First-line treatments for chronic lymphocytic leukaemia: interpreting efficacy data by network meta-analysis

Andrea Messori · Valeria Fadda · Dario Maratea · Sabrina Trippoli

Received: 12 September 2014 / Accepted: 27 January 2015 © Springer-Verlag Berlin Heidelberg 2015

Abstract When multiple treatments are available, network meta-analysis can synthesize evidence and rank relative effectiveness. We applied this approach to current treatments for previously untreated chronic lymphocytic leukaemia. Data search was conducted in PubMed and websites of regulatory agencies (year 2000 through present time). Our analysis included randomized controlled trials assessing treatments for previously untreated chronic lymphocytic leukaemia. The endpoint of the analysis was the rate of progression-free survival at 3 years. At least two reviewers abstracted study data and outcomes. Agents examined for their relative effectiveness included four monotherapies (chlorambucil, fludarabine, bendamustine, alemtuzumab) and four combination treatments (cyclophosphamide + fludarabine, cyclophosphamide + cladribine, cyclophosphamide + fludarabine + rituximab, cyclophosphamide + fludarabine + alemtuzumab). A Bayesian network meta-analysis was conducted to comparatively evaluate these treatments. Nine trials (3620 patients) were included in the analysis. Odds ratio (with 95 % credible intervals) was estimated for all direct and indirect comparisons. Combinations treatments were found to be significantly more effective than single-agent treatments. Ranking in effectiveness was as follows: (1) cyclophosphamide + fludarabine + rituximab, (2) alemtuzumab, (3) cyclophosphamide + fludarabine + alemtuzumab, (4) cyclophosphamide + fludarabine and (at same ranking) cyclophosphamide + cladribine, (6) fludarabine, (7) bendamustine and (8) chlorambucil. Bendamustine fared worse in our analysis than in its pivotal trial. Overall, the estimated rankings appeared to be robust according to probabilistic analysis. Numerous indirect comparisons were assessed in the absence of RCTs. Some progression-free rates were unavailable from the publications and were estimated graphically from the survival curve. We generated an updated synthesis of the effectiveness of these treatments and we ranked them according to a Bayesian probabilistic model. In our probabilistic analysis, cyclophosphamide + fludarabine + rituximab ranked first in the base case while the worst-case scenario of this analysis placed this treatment at a remarkable second place.

Keywords Meta-analysis · Chronic lymphocytic leukaemia · Chlorambucil · Fludarabine · Bendamustine · Alemtuzumab · Cyclophosphamide · Cladribine · Rituximab

Introduction

The literature on the effectiveness of treatments for previously untreated chronic lymphocytic leukaemia (CLL) covers more than 50 years and is therefore sufficiently settled [1]. However, a relatively large number of new agents have been approved in recent times [2–11], and for this reason, an updated comparative synthesis of effectiveness data can be of interest. In the present study, we examined the data of comparative effectiveness published after 2000 in randomized controlled trials (RCTs) and we applied network meta-analysis to synthesize this information. The main purpose of our study was to comparatively evaluate the effectiveness of current treatments and to rank them according to their effectiveness. In particular, our assessment of outcomes was focused on progression-free rates at 3 years in order to capture a sufficiently long time interval. Furthermore, our analysis differed from others...
recently published in this area because we did not include
convention combination treatments, the importance of which
has markedly lessened over the past decade.

Methods

Study design

Our study was designed according to a protocol registered on
the PROSPERO website of the University of York (protocol
CRD42014013596). The design was in agreement with the
PRISMA 2009 checklist [12].

Data source and search

We carried out a literature search aimed at identifying the
RCTs potentially eligible for our analysis. This search was
conducted in PubMed (last query run on 31 December 2014)
and covered the period from January 2000 to present date
according to PubMed’s definition of this interval. A single
search term (“chronic lymphocytic leukaemia”) was
employed in combination with the filter “randomized con-
trolled trials”. Since the number of citations extracted through
this procedure was small (less than 200), we analysed all of
these articles (by examining the abstract or, when necessary,
their full text) and we identified the RCTs that met our inclu-
sion criteria (see below). We also searched the Cochrane Li-
brary for any recent systematic review on the subject, the
ClinicalTrials.gov database and the websites of European
Medicines Agency (EMA) and US Food and Drug Adminis-
tration (FDA). The keyword chronic lymphocytic leukaemia
was employed also for these additional searches.

Inclusion criteria and study selection

Our analysis included the trials that met the following criteria:
(a) randomized design, (b) previously untreated patients with
CLL, (c) evaluation of treatments currently approved by EMA
and/or FDA, and (d) determination of progression-free surviv-
al (PFS) at 3 years for each treatment group. No restrictions
were employed in terms of physical fitness or age of the study
population.

Trials with designs in which patients were firstly given an
induction treatment were evaluated for response and then ran-
donized into two groups were excluded. Randomized studies
reported in duplicate publications were included once only.

Data extraction and quality assessment

For each trial, we extracted the basic information needed for
our analysis as well as a copy of the Kaplan-Meier curve of
PFS when available. Information on disease staging at
baseline was recorded as well. Extraction of this information
from the trials was made by AM, VF and ST.

The probability of PFS at 3 years (including the effect of
censoring according to the Kaplan-Meier approach) was
drawn from the text of the articles when this information
was explicitly reported in numerical form. In the remaining
cases, we determined these probabilities from a graphical
analysis of the Kaplan-Meier curves. This latter analysis was
carried out as previously described [13]; in addition, a specific
software of graphical analysis of $y$ vs $x$ curves was employed
to improve the quality of this procedure (software WinDIG
2.5 F by D. Lovy, University of Geneve, Switzerland). Extrac-
ction of probabilities of PFS at 3 years from each of the study
arms was made in duplicate by AM and Vf; discrepancies
were resolved by consensus.

Since the input required by the network meta-analysis was
represented by the rates of PFS at 3 years expressed in the
form of numerator and denominator for each of the study
arms, the denominators (total number of patients) were direct-
ly drawn from the reports of the RCTs while the numerators
were estimated by multiplying the total number of patients by
the probability of PFS at 3 years determined from the Kaplan-
Meier curve. This approximation ensures an acceptable ad-
justment for the presence of censored patients.

As regards the assessment of methodological quality,
two reviewers (AM and ST) assessed the risk of bias in
included studies by using the Cochrane Collaboration’s
tool [14] (which addresses six domains: sequence gen-
eration, allocation concealement, blinding of participants
and personnel, incomplete outcome data, selective out-
come reporting and other sources of bias). Studies with
adequate procedures in all domains were considered to
have a low risk of bias.

Data synthesis and analysis

We employed Bayesian network meta-analysis [15–17]. This
approach to make indirect comparisons is increasingly being
used and can be considered the current standard in this field.
As compared with the traditional frequentist approach [18],
the Bayesian approach has one main advantage in that all
treatments included in the comparisons are incorporated into
a single model; in contrast, in most frequentist approaches,
there are as many separate analyses as the number of compar-
sions being studied. Another advantage of the Bayesian ap-
proach is that this technique enables rank ordering of each
treatment. As opposed to traditional confidence intervals
adopted in frequentist analysis, the Bayesian output re-
ports credible intervals (CrIs), which can be directly
interpreted as the probability of an event residing in the
reported range.

The output of the network meta-analysis consisted of the
odds ratios (ORs) for all combinations of pairwise
comparisons. Both direct comparisons and indirect comparisons were considered. The values of OR were associated with their respective 2.5 to 97.5 % CrI (i.e. 95 % CrI), that reflects a formal level of statistical significance at 5 %. Direct comparisons are those for which at least a single clinical trial was available while indirect comparisons are those for which a “real” trial is lacking. The Bayesian model adopted for our analysis [17] has been developed by the NICE Support Unit (UK). Further technical details on our Bayesian model are presented in the Appendix.

Results

Literature search, identification of included studies and data extraction

Our literature search, which is summarized in Figure S1 (see Appendix), extracted a total of 176 citations. For a further scrutiny of the material eligible for our analysis, we examined the full text of 36 articles. After examining these papers, we selected a total of nine RCTs (total number of treatment arms: 19; total number of patients: 3620) that met our inclusion criteria. None of these trials had adopted a double-blind design. As regards the methodological quality, all studies were evaluated to have a low risk of bias. Details and on these points and further information on the trials is presented in Table S1 and Figures S1–S2 of the Appendix and later on in Table 2.

The treatments used in these RCTs were four single-agent regimens (chlorambucil (CHL) in four treatment arms, fludarabine (FLU) in four treatment arms, bendamustine (BEN) in one treatment arm and alemtuzumab (ALE) in one treatment arm) and four combination regimens based on cyclophosphamide (CYC), i.e. CYC + FLU in five treatment arms, CYC + cladribine (CLA) in one treatment arm, CYC + FLU + rituximab (RIT) in two treatment arms and CYC + FLU + ALE in one treatment arm. As regards disease stage, the inclusion criteria of the trials were as follows: Hallek et al. [2], Binet stage C or stages A–B with confirmed active disease; Catovsky et al. [3], Binet stages C or B or A-progressive; Knauf et al. [4] and Lepretre et al. [5], Binet stages B or C; Hillmen et al. [6], Rai stages I through IV with evidence of progression according to NCI-WG 1996 criteria; Eichhorst et al. [9] and Eichhorst et al. [10], Binet stage C and Binet stages B or A if patients had rapid disease progression or symptoms from enlarged lymph nodes and organs or had severe B symptoms; Robak et al. [7], Rai stage II through IV (with further selection criteria for stage II) and Flinn et al. [8], NCI-WG 1996 criteria with evidence of progression.

Bayesian network meta-analysis

Figure 1 shows the network of the comparisons for which “real” trials were available (direct comparisons). To carry out our analysis, firstly, we separately ran the fixed-effect and the random-effect models. According to the deviance information criterion, the fixed-effect model showed a better fit than the random-effect one. Therefore, our study was based on the results generated by the former.

Our results are shown in Table 1. Interestingly enough, the demonstrations of superiority resulting from the network meta-analysis were in agreement with the findings of individual RCTs. The only exception was BEN that ranked worse according to the network analysis than according to the effectiveness estimated from the RCT [4]; in addition, as regards the comparison between FLU and CHL, while our meta-analytic OR was significantly in favour of the former, only one [9] of the two trials by Eichhorst et al. [9, 10] had reached the threshold of statistical significance in favour of FLU.

In our probabilistic analysis (Fig. 2), CYC + FLU + RIT ranked first in the base case and, more interestingly, the worst-case scenario of this analysis placed this treatment at a remarkable second place (according to 95 % CrIs). On the other hand,
BEN and CHL had quite unfavourable ranks. The overall ranking in effectiveness was the following: (1) CYC + FLU + RIT; (2) ALE; (3) CYC + FLU + ALE; (4) CYC + FLU and, at the same ranking, CYC + CLA; (6) FLU; (7) BEN; (8) CHL. Rankings appeared robust according to probabilistic analysis, as confirmed by the histogram of rankings. Given that most of the pairwise comparisons included just a single RCT, we could not formally assess statistical heterogeneity and publication bias. Finally, running two further Markovian chains as a form of sensitivity analysis (data not shown) gave full confirmation of the results of our primary analysis.

**Discussion**

Our results generated an updated synthesis of the effectiveness data available for the main first-line treatments for CLL and was successful in defining their respective rankings. In a context where eight different treatments are available and have in fact been tested in RCTs, this comprehensive picture of current therapeutic evidence is of interest under several viewpoints. In particular, the information on relative rankings (along with the probabilistic analysis) represents—in our view—the most interesting result of our analysis.
The histogram of rankings (Fig. 2) clearly indicated that combination treatments fared much better than monotherapies. Among the various combination treatments, the three-agent regimen of CYC + FLU + RIT ranked first in the great majority of the simulations. This regimen was significantly more effective than CYC + FLU (OR=0.443 with 95% CrI of 0.335 to 0.590, which is in line with the study by Hallek et al. [2]) and was also superior to CYC + CLA (OR=0.478 with 95% CrI of 0.289 to 0.776 which is a comparison that was not tested by any real trial). The other three-agent regimen (CYC + FLU + ALE) fared numerically worse than CYC + FLU + RIT, but the difference was not statistically significant.

Fig. 2 Histogram of rankings generated by the Bayesian network meta-analysis. The graphs reflect a total of 20,000 iterations and consist of as many histograms as the treatments (N=8) included in the analysis. In each panel, the histogram shows the percent distribution of the simulations across ranks 1 (most effective treatment) through 8 (least effective treatment); the y-axis shows probability on a 0 to 1 scale. The overall ranking in effectiveness (with 95% CrIs in parenthesis) was the following: (1) CYC + FLU + RIT: 1 (1 to 2); (2) ALE: 2 (1 to 5); (3) CYC + FLU + ALE: 3 (1 to 6); (4) CYC + FLU: 4 (3 to 5) and, at the same rank in, CYC + CLA: 4 (2 to 5); (6) FLU: 6 (5 to 7); (7) BEN: 7 (6 to 8); (8) CHL: 8 (7 to 8). CHL chlorambucil, FLU fludarabine, BEN bendamustine, ALE alentuzumab, CLA cladribine
this likely reflects the safety problems with CYC + FLU + ALE that have previously been highlighted by Lepretr et al. [5].

Our results on BEN were at variance with the result of the pivotal RCT available for this agent (Knauf et al. [4]); in fact, the superiority found for BEN in this trial (as compared with CHL) was not confirmed by our Bayesian analysis. This latter finding can be explained because, in the trial by Knauf et al. [4], the control group had a worse PFS pattern than that of the other six patient arms given CHL included in our analysis. Another explanation is that the inclusion criteria adopted by Knauf et al. [4] (and also by Lepretr et al. [5]) selected a population with a slightly more advanced disease than that enrolled in the other six patient arms given CHL.

The combination of obinutuzumab plus chlorambucil has been tested in a three-arm randomized trial [19] that compared obinutuzumab plus chlorambucil or rituximab plus chlorambucil with chlorambucil monotherapy in 781 patients with previously untreated CLL and coexisting conditions (i.e. a score higher than 6 on the Cumulative Illness Rating Scale or an estimated creatinine clearance of 30 to 69 ml per minute). This trial did not meet the inclusion criteria of our analysis because the progression-free survival curve did not reach the time-point at 3 years requested by our analysis. These results however deserve to be narratively mentioned because obinutuzumab was found to be superior to rituximab when each was combined with chlorambucil (median progression-free survival: 26.7 months with obinutuzumab plus chlorambucil, 16.3 months with rituximab plus chlorambucil, 11.1 months with chlorambucil alone). More interestingly, the hazard ratio found in this trial for the comparison of obinutuzumab plus CHL vs CHL alone (hazard ratio, 0.18; 95% confidence interval: 0.13 to 0.24) was in the same range of the OR found in our analysis for the comparison of CYC + FLU + RIT vs CHL alone (OR, 0.156; 95% CrI, 0.102 to 0.240).

Most findings generated by our analysis refer to indirect head-to-head comparisons, i.e. comparisons for which no real trial has ever been conducted (Table 1). The limitations intrinsic to indirect comparisons are well known [20–23], and our study did incorporate all of them. Apart from these general considerations, one specific limitation of our study was represented by the need to reconstruct the rate of PFS at 3 years by graphical analysis of the Kaplan-Meier curves; this procedure was necessary in seven of the nine RCTs. However, for a long time, this technique has invariably shown a good performance [24–26].

Last but not least, several randomized studies available from the literature did not meet the inclusion criteria set by our meta-analytic protocol, mainly because these studies were old, i.e. published before 2000. However, since most of these studies evaluated conventional combination regimens that are no longer used (see Appendix Table 1 of the study by Terasawa et al. [11]), this exclusion did not substantially affect the main message resulting from our analysis. Of more concern is the fact that some studies published after 2000 were excluded because they did not present any information on progression-free rates at 3 years; however, the studies we missed for this reason were only a few, as clearly suggested by the comparison between our analysis on progression-free survival and that previously reported by Terasawa et al. [11] (namely Appendix Table 6 of the Japanese researchers).

The main message arising from our study is that, among combination regimens, the CYC + FLU + RIT regimen not only ranked first in effectiveness but its superiority was statistically significant and clinically relevant in comparison with the other treatments we tested (with the exception of alemtuzumab monotherapy). On the other hand, all single-agent treatments (with the exception of alemtuzumab) occupied the worse ranks; in particular, bendamustine monotherapy fared worse in our indirect comparisons than in the results suggested by its pivotal trial.

In conclusion, our study generated an updated synthesis of the effectiveness data available for these treatments and was successful in defining their respective rankings.

Conflict of interest The authors declare that they have no conflict of interest.

References


Q1. Please check captured article title, if appropriate.
First-line treatments for chronic lymphocytic leukemia: interpreting efficacy data by network meta-analysis

Appendix

Technical details on the Bayesian model of network meta-analysis

The Bayesian model of network meta-analysis [15-17] is increasingly being used to simultaneously make direct and indirect comparisons and can be considered the current standard in this field. As compared with the traditional frequentist approach [18], the Bayesian model has one main advantage in that all treatments included in the comparisons are incorporated into a single model; in contrast, in most frequentist approaches (e.g. the Bucher method[18]) there are as many separate analyses as the number of comparisons being studied. Another advantage of the Bayesian approach is that this technique enables rank ordering of each treatment. As opposed to traditional confidence intervals adopted in frequentist analysis, the Bayesian output reports credible intervals (CrIs), which can be directly interpreted as the probability of an event residing in the reported range.

The Bayesian analysis involves a formal combination of a prior probability distribution that reflects a prior belief of the possible values of the effect of interest, and the likelihood distribution of the effect based on the observed data, to obtain a posterior distribution. In the absence of real data, prior probabilities are assigned by using vague, flat, or non-informative priors (that are generally small numbers between 0 and 3).

The Bayesian model adopted for our analysis [17] has been developed by the NICE Support Unit (UK) and is available as fixed-effect model and random-effect model. Both employ a random sequence of chains, called Markov chain Monte Carlo simulation. Each chain must be run for a length of time sufficient to allow model convergence (burn-in) before estimating posterior probabilities. We created the model by using the binary outcome of how many subjects in each arm of each study were free of progression at 3 years. Randomization within each study was preserved by specifying each arm in each study separately, thus accounting for the effect of the comparator.

We planned to run both the fixed-effect model and the random-effect model and to choose the best one for our purposes on the basis of the deviance information criterion (which is a sort of goodness-of-fit test implemented in the software). Results were presented as odds ratio (OR).

Finally, as a sensitivity analyses, we changed the initial values from which each Markov chain Monte Carlo simulation began, as is customary in the Bayesian framework [15-17].
Recent advances in computing power and the development of sophisticated software have greatly facilitated the use of Bayesian statistics. All of our analyses were conducted by using the software package WinBUGS 1.4.3 (Cambridge, United Kingdom) in combination with the meta-analysis code developed by the National Institute for Health and Care Excellence [17].

For each model, we ran 20,000 iterations within a single Markovian chain and then another 20,000 iterations for the so-called “burn-in” of the parameters. In both cases, this series of iterations reached the convergence of assigned prior distributions. Calculations of rankings required an additional 20,000 iterations. According to the deviance information criterion, the fixed-effect model showed a better fit than the random-effect one. Therefore, our study was based on the results generated by the former.
Figure S1. PRISMA schematic.

This figure summarizes the phases of our literature search and their respective results. Phase A scanned MEDLINE (PubMed version) while Phase B scanned other potential sources of RCTs.

Note: This note explains why, in comparison with the literature search published by Terasawa et al.[11], fewer studies were included in our analysis than in that of Tarawawa and coworkers. The main reasons were the following three:

a) the numerous trials based on conventional combination regimens (reported by Terasawa et al. in Appendix Table 1 on their pages 9 and 10) were not included in our analysis for two reasons: i) in most cases because these trials were published before 2000 whereas our literature search ranged from 2000 to 2014; ii) in a few cases because the trials did not report any information on progression-free survival at 3 years;

b) the recent trial on obinutuzumab plus chlorambucil conducted by Goede et al. (New England Journal of Medicine, issue of 20 March 2014) was not included for two reasons: i) this treatment is not approved by EMA or FDA whereas our meta-analytic protocol included this criterion; ii) the information on progression-free rates at 3 yrs was not available because of an insufficient length of the follow-up; in fact. one pre-declared criterion of our meta-analytic protocol was that the RCT reported the information on progression-free survival at 3 yrs.

c) some trials published as abstracts, which were included in the review by Terasawa et al., were not eligible for our analysis because these trials were not available for the sources of literature employed in our study and also because no information on progression-free at 3 yrs was available.
Table S1. Characteristics of the 9 randomized trials included in the network meta-analysis and results of our estimation of patients free of progression at 3 yrs.

<table>
<thead>
<tr>
<th>First author</th>
<th>Patient groups</th>
<th>Patients without progression at 3 years*</th>
<th>Total number of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catovsky et al.⁵</td>
<td>CHL</td>
<td>89</td>
<td>387</td>
</tr>
<tr>
<td></td>
<td>FLU</td>
<td>60</td>
<td>194</td>
</tr>
<tr>
<td></td>
<td>CYC+FLU</td>
<td>111</td>
<td>196</td>
</tr>
<tr>
<td>Eichorst et al.¹⁶</td>
<td>CHL</td>
<td>29† (29%)</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>FLU</td>
<td>24† (26%)</td>
<td>93</td>
</tr>
<tr>
<td>Eichorst et al.¹⁷</td>
<td>FLU</td>
<td>94† (53.6%)</td>
<td>175</td>
</tr>
<tr>
<td></td>
<td>CYC+FLU</td>
<td>63† (35.7%)</td>
<td>176</td>
</tr>
<tr>
<td>Hillmen et al.⁶</td>
<td>CHL</td>
<td>18§</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td>ALE</td>
<td>52</td>
<td>149</td>
</tr>
<tr>
<td>Hallek et al.²</td>
<td>CYC+FLU</td>
<td>187</td>
<td>409</td>
</tr>
<tr>
<td></td>
<td>CYC+FLU+RIT</td>
<td>267</td>
<td>408</td>
</tr>
<tr>
<td>Knauf et al.⁴</td>
<td>CHL</td>
<td>116</td>
<td>157</td>
</tr>
<tr>
<td></td>
<td>BEN</td>
<td>120</td>
<td>162</td>
</tr>
<tr>
<td>Flinn et al.⁸</td>
<td>FLU</td>
<td>33</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td>CYC+FLU</td>
<td>68</td>
<td>141</td>
</tr>
<tr>
<td>Lepretre et al.⁵</td>
<td>CYC+FLU+RIT</td>
<td>68</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>CYC+FLU+ALE</td>
<td>60</td>
<td>83</td>
</tr>
<tr>
<td>Robak et al.⁷</td>
<td>CYC+FLU</td>
<td>71</td>
<td>212</td>
</tr>
<tr>
<td></td>
<td>CYC+CLA</td>
<td>74</td>
<td>211</td>
</tr>
</tbody>
</table>

†These values were estimated from the graph of the respective Kaplan-Meier curves because the article did not explicitly report the percentage of progression-free patients at 3 yrs.

*Estimated by multiplying the total number of patients by the value of PFS at 3 years determined from the Kaplan-Meier curve.

§In this arm, the probability of PFS, which was available at 34 months, was assumed to represent also the value at 36 months.

Abbreviations: CHL, chlorambucil; FLU, fludarabine; BEN, bendamustine; ALE, alemtuzumab; CLA, cladribine.

Note: Superscript numbers refer to the Reference List of our original article.
Figure S2. Application of the Cochrane Collaboration’s tool for assessing risk of bias in randomised trials.

The figure shows the summary of risk-of-bias assessments for the 9 randomized, controlled trials included in our analysis. Low risk of bias is represented by green circles (see reference 11 for further details). Note: Superscript numbers refer to the reference list of our original article.