Has the time for first-line treatment with second generation tyrosine kinase inhibitors in patients with chronic myelogenous leukemia already come? Systematic review and meta-analysis

Ronit Gurion,1,4* Anat Gafter-Gvili,1,4* Liat Vidal,1,4 Avi Leader,2 Ron Ram,1,4 Adi Shacham-Abulafia,1,4 Mical Paul,1,4 Isaac Ben-Bassat,4 Ofer Shpilberg,1,4 and Pia Raanani1,4

1Institute of Hematology, Davidoff Cancer Center, Rabin Medical Center; 2Internal Medicine Department A, Meir Hospital; 3Infectious Disease Unit, Rabin Medical Center; and 4Sackler School of Medicine, Tel Aviv University, Israel

© 2013 Ferrata Storti Foundation. This is an open-access paper. doi:10.3324/haematol.2012.063172

ONLINE SUPPLEMENTARY APPENDIX

Search terms used for the systematic review


Sokal and Hasford (EURO) prognostic scoring systems for newly diagnosed chronic myeloid leukemia

Various scoring systems have been developed in an attempt to predict disease outcome in CML patients.

(i) The Sokal prognostic score identified four clinical variables: spleen size, percentage of blasts, age, and platelet count over 700x10^9/L. The first three are continuous variables with progressively worse prognosis at higher values. A Sokal risk score calculator is available online at: http://www.leukemia-net.org/content/leukemias/cml/cml_score/

(ii) The Hasford or Euro score adds eosinophilia and basophilia to the four clinical variables of the Sokal score. It was developed for CML patients receiving treatment with interferon. A Hasford score calculator is available online at: http://www.leukemia-net.org/content/leukemias/cml/cml_score/

Statistical terms: fixed effect method

A meta-analysis yields an overall statistic that summarizes the effectiveness of the experimental intervention compared with a control intervention. A pooled intervention effect estimate is calculated as a weighted average of the intervention effects assessed in the individual studies. There are four commonly used methods of meta-analysis for dichotomous outcomes: three fixed-effect methods (Mantel-Haenszel, Peto and inverse variance) and one random-effects method (DerSimonian and Laird). The fixed effect method presumes that each study is estimating the same intervention effect. The Mantel-Haenszel method is the most common and is the preferred (the default) fixed-effect method of meta-analysis. Also, this method has been shown to have better statistical properties when there are few events, as in our case. The results are given as relative risks (RRs) and confidence interval. With positive outcomes (such as response and survival), the forest plot shows the investigational arm on the right-hand side of the plot. Thus, with positive results, RRs above 1 indicate that 2nd generation TKIs are better. With negative outcomes, for example, mortality and disease progression) the investigational arm is on the left-hand side of the forest plot. Thus, RRs less than 1 indicate that 2nd generation TKIs are better.

χ² a statistical measure to assess heterogeneity

It is important to examine heterogeneity in a meta-analysis. A χ² test evaluates whether observed differences in results are compatible with chance alone. A low P value (or a large χ² statistic relative to its degree of freedom) provides evidence of statistical heterogeneity of intervention effects (variation in effect estimates beyond chance). The χ² test has low power in meta-analysis when there is a small sample size or few studies, like in the present study. Also, some claim that there will always be heterogeneity in meta-analysis, whether this is detected or not. Methods have been developed to quantify inconsistency across studies. This moves the focus away from testing whether heterogeneity is present to assessing its impact on the meta-analysis. A useful statistic for quantifying inconsistency is I². This description of the percentage of the variability, in effect, estimates that it is due to heterogeneity rather than sampling error (chance). A rough guide for interpretation would be:

- 0-40%, might not be important;
- 50-60%, may represent moderate heterogeneity;
- 50-90%, may represent substantial heterogeneity;
- 75-100%, considerable heterogeneity.
References


