Indirect comparisons of second-generation tyrosine kinase inhibitors in CML: case study using baseline population characteristics

Kimbach Tran Carpiuc
Gianantonio Rosti
Fausto Castagnetti
Maarten Treur
Jennifer Stephens

1Pharmerit North America LLC, Bethesda, MD, USA; 2Department of Hematology and Oncology, S Orsola-Malpighi University Hospital, Bologna, Italy; 3Pharmerit Europe, Rotterdam, The Netherlands

Abstract: The use of indirect comparisons to evaluate the relative effectiveness between two or more treatments is widespread in the literature and continues to grow each year. Appropriate methodologies will be essential for integrating data from various published clinical trials into a systematic framework as part of the increasing emphasis on comparative effectiveness research. This article provides a case study example for clinicians using the baseline study population characteristics and response rates of the tyrosine kinase inhibitors in imatinib-resistant or imatinib-intolerant chronic myelogenous leukemia followed by a discussion of indirect comparison methods that are being increasingly implemented to address challenges with these types of comparisons.

Keywords: comparative effectiveness research, meta-analysis, BCR-ABL-positive chronic myelogenous leukemia, imatinib mesylate, nilotinib, dasatinib

Introduction

Health care providers and policy makers routinely rely on available evidence to make informed clinical and medical decisions on treatment interventions, whether on clinical effectiveness, safety, cost-effectiveness, or quality of life outcomes.

Where existing practice guidelines do not yet exist, data from randomized control trials (RCTs) provide the most reliable source for examining the relative effectiveness of different interventions. Well-conducted head-to-head RCTs of available treatments are generally accepted as the most valid evidence in comparing relative outcomes for competing interventions. Moreover, when large amounts of published and unpublished data exist, clinicians and health care policy makers often rely on systematic reviews or meta-analyses of different treatment options, where large amounts of published and unpublished data are synthesized from different clinical studies.

However, pairwise (or more) clinical trials of all eligible treatments for a particular indication are not always available or feasible. In many cases, there may be an indication where several placebo-controlled RCTs for the various treatments are available, however no studies in which the treatments have been compared directly with each other. Even if direct comparison studies have been conducted, the data may be of poor quality and/or insufficient.

Where direct comparative evidence is unavailable or limited, indirect comparisons using the available data from well-designed and conducted studies using common comparators, such as a placebo, no treatment, or other existing interventions, are recommended and may be very useful. While there is still some concern that indirect comparisons may be subject to more methodologic bias than direct comparisons,1,2
awareness and understanding of the underlying assumptions will help support the usefulness of indirect comparisons for evaluating competing health care interventions.3,4

Indirect comparisons
Comparisons of treatments not from head-to-head studies may come in several different forms; in a simple example, trying to compare evidence from separate placebo-controlled RCTs for treatment A and treatment B. In this scenario, a comparison between treatment A and treatment B is estimated relative to the effect of the common comparator, C (ie, the placebo-controlled group). In order to minimize bias, the baseline study populations in the respective studies should ideally be similar, as should the study design and methodology, follow-up, and statistical analyses, although adjustment to control for any potential differences is possible. In more complex situations, known as mixed or multiple treatment comparisons, also known as network meta-analysis,5 evidence for more than two treatments are compared. Multiple treatment comparisons may also refer to circumstances where direct and indirect evidence are combined to strengthen the inference regarding the relative treatment effects of two or more interventions.6 Whether comparisons are made for greater than two interventions, or use both indirect and direct comparisons,6,7 awareness of assumptions for examining comparisons across multiple studies is critical.

The use of indirect comparisons to evaluate the relative effectiveness between two or more treatments is widespread in the literature and continues to grow each year. While classic frequentist methods are still the most common approach used for indirect methods,4 there is an increasing popularity and use of approaches such as Bayesian meta-analyses, as seen in workshops and courses at scientific meetings, as well as in the number of publications in high-tier journals.8

Case study
Background
Imatinib mesylate (Gleevec®/Glivec®; Novartis Pharmaceuticals, Basel, Switzerland) was developed as the first available inhibitor with targeted activity against the constitutively active tyrosine kinase of the BCR–ABL chimeric fusion protein present in patients with chronic myelogenous leukemia (CML), and has become the current standard of care for CML due to remarkable long-term activity and a mild toxicity profile.9,10

While imatinib is an effective, frontline treatment for patients with chronic phase CML, some patients are unable to respond to, become resistant to, or are unable to tolerate some of the side effects of imatinib, leaving few treatment options available. Second-generation tyrosine kinase inhibitors with greater potency and less susceptibility to mutation-induced mechanisms of resistance have led the way in therapeutic approaches for patients who cannot be treated with imatinib.

Nilotinib (Tasigna®; Novartis Pharmaceuticals and dasatinib (Sprycel®; Bristol-Myers Squibb, Princeton, NJ) were developed as novel targeted therapies to address the challenges associated with developed resistance in patients treated with imatinib. In clinical trials, treatment with nilotinib or dasatinib has achieved strong hematologic and cytogenetic response in patients with CML with primary or secondary imatinib resistance or imatinib intolerance. Currently, no head-to-head studies are available comparing nilotinib with dasatinib, although data from single-arm studies suggest differences in both efficacy and safety between the two treatments.

Clinical design differences for tyrosine kinase inhibitors
While both nilotinib and dasatinib have been shown to be highly effective in chronic phase CML patients with imatinib resistance or intolerance, the pivotal investigations of these agents11,12 were nonrandomized Phase II studies, making direct comparisons difficult. Moreover, the patient populations and protocols for each study were different, further complicating both efficacy and safety comparisons.

Defining imatinib resistance
The approval of dasatinib 70 mg twice daily was based on the SRC/ABL Tyrosine Kinase Inhibition Activity Research Trial (START-C) Phase II registration study.12 However, more recently, a 100 mg once-daily dosing schedule was approved for chronic phase CML patients based on results from a Phase III dose and schedule optimization (2 × 2) study.13 Nilotinib was approved based on results of the 2101 trial.11 These studies used similar definitions of imatinib resistance that correspond with the European LeukemiaNet recommendations regarding failure to achieve hematologic and cytogenetic milestones or loss of responses,14 with imatinib resistance defined as no complete hematologic response by three months; no minor cytogenetic response by six months; no major cytogenetic response by 12 months; and a loss of complete hematologic response or major cytogenetic response at any time. However, in the dasatinib Phase III dose optimization study, patients who tolerated imatinib 400 mg/day but who did not achieve a complete cytogenetic
response after 12 months and were unable to tolerate imatinib dose escalation to 600 mg/day were still considered to be imatinib-resistant. Patients in a partial cytogenetic response or major cytogenetic response were thus eligible for the dasatinib dose optimization study. These patients were included in the imatinib-resistant population, despite the fact that many patients (20% in the 100 mg once-daily arm including intolerant patients) entered the study with a baseline major cytogenetic response.

There were also differences in the dose and requirements for duration of prior imatinib therapy between the trials. In the nilotinib trial, imatinib-resistant patients were required to have received dose-escalated imatinib therapy with ≥600 mg/day for at least three months prior to trial enrollment. In contrast, there were no imatinib dose-escalation requirements for enrollment in the dasatinib studies. Of chronic phase CML patients enrolled in the 2101 study, 72% received imatinib doses ≥600 mg/day (including 38% who received >800 mg/day) as the highest prior imatinib dose. For patients in the dasatinib study, 72% of patients enrolled reported >600 mg/day prior imatinib use, and 37% of patients in the dose optimization 2 × 2 study reported ≥800 mg/day imatinib prior to study enrollment. While the criteria between the nilotinib and dasatinib trials appear to be balanced with regards to imatinib resistance for enrollment, defining highest prior dose and duration of imatinib among the studies indicate differences that make comparisons challenging. Furthermore, the eligibility of patients with suboptimal responses to imatinib in the dasatinib 2 × 2 study also make it very difficult to compare efficacy data with the 2101 study data.

Defining imatinib intolerance

In the dasatinib START-C trial, imatinib intolerance was defined as at least Grade 3 nonhematologic toxicity or Grade 4 hematologic toxicity persisting for >7 days related to imatinib. However, in the dasatinib 2 × 2 study, intolerance was defined as Grade 3 or worse toxicity that led to discontinuation of imatinib therapy. Imatinib intolerance was defined more stringently in the nilotinib trial, i.e., patients without an major cytogenetic response and who discontinued for persistent Grade 3/4 adverse events despite optional supportive care, or Grade 2 adverse events related to imatinib despite optimal supportive care persisting for ≥one month, or recurring > three times with dose reduction or discontinuation. Importantly, imatinib-intolerant patients achieving a major cytogenetic response following imatinib therapy were not eligible for participation in the nilotinib study. Overall, 84% of imatinib-intolerant patients enrolled into the nilotinib study did not have a major cytogenetic response. Of the imatinib-intolerant patients entering the study with a major cytogenetic response, a partial or complete cytogenetic response was observed in 40% of patients. These eligibility differences resulted in many imatinib-intolerant patients with pre-existing major cytogenetic response at study entry, i.e., 44% of the imatinib-intolerant patients enrolled in the dasatinib START-C study and 20% of the patients (including some resistant patients) in the dasatinib dose optimization 2 × 2 study. Additionally, 51% of patients treated with dasatinib in the dose optimization study had a baseline complete hematologic response upon study entry, compared with 36% of nilotinib-treated patients in the 2101 study. Data from recent analyses suggest that patients with a complete hematologic response have a greater chance of achieving a cytogenetic response.

Clinical case study with response rates

Because there are currently no head-to-head studies between the two available second generation tyrosine kinase inhibitors, clinicians and medical decision-makers looking to examine the efficacy evidence from individual studies of nilotinib and dasatinib must look at individual studies.

For nilotinib, the key evidence comes from a paper published in 2007 by Kantarjian et al. This paper followed 280 patients with chronic phase CML with a median treatment duration of 245 days (eight months).

While initial approval indicated 70 mg twice-daily dosing for chronic phase CML based on the dasatinib START-C study, in late 2007 new product labeling changed the dosing to 100 mg once daily for chronic phase CML based on results from the Phase III Dose and Schedule Optimization (2 × 2) study reported by Shah et al. It is therefore important to examine the evidence of the currently recommended dosing for dasatinib from this 2 × 2 study rather than that from the previous pivotal study. Median duration of treatment was eight months.

In the following example, we examine the challenges of comparing efficacy evidence from these studies for treating imatinib-resistant and imatinib-intolerant patients with chronic phase CML.

The percentage of patients achieving a major cytogenetic response, which was the primary efficacy endpoint in both studies with a minimum follow-up of six months, is presented in the Table 1. In the pivotal nilotinib study, a major cytogenetic response was achieved in 50% of imatinib-resistant or imatinib-intolerant patients with chronic phase CML. By
comparison, those treated with dasatinib 100 mg once daily, the current recommendation for chronic phase CML patients, 59% of patients achieved a major cytogenetic response. Thus, dasatinib 100 mg once daily may appear more effective than nilotinib 400 mg twice daily by simply looking at these efficacy rates without regard to differences in study design or baseline patient clinical characteristics.

Comparison of response and progression rates over time between nilotinib and dasatinib must be considered in the context of the study inclusion criteria and the baseline characteristics of the patient populations as previously mentioned. Of particular note, patients with chronic phase CML treated with nilotinib (compared with the dasatinib 100 mg once-daily arm in the $2 \times 2$ dose and schedule optimization study) were more heavily pretreated with high-dose imatinib and interferon, less likely to have a baseline complete hematologic response or major cytogenetic response, more likely to have baseline mutations, and more likely to have p-loop mutations.\textsuperscript{11,13}

A quick adjustment of simply removing the patients that already were in a major cytogenetic response at the start of each respective study indicates substantial changes in response rates between the two treatments. Even not taking into account any other discrepancies between study and clinical characteristics in disease severity and pretreatment activity between patients treated with nilotinib and dasatinib, correcting only for baseline major cytogenetic response status demonstrates how important homogeneity of study design and patient characteristics are in comparing efficacy rates between different treatments. Reflecting those with new responses, dasatinib-treated CML patients (100 mg once daily) thus had only 38% achieving a new major cytogenetic response compared with the 46% of nilotinib-treated patients (see Table 1).

While similarity in patient demographics is ideal for comparing different treatments, it becomes even more important when certain baseline characteristics modify the outcome. For example, it has been shown that baseline mutation status influences the response rates for efficacy in clinical trials of both nilotinib and dasatinib. Patients harboring baseline mutations had major cytogenetic response rates that were 15%–50% lower.\textsuperscript{13,18-20} Additionally, those with baseline mutation rates had a 1.5-fold increased risk of disease progression.\textsuperscript{18,20} Specifically, those with p-loop mutations (ie, those falling between amino acids 248–256), are significant and may be associated with lower response rates and a faster rate of disease progression (eg, E255K/V).\textsuperscript{11,20}

### Discussion of methods and Bayesian meta-analysis

The example presented above is given to emphasize the importance of trial similarity prior to conducting indirect comparisons. While the information shown was primarily for informal discussion, the actual data from the respective nilotinib and dasatinib studies would need to be analyzed using more complex methodology. As the studies described do not have a common comparator (ie, the nilotinib study was a single-arm open-label study, whereas the dasatinib study was a $2 \times 2$ factorial design with four open-label dasatinib treatment arms), and there is a significant difference in clinical characteristics between the patients in each respective study, only a naïve indirect comparison was possible; simple adjusted methods for indirect comparison would not be appropriate. While existing second-line studies provide only an opportunity for naïve indirect comparisons, new studies for first-line treatment using second-generation TKIs in newly diagnosed CML have imatinib as a common comparator and would thus allow for adjusted indirect comparisons.

Adjusted indirect comparisons can preserve the strength of randomization because individual studies are considered the units of analysis, and not treated as if they had come from one large trial.\textsuperscript{21,22} They may contain fewer biases than direct head-to-head studies, by counterbalancing biases in

### Table 1

<table>
<thead>
<tr>
<th>Correction for baseline characteristics</th>
<th>Nilotinib 400 mg bid (2101)\textsuperscript{11}</th>
<th>Dasatinib 100 mg qd (2 $\times$ 2)\textsuperscript{13}</th>
<th>Dasatinib 50 mg bid (2 $\times$ 2)\textsuperscript{13}</th>
<th>Dasatinib 140 mg qd (2 $\times$ 2)\textsuperscript{13}</th>
<th>Dasatinib 70 mg bid (2 $\times$ 2)\textsuperscript{13}</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 291</td>
<td>n = 167</td>
<td>n = 168</td>
<td>n = 167</td>
<td>n = 167</td>
<td>n = 168</td>
</tr>
<tr>
<td>MCyR at study entry</td>
<td>11 (4%)\textsuperscript{11,13}</td>
<td>34 (20%)</td>
<td>23 (14%)</td>
<td>28 (17%)</td>
<td>31 (18%)</td>
</tr>
<tr>
<td>MCyR at six months reported in publications\textsuperscript{11,13}</td>
<td>145 (50%)</td>
<td>98 (59%)</td>
<td>90 (54%)</td>
<td>93 (56%)</td>
<td>93 (55%)</td>
</tr>
<tr>
<td>Adjusted MCyR to reflect new responders</td>
<td>134 (46%)</td>
<td>64 (38%)</td>
<td>67 (40%)</td>
<td>65 (39%)</td>
<td>62 (37%)</td>
</tr>
</tbody>
</table>

**Notes:** \textsuperscript{11} As noted in the nilotinib 2101 study. \textsuperscript{13} In addition, five patients entered the study with a complete cytogenetic response and maintained their response in the study; another three patients entered the study in partial cytogenetic response and also maintained their response in the study; and three patients had missing baseline assessment but achieved complete cytogenetic response during the study. Therefore, another 11 (4%) patients had documentation of major cytogenetic response during the study.

**Abbreviations:** qd, once daily; bid, twice daily; MCyR, major cytogenetic response.
clinical trials if the sets of trials are similarly biased.\textsuperscript{22} Patient baseline clinical characteristics should be similar across the different treatment interventions as should the protocols of the different trials. However, in reality, there are usually significant differences in study and patient characteristics, which may or may not be assessed and adjusted for in regression models.

Several methodologic challenges in the use of indirect comparisons have recently been highlighted by Song et al in their survey of the published systematic reviewed literature.\textsuperscript{4} Most of these problems surround issues related to assumptions of homogeneity underlying the comparisons. The validity of adjusted indirect comparisons weighs heavily on the internal validity of each study, as well as the similarity of the studies with regard to patient population, methodology, and analysis.\textsuperscript{2}

Bayesian meta-analysis is one approach available for indirect comparisons that can accommodate multiple common comparators and/or require less restrictive study assumptions and scenarios. Bayesian statistics are increasingly being used in the research community as an important method for combining evidence. Rather than focusing on a single comparison of two interventions, these methods include and treat all included interventions equally.\textsuperscript{4} Bayesian methods in meta-analysis are useful when homogeneity between various studies does not exist, and offer the greatest flexibility in models evaluating evidence from indirect comparisons, particularly in random effects models that can account for the heterogeneity of between-trial variations, and lends itself to a decision framework that supports medical decision-making.\textsuperscript{23–26} Bayesian methods have been applied across a variety of therapeutic and disease areas, including methicillin-resistant \emph{Staphylococcus aureus} skin infections, hemophilia, malaria, colony stimulating factors, drug-eluting stents, hyperlipidemia, and breast cancer.\textsuperscript{27–33} Moreover, Bayesian statistics allow pooling of information from all relevant studies, whether comparative or single-arm studies, while adjusting multiple variables and study differences. In CML studies, using Bayesian meta-analysis methods may minimize bias and provide a meaningful comparison of treatment options, while also allowing input on a variety of inclusion and external information, including prior knowledge of Sokal risk, baseline mutation status, and other variables that might help predict treatment success or failure in patients with CML.

Bayesian methods for evaluating indirect comparisons require more complex statistical techniques, although recent developments in software, such as WINBUGS (Bayesian inference Using Gibbs Sampling), have made for easier implementation of Bayesian methods.\textsuperscript{34} However, expert statistical understanding is still required to appropriately use and interpret the data.

**Conclusion**

In the context of CML, application of indirect comparison methods would be useful and timely. The natural history of the disease has been radically changed over the past decade and will continue to evolve as newer targeted therapies show evidence of effectiveness. As new interventions are developed, the number of clinical trials to measure the effectiveness and safety of these treatments will also continue to grow. Clinicians and policy makers need ways to synthesize the increasing amount of evidence generated from such studies, especially where competing interventions exist. In the absence of head-to-head RCTs of the newer tyrosine kinase inhibitors, indirect comparisons, especially advanced techniques such as Bayesian meta-analysis, are valuable methodologic tools that can provide health care decision makers with useful information on the comparative effectiveness of CML therapies.

**Disclosure**

Funding for this manuscript was provided by Novartis Pharmaceuticals.

**References**


