% at 72-month follow-up) from IFN-α plus cytarabine to imatinib [50], and 2) Hehlmann et al. [48] compared the 400 mg/day imatinib with the high-dose therapy (i.e., 800 mg/day) or combined therapy with interferon, which were not among the treatment options under comparison in our analysis.

**Study quality and hierarchy of surrogate evidence**

The included studies consistently showed moderate to good internal validity (see Table 1 in Supplemental Material found at http://dx.doi.org/10.1016/j.jval.2013.07.004) and were therefore all considered in the base-case analysis. In the two RCTs [43,44,48], the association between CCyR and MMR and overall survival was examined as a stratified comparison of overall survival in MMR and CCyR responders versus nonresponders for the imatinib 400 mg/day arm only. Thus, the level of surrogate outcome evidence identified by this review was entirely observational, that is, “level 2” evidence according to the three-level surrogacy evaluation scheme proposed by Elston and Taylor [32].

**Association between CCyR or MMR at 12 months and overall survival**

Table 2 shows the weighted pooled mean overall survival (95% CI) at yearly intervals, up to 7 years after the initiation of imatinib treatment, according to achievement of CCyR and MMR (or not) at 12 months [43–48]. For imatinib-treated patients with CML, the impact of achieving a CCyR at 12 months progressively translates into a
survival benefit compared with patients who do not achieve such response. The advantage of achieving an MMR at 12 months in terms of overall survival rates, however, is less clear. Overall survival rate estimates for imatinib-treated patients are slightly lower than the level of overall survival seen in the cohort of patients treated with nilotinib or dasatinib who achieve a CCyR at 12 months after treatment initiation in Jabbour et al. [45], although without reaching statistical significance (logrank test: \( P = 0.80 \)).

**Prediction of long-term overall survival**

The estimated long-term overall survival for responders and nonresponders compared with our literature-based synthesized historical data and survival for the general population in England and Wales is shown in Figure 2. After considering CCyR and MMR response rates by 12 months for each of the three treatment arms under comparison (see Table 2 in Supplemental Material found at http://dx.doi.org/10.1016/j.jval.2013.07.004), we were able to stratify long-term extrapolated overall survival by therapy (Fig. 3).

The extrapolated long-term overall survival curves for imatinib, dasatinib, and nilotinib treatment appear very comparable. The estimated mean overall survival across the three therapies (see Table 3 in Supplemental Material found at http://dx.doi.org/10.1016/j.jval.2013.07.004) ranges from 2.7 to 4.1 years less than the life expectancy of the general population of England and Wales. Taking into account the potential uncertainties in the estimation of trial-based response rates, according to our model, people treated with first-line imatinib are expected to survive 21.3 years on average, 1.4 years less than people treated with nilotinib and dasatinib, using the CCyR data. They are expected to survive 22.0 years, 0.6 years less than patients treated with nilotinib and dasatinib, using the MMR data. Results from the sensitivity analyses showed little or no impact on estimated mean overall survival when data from the IRIS trial were excluded [43,44] or when data from the dasatinib and nilotinib cohort [45] were included, thus supporting the consistency of the surrogate to final outcome relationship across the interventions.

We checked the accuracy of our estimates by comparing modeled overall survival and trial-based overall survival from the two RCTs of first-line dasatinib [21] and nilotinib [24] and from the imatinib arm of the IRIS trial [43]. It appears that the modeled overall survival is consistent with these data:

- at 2-year follow-up, dasatinib overall survival observed in DASISION was 95% compared with 97% estimated by our model [42];
- at 2-year follow-up, nilotinib overall survival observed in ENESTnd was 97% [40] compared with 97% in the model; at 3-year follow-up, the overall survival observed in the trial was 95% [41] compared with 95% estimated by the model;
- imatinib observed overall survival was 95% and 96% in DASISION and ENESTnd, respectively, compared with 97% in the model based on the CCyR surrogate relationship and 97% based on the MMR surrogate relationship [40,42].

In addition, the estimated overall survival for the imatinib arm closely predicts the actual overall survival in the imatinib arm of the IRIS RCT (see Fig. 1 in Supplemental Material found at http://dx.doi.org/10.1016/j.jval.2013.07.004). This is not a completely independent verification of overall survival through this method, because some of the data on overall survival for imatinib responders and nonresponders from the IRIS RCT were also used to estimate the overall survival surrogate relationships. Nonetheless, it serves as useful calibration of the model’s survival outputs because other historical data also heavily influenced the surrogate overall survival estimates.

**Discussion**

The molecular biology of CML supports the adoption of both CCyR and MMR as potential markers for monitoring of disease progression [13,51]. It has also been shown, however, that TKIs can have potential unexpected off-target effects (i.e., stem cell chromosomal instability, inhibition of proinflammatory functions [52-55]) that may call into question the ability of these
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
<td>UK</td>
<td>US</td>
<td>US</td>
<td>Germany</td>
<td>International</td>
<td>International</td>
</tr>
<tr>
<td>Study type</td>
<td>Cohort</td>
<td>Cohort</td>
<td>Cohort</td>
<td>RCT</td>
<td>RCT</td>
<td>RCT (retrospective comparison)</td>
</tr>
<tr>
<td>N patients</td>
<td>204</td>
<td>276</td>
<td>73</td>
<td>154</td>
<td>324</td>
<td>476</td>
</tr>
<tr>
<td>(TKI arm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median age (y)</td>
<td>46 (18–79)</td>
<td>48 (15–84)</td>
<td>48 (15–78)</td>
<td>47 (18–85)</td>
<td>54 (16–88)</td>
<td>50 (20–69)</td>
</tr>
<tr>
<td>(range)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention</td>
<td>Imatinib 400 mg/d</td>
<td>Imatinib 400 mg/d</td>
<td>Imatinib 400 mg/d</td>
<td>Dasatinib or nilotinib</td>
<td>Imatinib 400 mg/d</td>
<td>Imatinib 400 mg/d</td>
</tr>
<tr>
<td>Comparator</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>Imatinib 400 mg/d combined with IFN Imatinib 800 mg/d</td>
<td>IFN-α plus cytarabine</td>
</tr>
<tr>
<td>Median follow-up (mo)</td>
<td>38</td>
<td>48</td>
<td>110</td>
<td>28</td>
<td>43</td>
<td>77</td>
</tr>
<tr>
<td>Comparison</td>
<td>CCyR vs. no CCyR</td>
<td>CCyR vs. minor CyR †</td>
<td>CCyR vs. no CCyR</td>
<td>CCyR vs. no CCyR</td>
<td>MMR vs. no MMR</td>
<td>MMR vs. no MMR</td>
</tr>
<tr>
<td>between</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>responders vs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nonresponders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male sex, n (%)</td>
<td>116 (57)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>194 (60)</td>
<td>296 (62) (within 6 mo before study entry)</td>
</tr>
<tr>
<td>Median time</td>
<td>1.7</td>
<td>1.1</td>
<td>1</td>
<td>1</td>
<td>0.6</td>
<td>341 (62) (within 6 mo before study entry)</td>
</tr>
<tr>
<td>from diagnosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>to treatment (mo)</td>
<td>Low: 59 (29)</td>
<td>Low: 179 (65)</td>
<td>Low: 50 (68)</td>
<td>Low: 118 (77)</td>
<td>EuroSCORE Low: 113 (35)</td>
<td>Low: 175 (37)</td>
</tr>
<tr>
<td>Sokal risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>Ph+ CML in CP</td>
<td>Ph+ CML in CP</td>
<td>CML in CP</td>
<td>CML in CP</td>
<td>CML in CP</td>
<td>Ph+ CML in CP</td>
</tr>
</tbody>
</table>

Ccyr, complete cytogenetic response, CP, chronic phase; EuroSCORE, European System for Cardiac Operative Risk Evaluation; IFN-α, interferon-alpha; IRIS, International Randomized Study of Interferon and STI571; MMR, major molecular response; Ph+, Philadelphia chromosome positive; RCT, randomized controlled trial; TKI, tyrosine kinase inhibitor.

* Jabbour et al. [45] studied three cohorts of patients treated, respectively, with imatinib 400 mg daily, imatinib 800 mg daily, and second-generation TKIs (i.e., nilotinib and dasatinib). We excluded the imatinib 800 mg daily arm for the purpose of this analysis as nonlicensed dose. Data from imatinib 400 mg daily cohort were included in our base-case analysis; data from the second-generation TKIs cohort were used in a sensitivity analysis.

† The group of people achieving a minor cytogenetic response at 12 mo after the first-line treatment initiation (N = 5) in Kantarjian et al. [47] study report was excluded from the pooled overall survival average estimate.

‡ Sokal score is a prognostic classification system for patients with chronic myeloid leukemia that is designed to identify patients with low, intermediate, or high risk of poor outcome.

§ More than 85% of the patients are Ph+.
markers to fully capture the efficacy and safety pathways of the TKI therapy. In this article, we assessed the empirical evidence for the use of CCyR and MMR at 12 months as surrogate outcomes for overall survival in patients with chronic phase CML receiving first-line TKI treatment.

The results confirm the current adoption of CCyR at 12 months as a gold standard for a good response, whereas MMR provides a measure of success rather than a measure of failure (i.e., not achieving high levels of molecular response does not constitute treatment failure in patients with CML) [25]. Our systematic review identified three cohort studies [45–47] and two RCTs [43,44,48] examining the association between these two biomarkers and overall survival in patients with chronic phase CML receiving first-line imatinib. While these studies showed a consistent association between CCyR and MMR and long-term (i.e., 1–7 years) overall survival, this was based on observational analyses comparing responders versus nonresponders. Based on the pooled observational association between CCyR and MMR and overall survival, our modeling showed comparable predicted mean duration of survival (21–23 years) following first-line treatment with imatinib, dasatinib, or nilotinib.

Although a plausible biological rationale constitutes a basic step toward the identification of a surrogate outcome [32,56,57], it is not sufficient by itself to prove that the treatment effect on the substitute end point may predict the treatment effect on the final patient-relevant outcome. Empirical evidence of an association between these end points and final patient-relevant outcome, as well as between treatment effects on them, is also needed. Ideally, this evidence should be in the form of multiple RCTs that assess the effects of the treatment on both the end point marker and final patient-relevant outcome at relevant follow-up time [58].

Findings of Previous Studies
To our knowledge, this is the first systematic review and meta-analysis to examine the scientific basis of the validation of CCyR and MMR as surrogates for long-term overall survival in patients with chronic phase CML treated with TKIs. Two previous studies have examined this question in patients treated with IFN-α.

Anstrom et al. [59] fitted a proportional hazards model to estimate long-term survival by conditioning on CCyR at 2 years landmark time point. Their data sources were the IRIS trial at 19-month follow-up [14] and four clinical trials [60–63] assessing patients treated with IFN-α plus low-dose cytarabine. They predicted a residual life expectancy after CCyR at 2 years of 16.7 years and of 5.8 years for non-CCyR cohorts. In comparison, our estimates were 24.5 and 14.3 years, respectively. The life expectancy estimates of Anstrom et al. [59] for first-line imatinib were 15.3 years compared with our estimate of 21.3 years using CCyR as a surrogate. The differences in the estimates can be explained first by the different landmark time considered for the CCyR (i.e., 1 year vs. 2 years), second by the choice of the baseline survival functions (i.e., life-table estimates for the general U.S. population weighted according to baseline age and sex distributions of the IRIS population vs. UK life-table estimates weighted by the number of responders or nonresponders in each trial, as appropriate), and third by the assumption that long-term survival estimates for imatinib-treated patients are similar to those derived from two cohorts of IFN-α-based regimens with [61] or without [60] CCyR at 2 years.

Schrover et al. [64] assumed that prolonged survival after attaining a major cytogenetic response (i.e., 0%–35% Ph+ cells among at least 20 cells in metaphase in a bone marrow aspirate) may be independent of treatment and developed a survival model for patients with chronic phase CML using a logistic regression to predict survival according to major cytogenetic response rate. They estimated an average difference in survival between responders and nonresponders at 2 and 4 years after landmark of 15.0% and 25.8%, respectively, and predicted a proportion of patients alive at 5 years of 70%. Our model predicts more favorable outcomes for the patients (i.e., fitted 5-year survival in patients without CCyR is 77%, 98% in patients with CCyR); however, our results cannot be easily compared with those reported by Schrover et al. [64], given that they considered a different surrogate end point and data derived from IFN-α-based regimens.

Recently, a systematic review and meta-analysis has evaluated the efficacy and safety of second-generation TKIs (including bosutinib) versus imatinib [65]. The inclusion criteria for this study were slightly different from those in our systematic review: Gurion et al. [65] considered only RCTs, also published as conference abstracts, with no restriction on adult population. Although the objective was not that of validating CCyR and MMR at 12 months as surrogates for overall survival, they observed no statistically significant difference between first- and second-generation TKIs groups in all-cause mortality rates at 12 months (relative risk [RR] 0.76; 95% CI 0.42–1.37) despite a general improvement in the CCyR rate at 12 months (RR 1.16; 95% CI 1.09–1.23) and MMR at 12 months (RR 1.68; 95% CI 1.48–1.91) in patients allocated to the second-generation TKIs arm as compared to patients allocated to the imatinib arm.

Strengths and Limitations
In a situation in which only evidence about clinical effectiveness measured in terms of cytogenetic or molecular response and

<table>
<thead>
<tr>
<th>Time (mo)</th>
<th>CCyR</th>
<th>No CCyR</th>
<th>OS % (95% CI)†</th>
<th>P†</th>
<th>MMR</th>
<th>No MMR</th>
<th>OS % (95% CI)†</th>
<th>P†</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>100</td>
<td>100</td>
<td>100 (99.4–100)</td>
<td>0.15</td>
<td>100</td>
<td>100</td>
<td>100 (99.4–100)</td>
<td>1.00</td>
</tr>
<tr>
<td>24</td>
<td>98.3</td>
<td>94</td>
<td>94 (89.7–96.5)</td>
<td>0.03</td>
<td>99.2</td>
<td>98.0</td>
<td>98.0 (99.8–99.5)</td>
<td>0.42</td>
</tr>
<tr>
<td>36</td>
<td>97.7</td>
<td>88.9</td>
<td>88.9 (83.9–92.2)</td>
<td>0.02</td>
<td>96.7</td>
<td>95.0</td>
<td>95.0 (97.0–97.9)</td>
<td>0.35</td>
</tr>
<tr>
<td>48</td>
<td>98.3</td>
<td>No data</td>
<td>Not calculable</td>
<td>96.2</td>
<td>94.4</td>
<td>94.4</td>
<td>94.4 (97.9)</td>
<td>0.20</td>
</tr>
<tr>
<td>60</td>
<td>97.4</td>
<td>73.6</td>
<td>73.6 (62.4–82.4)</td>
<td>0.02</td>
<td>96.9</td>
<td>95.1</td>
<td>95.1 (98.1)</td>
<td>0.11</td>
</tr>
<tr>
<td>72</td>
<td>98</td>
<td>No data</td>
<td>Not calculable</td>
<td>96.0</td>
<td>93.2</td>
<td>93.2</td>
<td>93.2 (97.5)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>84</td>
<td>94</td>
<td>92.5</td>
<td>92.5 (87.6–95.9)</td>
<td>89.2</td>
<td>83.5</td>
<td>83.5</td>
<td>83.5 (93.4)</td>
<td>NS</td>
</tr>
</tbody>
</table>

CCyR, complete cytogenetic response; CI, confidence interval; MMR, major molecular response; NS, nonsignificant; OS, overall survival.

† P value for difference in the OS rate between responders and nonresponders. Where more than one studies are available, P values derived from t-statistic meta-regression analysis. Where only one study is available, P values derived from the study publication.
immature data about overall survival were available, we systematically looked for evidence supporting the adoption of both CCyR and MMR at 12 months as reliable predictors of overall survival by looking at TKIs-treated patients data, naive to previous pharmacological therapies for CML. In fact, although initial marketing-authorization for imatinib was granted on the primary efficacy end point of the proportion of patients achieving major cytogenetic response, based on the strong association between cytogenetic response and survival observed with IFN-α, the accompanied scientific discussion document specified that the mechanisms of action of imatinib and IFN-α were different and the association between survival improvement and achievement of cytogenetic response needed to be confirmed for imatinib and following TKI drugs [66]. Nonetheless, we acknowledge some potential limitations in our analysis. First, we were able to examine the validity of the two biomarkers as surrogate end points only on the basis of aggregate data. Access to individual patient data has the advantage of standardizing the methods of statistical analysis not only across but also within studies [67]. Second, we relied on a landmark analysis, which

Fig. 2 – Observed vs. fitted overall survival for patients (a) with and without a CCyR (upper panel) and (b) with and without a MMR (lower panel) at 12 months. CCyR, complete cytogenetic response at 12 months; MMR, major molecular response at 12 months. Observed overall survival rates derived from Table 2.

Dasatinib was either dominated by nilotinib (i.e., dasatinib was less effective and more costly) or in comparison to imatinib was judged to have a cost per QALY in excess of €200,000. Based on NICE’s recommendation, it is estimated that about 509 new patients each year in the United Kingdom will receive first-line therapy with imatinib and nilotinib [26].

The use of surrogate validation evidence in this case by NICE has policy implications for cancer therapy beyond the management of CML. With pressure for faster patient access to innovative therapies, surrogate or intermediate outcomes (such as tumor response, event-free survival) are increasingly being used as primary outcomes in licensing trials of new cancer therapies. Policymakers are also facing reimbursement [28,69] decisions on these new treatments on the basis of evidence on impact on intermediate outcomes with little or no definitive data on the impact of therapy on overall survival. In response, health technology assessment groups and national or regional agencies responsible for drug coverage are beginning to introduce access restrictions on the basis of surrogate outcome data [70,71]. Basically, new guidelines for technology appraisals state that it is no longer sufficient for new therapies to claim effectiveness on surrogates accepted on the basis of the biological plausibility from pathophysiological studies or the understanding of the disease process. Instead, such claims need to be grounded on “validated” surrogate outcomes, that is, outcomes that have proven association and predictive capacity to the final patient-relevant outcome.

Although the highest level of evidence should come from a meta-analysis of RCTs demonstrating consistent association between treatment effects on the surrogate outcomes and treatment effects on the patient-relevant outcomes, the recent NICE evaluation of first-line therapies for chronic phase CML demonstrates not only that evidence of surrogate outcome validation can be central to a positive listing but also that observational-level evidence may suffice for the new drug to be included in public formularies.

Ongoing and future RCTs of TKI first-line therapies in chronic stage CML should continue to follow up patients to provide the necessary data to examine the strength and consistency of the relationship between treatment changes in CCyR and MMR and long-term overall survival.

**Fig. 3** Overall survival for each treatment arm estimated by surrogate relationship based on CCyR and MMR separately. CCyR, complete cytogenetic response at 12 months; MMR, major molecular response at 12 months.

**Implications for Policy and Practice**

In March 2012, NICE issued guidance recommending nilotinib and standard-dose imatinib for the first-line treatment of Ph+ CML [26]. Although the Appraisal Committee concluded that dasatinib and nilotinib provided superior clinical benefit over standard-dose imatinib, as measured by surrogate outcomes, the cost per quality-adjusted life-year (QALY) gained for nilotinib compared with standard-dose imatinib was estimated to fall below NICE’s willingness-to-pay threshold of €20,000 to €30,000/QALY, while dasatinib was either dominated by nilotinib (i.e., dasatinib was less effective and more costly) or in comparison to imatinib was judged to have a cost per QALY in excess of €200,000. Based on NICE’s recommendation, it is estimated that about 509 new patients each year in the United Kingdom will receive first-line therapy with imatinib and nilotinib [26].

The use of surrogate validation evidence in this case by NICE has policy implications for cancer therapy beyond the management of CML. With pressure for faster patient access to innovative therapies, surrogate or intermediate outcomes (such as tumor response, event-free survival) are increasingly being used as primary outcomes in licensing trials of new cancer therapies. Policymakers are also facing reimbursement [28,69] decisions on these new treatments on the basis of evidence on impact on intermediate outcomes with little or no definitive data on the impact of therapy on overall survival. In response, health technology assessment groups and national or regional agencies responsible for drug coverage are beginning to introduce access restrictions on the basis of surrogate outcome data [70,71]. Basically, new guidelines for technology appraisals state that it is no longer sufficient for new therapies to claim effectiveness on surrogates accepted on the basis of the biological plausibility from pathophysiological studies or the understanding of the disease process. Instead, such claims need to be grounded on “validated” surrogate outcomes, that is, outcomes that have proven association and predictive capacity to the final patient-relevant outcome.

Although the highest level of evidence should come from a meta-analysis of RCTs demonstrating consistent association between treatment effects on the surrogate outcomes and treatment effects on the patient-relevant outcomes, the recent NICE evaluation of first-line therapies for chronic phase CML demonstrates not only that evidence of surrogate outcome validation can be central to a positive listing but also that observational-level evidence may suffice for the new drug to be included in public formularies.

Ongoing and future RCTs of TKI first-line therapies in chronic stage CML should continue to follow up patients to provide the necessary data to examine the strength and consistency of the relationship between treatment changes in CCyR and MMR and long-term overall survival.
Acknowledgments

Source of financial support: Ciani Oriana is currently receiving a Peninsula College of Medicine and Dentistry PhD studentship. This work was supported by the National Institute for Health Research (NIHR) health technology assessment program (project number 08/226/01). The authors thank Rob Anderson and Obioha Ukoumunne for their valuable contribution to the study. R.G. is partially supported by the NIHR Collaboration for Leadership in Applied Health Research and Care (CLAHRC) for the South West Peninsula and by the European Regional Development Fund and the European Social Fund Convergence Programme for Cornwall and the Isles of Scilly. The views expressed in this publication are those of the authors and not necessarily those of the National Health Service, the NIHR, the Department of Health in England, or the European Union.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at http://dx.doi.org/10.1016/j.jval.2013.07.004 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES
