FLUDARABINE PHOSPHATE FOR THE FIRST-LINE TREATMENT OF CHRONIC LYMPHOCYTIC LEUKAEMIA

THE EVIDENCE REVIEW GROUP'S REPORT

The Evidence Review Group consists of:

Centre for Health Economics, University of York & **NHS Northern and Yorkshire Regional Drug and Therapeutics Centre**

S Walker, S Palmer

S Erhorn, S Brent, A Dyker, L Ferrie, W Horsley, K Macfarlane, S White, S Thomas

About the authors:

The ERG is a collaboration between two centres, the NHS Northern and Yorkshire Regional Drug and Therapeutics Centre (RDTC) and the Centre for Health Economics (CHE), University of York.

The RDTC was established in 1991 to promote safe effective prescribing and economical drug usage, and to provide a source of independent authoritative advice on pharmaceutical and therapeutic issues throughout the former Northern and Yorkshire region. The RDTC co-ordinates prescribing activities, provides a poisons and medicine information service and is the teratology information service for the UK. The Centre is one of four NHS regional monitoring centres for the Medicines and Healthcare Regulatory Authority (MHRA).

CHE is a research unit of the University of York. The Centre's aim is to undertake high quality research that is capable of influencing health policy decisions. The largest programme of work at CHE is that on economic evaluation and health technology assessment which focuses on a range of methodological and applied work. This includes full technology assessment reviews and evidence review reports for the National Institute for Health and Clinical Excellence (NICE). Recent assessment reports for NICE include treatments for prostate and ovarian cancer, psoriasis and psoriatic arthritis.

Contact the authors:

Regional Drug and Therapeutics Centre
Wolfson Unit, Claremont Place, Newcastle upon Tyne, NE2 4HH

Centre for Health Economics
University of York, Heslington, York, YO10 5DD

Acknowledgements:

The ERG received clinical advice on the current management of CLL in the UK from Dr Geoff Summerfield (Gateshead Health NHS Trust, Dept of Heamatology, Queen Elizabeth Hospital). The interpretation and views expressed in this document, and any errors, relate to the ERG authors alone.

Conflicts of interest:

The authors to this report have no conflicts of interest.

CONTENTS

TABLES

FIGURES

Chapter 1
Summary

1. Introduction

This document critically evaluates the evidence submission, from Schering Health Care Ltd (SHC), on the clinical and cost-effectiveness of fludarabine phosphate (Fludara®) (F) or fludarabine plus cyclophosphamide (FC) for the first-line treatment of chronic lymphocytic leukaemia.[1] This report identifies the submission's strengths and weaknesses, supplemented, where appropriate, with our own analysis. Clinical experts were asked to advise the Evidence Review Group (ERG) to help inform the review.

1.1 Scope of the submission

The perceived aim of the SHC submission was to evaluate the clinical and cost-effectiveness of fludarabine (F), or fludarabine plus cyclophosphamide (FC), relative to chlorambucil (Chl) in the first-line treatment of patients with Binet Stage A progressive, or Binet Stages B and C, Chronic Lymphocytic Leukaemia (CLL) with sufficient bone marrow reserve.[2]

1.2 Summary of submitted clinical evidence

Of the seven studies included in the submission, only two were fully published [3, 4] and the remaining five studies were available in abstract form only.[5-9] Fludarabine and fludarabine plus cyclophosphamide were compared with chlorambucil (Chl) in five studies.[4-6, 8, 9] Two studies compared fludarabine with fludarabine plus cyclophosphamide.[3, 7] Only one study compared all three regimens.[5, 10] Fludarabine was also combined with chlorambucil in the Rai *et al* study; however assignment of patients to the fludarabine plus chlorambucil arm was stopped when a planned interim analysis revealed excessive toxicity and a response rate that was no better than the rate with fludarabine alone.[4] The Evidence Review Group felt it was appropriate to disregard this as a likely therapeutic option for previously untreated chronic lymphocytic leukaemia patients.

All studies, with one exception [9], showed an improvement in overall response (OR) in those patients receiving fludarabine compared to those receiving Chl.[3-8] In all but one [8] of the studies comparing fludarabine or fludarabine plus cyclophosphamide with Chl, there was a higher complete response (CR) rate for the fludarabine containing arms.[4-6, 9] Although progression-free survival (PFS) was stated as a primary outcome measure in five

studies [3-7], this outcome was fully reported in only three.[3, 4, 7] In one study comparing differences in median progression-free survival between the fludarabine and chlorambucil regimens, there was a significantly longer duration of response in the fludarabine arm.[4] Two studies demonstrated a significantly longer duration of response with the fludarabine plus cyclophosphamide combination compared to fludarabine alone.[3, 7] At present the follow-up periods of the studies included in the submission are too short to demonstrate any significant improvement in overall survival (OS). Therefore, fully matured survival data are necessary to ascertain whether any improvement in progression-free survival translates in to an increase in OS.

Three studies included quality of life (QoL) analyses. However, only limited data from the CLL4 study are presented.[10] In this study quality of life was the same for each treatment group at baseline and at 12 months and correlated with the quality of response. It is anticipated that the results of further QoL analyses are likely to become available within the next year.

Because five of the studies included in the submission are not fully published and report only preliminary results in abstract form, there are insufficient data presented to fully assess the validity of these studies.[5-9] Although the unpublished CLL4 study [5] is supplemented with additional patient level data [10] provided by the manufacturer to support the health economic analyses, this supplemental data are not in the public domain and therefore cannot be verified externally. Until these studies are fully published and the complete data made available for evaluation, these results must be interpreted with caution.

1.3 Summary of submitted cost-effectiveness evidence

Two papers were identified in both the manufacturer's submission and the ERG searches which reported on the cost-effectiveness of fludarabine monotherapy in comparison to chlorambucil in the management of CLL in previously untreated patients.[11, 12] Neither of the studies was considered particularly relevant due to the limited clinical and economic evidence on which the studies were based (mainly due to the limited evidence available at the time these studies were undertaken) and the restricted range of comparators considered. Neither of these studies considered the cost-effectiveness of fludarabine combined with cyclophosphamide as a first-line treatment for CLL. Consequently the submission by the manufacturer was considered to comprise the most relevant evidence to consider for the purposes of this STA.

The manufacturers submission included on a '*de-novo*' decision analytic Markov model to

estimate the cost-effectiveness of treatment with (i) fludarabine monotherapy (F), (ii) fludarabine in combination with cyclophosphamide (FC) and (iii) chlorambucil (Chl). The model used individual patient data from the CLL4 trial to model transition probabilities related to first-line treatment with these treatments. The costs of first-line treatment were derived from an audit of UK patients from the CLL4 trial. The model was based on a lifetime time horizon and included the costs and consequences of further treatments required after first-line treatment had failed. Data on the costs and effects of further treatment (including re-treatment, second-line and salvage therapies) were derived from a combination of secondary sources and assumptions by the manufacturers. Results were presented in terms of cost per QALY gained, with quality of life estimates informed by a separate systematic review.

In the original submission by the manufacturers, the incremental cost-effectiveness ratio (ICER) of FC compared to Chl was £2,602 per additional QALY. FC was reported to dominate F (i.e. was less costly and more effective). These results were based on an approach which assumed that median (as opposed to mean) survival was equal in all treatments. An addendum was submitted by the manufacturers which presented similar results based on an approach which equalised mean survival. This latter approach was considered by the ERG to be a more appropriate assumption. The results presented in the addendum increased the ICER of FC compared to Chl to £3,244 per additional QALY. FC continued to dominate F. The results of the sub-group analysis presented by age and Binet stage did not substantially alter these results. Similarly, the results were reported to be robust to a wide range of sensitivity analyses undertaken by the manufacturers. The results were most sensitive to the time horizon of the model, such that FC did not appear cost-effective at a time horizon of 5 years.

1.4 Commentary on the robustness of submitted evidence

1.4.1 Strengths

The ERG felt that the SHC submission was generally of good quality. There were no major errors or omissions in the clinical evidence. The majority of the data quoted within the submission was a fair and accurate representation of the original reference data. The ERG noted the limitations of existing cost-effectiveness studies in this area and considered the economic model submitted by the manufacturers to be the most relevant source for the purpose of informing this STA. The economic model structure (including the comparators) was considered appropriate for the decision problem, and the data sources used to inform the model were deemed appropriate from a UK NHS perspective. A range of sub-groups was considered and uncertainty in parameter estimates was addressed using probabilistic approaches.

1.4.2 Weaknesses

The majority of the reference data presented in the submission were not fully published and only available in abstract form. Therefore, the ERG felt that until these studies are fully published and the complete data made available for evaluation, these results must be interpreted with caution.

The ERG identified a number of potential sources of weakness in the manufacturer's economic submission. In particular, a number of issues were identified which may have introduced possible bias into the results. Most of these issues appeared to act in favour of the FC regimen such that it is likely that the manufacturer's results are overly optimistic towards this regimen. The robustness of the manufacturer's results to some of these issues was explored in additional work undertaken by the ERG. The cost-effectiveness of FC appeared relatively robust to wide variation in several of the key assumptions made by the manufacturers. The ERG was concerned with the approach the manufacturer used to estimate a number of key probabilities derived from the CLL4 trial data. Due to the structure of the model it was not possible to fully explore the potential robustness of the manufacturer's results to alternative assumptions. However, work undertaken by the ERG brought into question the validity of the assumptions underpinning the extrapolation of data over a lifetime time horizon.

1.4.3 Areas of uncertainty

The fludarabine summary of product characteristics (SPC) does not mention the use of fludarabine in combination with other chemotherapeutic agents. The dose for oral therapy in combination with cyclophosphamide does not appear to be a licensed dose and is not mentioned in the SPC.

The SPC for cyclophosphamide states that it is frequently used in combination chemotherapy regimens involving other cytotoxic drugs and that it is recommended that the calculated dose be reduced at the discretion of the clinician when it is given in combination with other antineoplastic agents or radiotherapy, and in patients with bone marrow suppression. However, the ERG feels that the efficacy of the FC regimen is still under investigation and that the recommendations outlined in the BSCH guidelines are expected to be revised following the outcomes of the CLL4 study. Therefore, the ERG sought clarification on this matter from the company.

The manufacturer believes the proposed regimen falls with the current licenses and state they are not, therefore, considering an extension to the fludarabine license. The dosing

regimen for the FC combination was agreed by expert clinicians within the MRC/LRF UK-CLL group. However, independent expert advice given to the ERG confirms that the FC regimen is increasingly used for the first-line treatment of CLL and that the dosing regimen chosen also reflects current practice.[13]

Chapter 2
Background

2.1 Fludarabine for chronic lymphocytic leukaemia

Chronic lymphocytic leukaemia (CLL) is defined as a slow progressive form of leukaemia characterised by an increased number of lymphocytes.[14] The majority of lymphocytes have a common appearance, being of small or medium size, with clumped nuclear material (chromatin), an indistinct or absent nucleoli and little cytoplasm.[15] The other type of lymphocyte commonly observed in approximately 15% of patients is a cell called a prolymphocyte which appears large with a prominent nucleolus.[15, 16] The general symptoms of CLL are tiredness, night sweats, weight loss, anaemia and associated symptoms, and increased susceptibility to infection.[15] The lymphocytes may also accumulate in the lymph nodes and spleen resulting in lymphadenopathy, splenomegaly, and other abdominal masses.[15, 16] Frequently the condition is identified by chance during a routine blood test in the absence of specific symptoms or physical signs. At the point of diagnosis CLL is usually widespread and with some degree of bone marrow involvement. With the exception of blood and marrow transplantation, the condition is inherently incurable with treatment emphasis on maintaining an acceptable state of health and inducing remission when required.[16]

2.1.1 Incidence

B-cell CLL is reported to be the most common leukaemia, representing approximately 25% of all cases of leukaemia.[17] In England and Wales in 2003 there were 6,198 cases of leukaemia; [18, 19] assuming that 25% of these are B-cell CLL means that there were approximately 1550 new cases of B-cell CLL diagnosed in 2003. This indicates a crude incidence in this population of approximately 3 per 100,000 per year; [18-20] however, this belies the demographics of its incidence. CLL is rare below the age of 30 years with 20-30% of patients presenting under the age of 55 years.[15] The peak incidence is between 60 and 80 years and increases up to almost 50 per 100,000 per year after the age of 70 years.[17] It is male-dominant, occurring with a male to female ratio of 2:1.[15, 21]

2.1.2 Diagnosis

Despite recent discoveries of several novel molecular and genetic markers [8] that may indicate the presence and severity of CLL, [15, 17, 21] it is still common practice to rely on established blood counts, serum screens, physical examination and more readily available immunophenotyping.[15, 17, 21] The most common diagnostic criteria, and those advocated

by the British Committee for Standards in Haematology and the European Society for Medical Oncology, are:

- Absolute lymphocyte count ≥ $5x10^9$/L
- Predominance of small, morphologically mature lymphocytes in the blood
- Physically palpable masses in the lymph nodes or spleen or liver
- Immunophenotyping

Additional tests and examinations that may further aid the diagnosis or provide information concerning the prognosis include marrow examination, lymph node biopsy, fluorescence *in situ* hybridization analysis, computed tomography or ultrasound scan and, where available, identification of novel molecular markers and genes.

2.1.3 Prognosis

Two established clinical staging systems are in use to determine a prognosis; the Rai system which was introduced in 1975 [22] and has since been further refined [23], and the Binet system which was introduced in 1981 [24] (see Appendix 1). Factors that are associated with a worse prognosis are: male sex, Binet stage B or C, Rai stage II, III or IV, atypical lymphocyte morphology, lymphocyte doubling time of less than 12 months, and raised serum markers.[15] Additionally, much interest is currently focused on the identification of novel serum markers and specific gene expression [21, 25] for example CD38, β-microglobulin, IgVH gene status and zeta-associated 70 protein.[15, 21, 25] However, these are not yet universally advocated due to variations in measurement and unproven correlation. Appendix 1 also shows the median survival associated with the stages of each of the classifications.

2.1.4 Treatment [17]

Treatment is not advocated for Binet stage A or Rai stage 0 disease. Where specific poor-prognostic markers have been identified for patients in these stages, treatment should only be initiated as part of a trial. For all other patients in these stages, only regular check-ups to assess disease progression are usually required. It has been suggested that approximately a third of patients never require treatment. For patients with advanced or progressive disease, treatment should be initiated depending on the strategy sought; either disease/symptomatic control or long-term survival. For the former, chlorambucil or fludarabine is recommended and, for the latter, treatment should be commenced with fludarabine in combination with either chlorambucil or chlorambucil with rituximab. As the nature of the condition is progressive, relapse is inevitable. For some patients, autologous or

allogeneic stem cell transplant may be feasible and the latter option may even be curative. Such procedures are normally reserved for younger patients who have failed on several other therapies.

2.1.5 Fludarabine [2]

Fludarabine is a deoxyadenosine derivative, which mimics the structure of the natural adenosine substrate. After some metabolic activation within cells the drug is incorporated into DNA. However fludarabine does not permit chain elongation by DNA polymerases, ultimately inducing cell apoptosis. This is the primary mode of action of fludarabine; additionally, it is thought to inhibit ribonucleotide reductase which reduces the available pool of nucleotides and thus enhances the cytotoxicity of the drug. Fludarabine is available commercially as Fludara® (Schering Health Care Limited) in a 10mg tablet and a 50mg vial containing powder for reconstitution for parenteral administration. The licensed dose for one cycle of fludarabine in CLL is 25mg/m^2 body surface area (BSA), for five consecutive days out of 28. The oral dose is 1.6 times the parenteral dose (i.e. 40mg/m^2 BSA). The number of cycles is not explicitly defined; treatment should be continued until the best objective response is achieved and this usually takes six cycles.

2.2 Critique of the manufacturer's description of the background

The SHC submission provided a very comprehensive and detailed background. The disease and current treatment options were discussed in detail. The rationale for the development of the technology and its proposed place in therapy were clearly defined. The description of the technology under assessment was detailed and appropriate and covered all the relevant aspects. However, the description of the relevant comparators and the justification for selection was lacking the detail of previous sections.

Chapter 3
Defining the Decision Problem

3.1 Scope

The scope for this single technology assessment (STA) was clearly defined in the SHC submission. The decision problem considered was the clinical and cost-effectiveness of fludarabine (F), or fludarabine plus cyclophosphamide (FC), relative to chlorambucil (Chl) in the first-line treatment of patients with Binet Stage A progressive, or Binet Stages B and C, Chronic Lymphocytic Leukaemia (CLL) in England and Wales.

3.2 Intervention

The intervention considered in the decision problem was F used as monotherapy or in combination with cyclophosphamide (FC). The submission states that various other drugs have been tried in combination with F, but C was identified as the most promising with laboratory studies confirming a synergistic effect. However, C is not the only agent to have demonstrated synergistic effects with F, other agents such as epirubicin have shown similar results in preliminary studies.[26] Other agents are currently under investigation in combination with F and/or FC.[17, 27] The available literature supports the statement that the FChl combination does not improve response rates, and is associated with life-threatening toxic events when compared to single agent F.[4]

Fludarabine phosphate (Fludara®) is manufactured by Schering Health Care Ltd [2] and is not available generically. Cyclophosphamide is manufactured in the UK as Endoxana® (Baxter Healthcare Ltd [28]) and is available generically (Pharmacia Ltd [29]). List prices stated are correct at the time of writing.[30] It is unclear whether the FC combination regimen is licensed (see section 1.4.3).

3.3 Patient population

The manufacturer stated that the patient population considered in the decision problem should be in line with the fludarabine license and, therefore, chemotherapy-naïve patients with B-cell CLL and ‘sufficient bone marrow reserves’. First-line treatment should only be initiated in those with advanced disease (Binet stage C), or Binet Stage A/B with disease-related symptoms or evidence of progression.[2] The statement concerning the stage of patients in whom first-line treatment should be initiated is in concordance with the BSCH and ESMO guidelines.[15, 17] However, the CLL4 study allowed enrolment of patients with Binet Stage B without progressive features, and did not specify that patients should have

'sufficient bone marrow reserves'. However, independent expert opinion given to the ERG indicates that these differences should not have a significant bearing on the outcome of the CLL4 study, as in practice very few Binet Stage B patients would present without disease related symptoms.[13] Therefore, the patient population considered is in line with UK practice.

3.4 Comparators

The manufacturer chose Chl as the most relevant comparator in the first-line setting, stating that prior to the advent of F and FC, Chl, was recognised as the standard treatment for initial therapy in CLL patients. In the UK, Chl is used first-line to treat 60% of patients with CLL. The figures quoted as to the frequency with which the various treatments are used are derived from a single table of data provided by a specialist healthcare data company.[31] Additional information provided by the company in response to concerns regarding this data stated "these figures are taken from the IMS Oncology Analyser data which is an independently provided data source that it is collected retrospectively from patient case records from a representative panel of 121 UK treating clinicians. All drug therapies used in the treatment of patients' cancer are recorded from 1999 onwards thereby enabling the analysis of usage of specific therapies and treatment pathways within specific oncology areas". Although these data are unpublished and, therefore, cannot be externally validated, independent expert clinical advice given to the ERG confirms Chl to be the most relevant comparator for the decision problem.[13] Furthermore, the moderately intensive dose of Chl (70mg/m^2) used in the CLL4 study is adequate and in line with that used in UK practice. This is supported by the BCSH and ESMO guidelines and the available literature.[15, 17] Discussions with the ERG's clinical expert suggest that alemtuzumab (Campath®) may become a major option for first-line treatment of CLL in the future. However, at present, there are only limited interim data available on its use first-line [32], and it is currently only indicated for the treatment of B-CLL in patients who have been treated with alkylating agents and who have failed fludarabine therapy.[33]

3.5 Trial Outcomes

There is an issue with measuring overall survival (OS) as a study outcome because of the long follow-up period required. This is particularly the case for CLL where median survival is of the order of 10 years. PFS is often taken to be a suitable surrogate non-patient orientated endpoint, although exactly how well this correlates with OS is unknown. The Quality of Life (QoL) measure referred to is health-related only and may, therefore, omit certain aspects of patients QoL. Health-Related Quality of Life (HRQoL) is a subjective measure and comparisons in values between trials may be compounded by subtle trial differences.

3.6 Key issues

All the points listed as key issues are appropriate and are supported by the BCSH guidelines [15] and the available literature.

Chapter 4
Clinical Effectiveness

4.1 Search Strategy

A systematic literature search was undertaken by the ERG to verify the completeness of the methodology used by the manufacturer to retrieve relevant clinical studies presented in the submission. The inclusion and exclusion criteria and the search strategy used by the ERG are included in Appendix 2.

Table 4.1 Summary of trials included in the manufacturer's submission

Trial	Interventions			Key Issues
Catovsky 2005 [5]	Chl	F	FC	• Abstract • FC significantly improved PFS vs. F and Chl
CLL4 Patient level data [10]	Chl	F	FC	• Report (Additional data provided by manufacturer) • OR higher with F & FC vs. Chl. [redacted]
Eichhorst 2006 [3]		F	FC	• Fully published • FC significantly improved OR & PFS vs. F. No difference in OS between groups.
Eichhorst 2005 [6]	Chl	F		• Abstract • OR similar in both arms. OS & PFS significantly shorter in elderly patients.
Finn 2004 [7]		F	FC	• Abstract • FC significantly improved OR, CR & PFS vs. F.
Karlsson 2004 [8]	Chl	F		• Abstract • No significant difference in OR, CR, nPR or PR between groups.
Rai 2000 [4]	Chl	F	FChl	• Fully published • FC significantly improved OR, CR, PR & PFS vs. Chl. No difference in OS between groups.
Spriano 2000 [9]	Chl	F		• Abstract • FC significantly improved OR, CR, PR & PFS vs. Chl. No difference in OS between groups.

Table 4.2 summary of trials

Abbreviations key: AIHA: autoimmune haemolytic anaemia; **C**: cyclophosphamide; **CdA**: cladribine; **Chl**: chlorambucil; **CLL**: chronic lymphocytic leukaemia; **CR**: complete response; **CTC**: common toxicity criteria; **ECOG**: Eastern Cooperative Oncology Group **F**: fludarabine; **FC**: fludarabine plus cyclophosphamide; **ITT:** intention-to-treat; **NCI**: National Cancer Institute; **nPR**: nodular partial response; **NR**: no response; **O/E**: observed to expected ratio; **OR**: overall response; **OS**; overall survival; **PD**: progressive disease; **PFS**: progression-free survival; **PR**: partial response; **QoL**: quality of life; **RCT**: randomised control trial; **SD**: stable disease; **TFS**: treatment free survival.

Table 4.2.1 Summary of trial: Catovsky 2005 [5]

Reference	Design	Intervention	Inclusion criteria	Exclusion Criteria	Outcomes	Results	Adverse Effects
Catovsky D *et al*. [5]	This abstract is from the LRF CLL4 trial, a RCT comparing chlorambucil (Chl), fludarabine (F) and fludarabine plus cyclophosphamide (FC) and presents early results.	783 patients were randomised to either Chl (n=387), F (n=194) or FC (n=196) F & FC = min 3 months, max 6 (exceptionally those experiencing continuous response may receive for up to 8 months) Chl = until maximum response achieved (exceptionally, may continue for a few more months)	Not stated in abstract The trial protocol (Submission ref 31) states all patients with B-cell CLL, previously untreated who require treatment with Binet stage A progressive, stage B or stage C disease	Not stated in abstract The trial protocol states: Other life-threatening diseases, renal failure, hepatic enzymes and bilirubin >twice the upper limit of normal (unless due to CLL), pregnancy or risk of pregnancy, patients not expected to complete the study due to other reasons.	CR, nPR, PR and NR	Progression-free survival showed fewer events with FC (O/E 0.5) than F (O/E 1.1) and Chl (O/E 1.3); F+FC v Chl 0<0.00005; FC v F p<0.0005. PFS at 3 years was Chl 23%, F 31% and FC 62%. No difference in overall survival was demonstrated.	Neutropenia: FC 55%, F 40%, Chl 29% Haemolytic anaemias: Chl 13%, F 10%, FC 4% Nausea and vomiting and alopecia; increased frequency with FC than with other regimens.

Table 4.2.2 Summary of trial: CLL4 patient level data [10]

Reference	Design	Intervention	Inclusion criteria	Exclusion Criteria	Outcomes	Results	Adverse Effects
CLL4 Patient level data (2006) [10]	LRF CLL4 trial, a RCT comparing chlorambucil (Chl), fludarabine (F) and fludarabine plus cyclophosphamide (FC). Additional data provided by SHC.	783 patients were randomised to either Chl (n=387), F (n=194) or FC (n=196) F & FC = min 3 months, max 6 (exceptionally those experiencing continuous response may receive for up to 8 months) Chl = until maximum response achieved (exceptionally, may continue for a few more months)	Previously untreated B-cell CLL, diagnosed by persistent lymphocytosis (>10x 10^9/L) and bone marrow infiltration of at least 40% who require treatment with Binet stage A progressive, stage B or stage C disease.	Other life-threatening diseases, renal failure, hepatic enzymes and bilirubin >twice the upper limit of normal (unless due to CLL), pregnancy or risk thereof, patients not expected to complete, diagnosis other than CLL after central review of markers and morphology.	Overall survival (OS), Complete response (CR) rate, nodular partial response nPR), partial remission (PR) or no response (NR). Quality of life (EORTC QLQ-C30 questionnaire.	ITT (n=777) [redacted]; PFS not reported. Evaluable (n=720) [redacted]; PFS not reported. [redacted] at 45 months follow-up QoL results were the same for each treatment group at baseline and at 12 months	ITT: neutropenia: F [redacted] thrombocytopenia: F 11%, FC 15% & Chl 12%. Evaluable: neutropenia: [redacted] thrombocytopenia: [redacted] Nausea and vomiting and alopecia; increased frequency with FC than with other regimens

Table 4.2.3 Summary of trial: Eichhorst 2006 [3]

Reference	Design	Intervention	Inclusion criteria	Exclusion Criteria	Outcomes	Results	Adverse Effects
Eichhorst B F *et al*. [3]	RCT designed to determine whether fludarabine plus cyclophosphamide in combination (FC) was more effective than fludarabine alone (F). Not stated whether blinded.	375 patients randomly assigned to F (n=182) or FC (n=180); 13 excluded (due to inclusion criteria violations). F alone 25 mg/m^2 intravenously (iv) for 5 days. F 30 mg/m^2 iv plus cyclophosphamide 250 mg /m^2 iv for 3 days. Both regimens repeated every 28 days for a maximum of 6 courses. Median of 6 courses in each arm administered; 70.7% and 64% completed 6 courses of F and FC, respectively. 11 patients lost to follow-up.	Patients aged 18-65 years with previously untreated advanced CLL, ie patients in Binet stage C, stage B if they had rapid disease progression or symptoms or if they had B symptoms, and stage A if they had B symptoms. Patients had to have a life expectancy > 6 months and an ECOG performance status of 0, 1 or 2.	Severe organ dysfunction, concomitant or previous neoplasms, autoimmune haemolytic anaemia (AIHA), thrombocyto-penia.	CR, OR, OS, PFS and TFS. Not stated which is/are primary outcome(s).	CR: 23.8% with FC (39/164), 6.7% with F (11/164; p < 0.001). OR; 94.5% with FC (155/164), 82.9% with F (136/164; p < 0.001). OS: No significant difference between FC (n = 176) and F (n = 175). Median follow-up = 22 months (too short for OS to be reached). PFS: 48 months with FC (n = 168), 20 months with F (n = 171; p = 0.001). TFS: 37 months with FC (n = 175), 25 months with F (n = 169; p < 0.001)	Toxicity data available for 173 patients in each arm. Two treatment-related deaths (1.2%) in FC arm (one each of severe AIHA and tumour lysis syndrome) and 3 (1.7%) in F arm (one each of pneumonia with sepsis, cerebral bleeding due to thrombocytopenia and AIHA). CTC grades 3 or 4 toxicity occurred in 54.0% and 72.6% of patients in the F and FC arms, respectively (p = 0.001). Grade 3 & 4 myelotoxicity: F 39.3% vs. FC 64.2% (p = 0.001). Grade 3 & 4 leucocytopenia: F 26.0% vs. FC 55.5% (p < 0.001). Grade 3 & 4 GIT side effects: F 1.7% vs. FC 5.8% (p = 0.05).

Table 4.2.4 Summary of trial: Eichhorst 2005b [6]

Reference	Design	Intervention	Inclusion criteria	Exclusion Criteria	Outcomes	Results	Adverse Effects
Eichhorst BF *et al*. [6]	This study was a meta-analysis designed to assess the efficacy and toxicity of fludarabine when administered to younger and elderly patients, within two phase III trials. Neither of the phase III trials was fully published, nor was this meta-analysis. This study was published as an oral-session, therefore did not provide details on study design.	'Younger' patients (n=362), median age 59 years), were randomised to receive either fludarabine (n=182) or fludarabine plus cyclophosphamide (n=180) within the CLL4 trial. In the CLL5 protocol, 191 elderly patients (median age 71 years) received fludarabine (n=92) or chlorambucil (n=99).	Inclusion criteria were stated as being identical in both trials except for age limits, but no details given. All patients were previously untreated and in advanced stage Binet C or Binet B with symptoms which require therapy or Binet A with severe B-symptoms.	Not stated	Efficacy and toxicity were compared in the two patient groups, primary outcomes were not specifically stated but outcomes assessed were response rates, remission rates, progression-free survival and overall survival. Side effect incidence was assessed.	OR rate similar for both arms, 82.9% in the younger group and 85.7% in the elderly. The complete remission rate was 6.7% in the younger patients and 10.4% in the elderly (p=0.3). After 24 months follow up, the progression-free survival was significantly shorter in the elderly group (18.7 months) compared to 19.8 months in the younger group after 22 months observation time (p=0.03). Overall survival was lower in elderly patients as well (29 months versus median not reached, p<0.001).	3 treatment related deaths in each group due to infection or haemolysis. Side effect incidence was similar in both groups. Severe CTC grade 3 and 4, myelosuppression occurred in 39% of younger and 41% of elderly patients. There was no difference in the rate of leukocytopenia, thrombocytopenia or anaemia. The incidence and severity of infections was similar in both groups (24% vs 32% all and 8.7% vs 6.9% CTC grade 3 and 4). The incidence of second neoplasia was significantly higher in the elderly patients (2.2% vs 12.2%, p=0.001).

Table 4.2.5 Summary of trial: Finn 2004 [7]

Reference	Design	Intervention	Inclusion criteria	Exclusion Criteria	Outcomes	Results	Adverse Effects
Flinn I W et al. [7]	This study was a RCT and was designed to further evaluate the efficacy and toxicity of FC treatment versus F treatment alone.	A total of 278 patients were enrolled in the study with 141 patients assigned to the FC arm and 137 to the F arm.	No inclusion criteria are given except that patients had previously untreated CLL. The median age of patients was 62 years (34-86), and the median performance status was 1 (0 to 2). 70% of patients were male (194) and 30% were female (83). At study entry, 56% of cases were Rai stages, 0, 1, or 2, while 44% were in stages 3 or 4.	No exclusion criteria are given. Four patients declined to receive protocol treatment, including one who was later found to be ineligible. Five additional patients were also deemed ineligible.	Primary outcome not stated Outcomes measured: CR, PR, OR and PFS.	CR rates were 22.4% (28/125 patients) and 5.8% (7/121 patients) in the FC versus F treatment arms, respectively ($p = 0.0002$). PR rates were 48.0% (60/125 patients) in the FC arm and 43.8% (53/121 patients) in the F alone arm. Preliminary estimates of the median progression-free survival time are 41.0 months for the FC arm and 17.7 months for the F alone arm ($p < 0.001$).	There were two deaths due to infection with grade 3 or 4 neutropenia (one in each treatment arm). Grade 4 or higher non-haematologic toxicities were seen in 17% of FC patients and 13% in F regimen patients ($p = 0.48$). In terms of general infections 17% was seen in the FC regimen versus 11% in the F regimen ($p = 0.21$). There is no reference to any other forms of toxicity or adverse reactions.

Table 4.2.6 Summary of trial: Karlsson 2004 [8]

Reference	Design	Intervention	Inclusion criteria	Exclusion Criteria	Outcomes	Results	Adverse Effects
Karlsson. K. *et al*. [8] International Phase III-Trial in B-cell CLL – Protocol. [34]	This study was an RCT designed to compare cladribine (CdA) and fludarabine (F) as first-line treatment of symptomatic B cell chronic lymphocytic leukaemia (B-CLL) with high dose intermittent chlorambucil (Chl) as control.	A total of 150 patients were randomly assigned to CdA, F or Chl. Exact numbers not stated. 139 patients evaluable; CdA (47), Chl (47) or F (45).	Patients were assessed according to NCI criteria. No other details were reported in abstract. Study protocol states: CLL of B-cell origin. Binet stage B or C or progressive symptomatic stage A. judged to be in need of systemic therapy. Age 18-75 years. No previous chemotherapy or cytotoxic immune therapy.	Not reported in abstract. Study protocol states: On-going infection. Liver failure. Renal failure. Severer heart failure. ECOG >2. Other malignancy.	The primary endpoint was not reported in abstract. Outcome measures stated in the study protocol: OR, CR, PR & OS.	Results were analysed by ITT and quoted for CR/nPR and PR (complete response, nodular partial response and partial response). The respective values obtained for CR, nPR and PR were 4.2%, 6.4% & 47.0% for Chl, 0.0%, 4.4% & 62.2% for F, and 4.2%, 6.4% & 64% for CdA. OR to Chl, F and CdA were reported to be 57%, 67% and 74% respectively.	Grade 3/4 haematological toxicity was higher in the CdA group (36% and 11%) compared to the Chl and F groups (25% and 20% vs.24% and 5% respectively). Serious grade 3/5 infections were seen in 30%, 28% and 25% of CdA, F & Chl patients, respectively.

Table 4.2.7 Summary of trial: Rai 2000 [4]

Reference	Design	Intervention	Inclusion criteria	Exclusion Criteria	Outcomes	Results	Adverse Effects
Rai KR *et al*. [4]	This study was an RCT designed to compare the efficacy of fludarabine (F) with that of chlorambucil (Chl) and a combination regimen of fludarabine plus chlorambucil (FChl) in patients with previously untreated chronic lymphocytic leukaemia (CLL).	A total of 544 patients were randomly assigned to F (195), Chl (200) or FChl (149).	Previously untreated CLL, assessed according to NCI guidelines and the modified Rai staging system. All patients with high-risk (Rai stage III or IV) and Intermediate-risk Rai stage I or II patients if they had at least one of the following: weight loss, night sweats, extreme fatigue, lymphadenopathy, splenomegaly, hepatomegaly, >50% increase in lymphocytes over 2 months. Additional criteria: >18 years of age; ECOG performance status of 0-2; baseline liver or kidney function ≤1.5x upper limit of normal and a negative direct antiglobulin (Coombs) test.	Any previous treatment for CLL.	The primary endpoint was PFS. Secondary endpoints were response according to stage, OS, and safety.	The FChl group was stopped after interim analysis revealed excessive toxicity and a non-superior response rate to F. PFS: F 20 months vs. FC 14 months, p<0.001. OR: F 63%, FChl 61% and Chl 37%; CR: F 20%, FChl 20% and Chl 4%; PR: F 43%, FChl 41% and Chl 33% (p<0.001 for all comparisons) OS: F 66 months, Chl 56 months and FChl 55 months, p=0.21 (F vs. Chl, p=0.01).	Grade 3/4 neutropenia & infections were higher in the F group vs. Chl group (27% vs. 19%, p=0.007, and 16% vs. 9%, p=0.01, respectively). Overall the incidence of all grade 3/4 side effects was significantly greater with F than Chl (44% vs. 55%, p<0.001).

Table 4.2.8 Summary of trial: Spriano 2000 [9]

Reference	Design	Intervention	Inclusion criteria	Exclusion Criteria	Outcomes	Results	Adverse Effects
Spriano M et al. Haematology and Cell Therapy 2000; 42 (1:93)	Randomized, prospective, multi-centre study which compared the response rate and safety of fludarabine (F) against chlorambucil and prednisone (ChL + P).	A total of 150 patients were enrolled in the study with 75 patients assigned to the F arm (25 mg/m^2 by 30 minute i.v. infusion daily for 5 consecutive days every 4 weeks) and 75 to the ChL + P arm (30 mg/m^2 orally on days 1 and 15 plus prednisone 40 mg/m^2 i.m. on days 1-5 and 15-19 every 4 weeks).	No inclusion criteria are given except that patients had previously untreated active B-CLL with Rai intermediate or high-risk stages.	No exclusion criteria are given. Eight patients could not be evaluated.	Response rates (CR + PR) and response duration.	The response rate (CR + PR) was 71% (46 CR + 25 PR) in the fludarabine arm and 71% (37 CR + 34 PR) in the chlorambucil and prednisone arm. Refractory CLL was seen in 19% (10 SD + 9 PD) and 18% (11 SD + 7 PD) of patients respectively. Response duration was longer in the fludarabine arm (28 months versus 21 months; p = 0.007).	Toxicity was comparable in the two treatment groups; however no data are given and there is no reference to any specific forms of toxicity or adverse reactions.

4.2 Submission Trial Analysis

All studies included in the clinical evidence section of the SHC submission were subjected to a detailed critical appraisal. The resultant appraisals were then compared to the data presented in the submission.

4.2.1 Catovsky 2005 [5]

Trial summary

This is presented as a written abstract from an oral conference session presenting early results from the LRF CLL4 trial, although the trial protocol is available separately. Details relating to the population studied, the interventions and the primary outcome considered are limited in the abstract itself.

The trial compared the effects of F alone, FC in combination, and Chl alone on clinical response rates, progression-free survival and overall survival in patients. 783 patients were randomised with 6 exclusions but the abstract does not state the method of randomisation or the reason for exclusion. No detail is given around the balancing of groups with respect to population characteristics. Overall the male: female ratio was 2.8:1 and the distribution by Binet stages was A progressive 25%, B 45% and C 30%. One-third of cases were aged <60 years and one-third 70 years or over. No mention is made of the remaining patients' ages. A structured critical appraisal of this trial is presented in Appendix 3.

Analysis of PFS showed fewer events with FC (observed to expected ratio (O/E) 0.5) than F (O/E 1.1) and chlorambucil (O/E1.3). F + FC v Chl $p<0.00005$; FC vs F $p<0.0005$. PFS at 3 years was reported to be 23% Chl, 31% F and 62% FC. No difference is reported between the groups for OS. Furthermore, younger patients appear to benefit more from treatment across all treatment groups.

Important trial points

Key trial points are outlined below:

- No mention is made of blinding within the abstract; however, the protocol states that patients and clinicians were not blinded. Responses were made by bone marrow biopsy and it is unclear from the abstract whether interpretation of results was carried out by blinded staff.

- At the point of reporting these results, data from only 661 patients were available. No information is given as to the reasons for lack of data from the remaining 116 patients or the impact this may have had on results. A significantly higher proportion of

patients in the Chl arm (78/387 or 20%) did not have data available compared to F alone (18/194, 9%) or the FC combination (20/196, 10%). This may have the effect of underestimating the effectiveness of the control arm. No information is given with respect to analysis using an intention-to-treat basis.

- No detail is given of any power calculation in the abstract; however, this is included in the protocol. 500 patients (250 allocated to Chl and 250 to F based treatment) would provide more than 90% power to detect an absolute difference of 15%, from 40% to 55% in survival at 5 years using a 2-sided p-value. There would be 65% power to detect a difference of 10%. This would also be the power in detecting a difference between the FC and F arms.

- More neutropenia was reported with FC (55%) than F (40%) and Chl (29%) with an increased number of hospitalisations in the fludarabine-containing groups. Conversely, more haemolytic anaemias were reported in the Chl group (13%) compared with F (10%) and FC (4%). More nausea/vomiting and alopecia were reported with FC than other regimens although precise figures are not quoted..

- Confidence intervals are not reported although p values are for all F containing combinations vs. Chl and FC vs. F. It is difficult to draw firm conclusions from the data provided

- The population recruited into this trial appear to reflect the population in which the treatments would be used in the UK. 88% of patients were from the UK and the male/female and age balance would seem appropriate. It is not clear though whether the groups were balanced once outcomes were analysed and until these data are available it is not certain the results can be applied to the general population of patients with CLL.

Critique of the SHC submission

The majority of references to the CLL4 trial within the submission refer to the unpublished patient-level data obtained directly from the investigators. Where the abstract was referred to specifically, most quoted information was correct with the exception of:

- Throughout the tabulated data, percentages for composite response rates appear to have been calculated through summation of the individual response rates and may differ from response rates if calculated from the originating patient-level data.

- Throughout the tabulated data, 178 patients in the fludarabine group were stated to have data available. Within the abstract 176 patients were evaluable in this group. It is unlikely, however, that this would make any difference to interpretation.

- Q61. The submission referenced this abstract in support of the statement that significant improvements in PFS and time without treatment were associated with the increased quality of life observed in the fludarabine-containing regimens, but makes no reference to the impact of the increased hospitalisations within these groups reported within the abstract.

- Page 66 references this study in support of the statement 'whereas fludarabine response typically lasted about 20 months, patients treated with the FC combination have a significantly longer response to therapy without relapse of over 40 months'. However, the median follow up in this study is only 21 months.

Summary

The Catovsky *et al* trial is a relevant trial to be included in the submission, despite some submission inaccuracies and the abstract format. These preliminary findings from the CLL4 study showed that FC was associated with an improvement in PFS at 3 years compared to F and Chl. The data presented in this abstract are largely superseded by the additional patient-level data presented within the submission. However, follow-up is ongoing and, until this study is fully published and the complete data made available for evaluation, these results must be interpreted with caution.

4.2.2 CLL4 patient level data [10]

Trial summary

These are supplemental patient level data presented by the manufacturer as 'academic in confidence' alongside the abstract published by Catovsky *et al* [5] to support the health economic analyses. The trial compared the effects of fludarabine alone (F), fludarabine in combination with cyclophosphamide (FC) and chlorambucil (Chl) alone on clinical response rates, PFS and OS in patients. At a median follow-up of 45 months, overall complete

responses were higher in those patients receiving fludarabine (F or FC) than those receiving chlorambucil. [redacted].

Important trial points

Key trial points are outlined below:

- The report states that patients and clinicians were not blinded. Randomisation was stated to be secure and treatment was allocated by computer, balancing treatment with groups by age (<60, 60-69/ 70+). Responses were assessed by bone marrow biopsy, but it is unclear from the abstract and study protocol whether interpretation of results was carried out by blinded staff.

- No detail is given of any power calculation in the report; however, this is included in the study protocol.[35] Five hundred patients (250 allocated to Chl and 250 to F based treatment) would provide more than 90% power to detect an absolute difference of 15%, from 40% to 55% in survival at 5 years using a 2-sided p-value. There would be 65% power to detect a difference of 10%. The power in detecting a 15% difference between the FC and F arms is also 65%.

- At the point of reporting these results, data from only 720 patients of the 777 randomised were evaluable for the economic analysis. However, data are also presented on an ITT basis.

- ITT analyses show that, at 45 months median duration of follow-up, OR and CR, were higher in those patients treated with F ([redacted] respectively) and FC ([redacted]) than those treated with Chl ([redacted]). PR was higher in those receiving F alone ([redacted]) than those receiving Chl ([redacted]) or FC ([redacted]). [redacted].

- More neutropenia and thrombocytopenia (all grades) was reported with FC ([redacted] & 15%) than F ([redacted] & 11%) and Chl ([redacted] & 12%) (ITT analyses). Conversely, more haemolytic anaemias (grade not stated) were reported in the Chl group [redacted] compared with [redacted] and FC [redacted]. More non-haematological toxicity (grade not stated) was reported with F and FC ([redacted], respectively) than with Chl ([redacted]).

- The sample recruited into this trial appears to reflect the population in which the treatments would be used in the UK. Eighty eight percent of patients were from the UK and the male/female and age balance would seem appropriate.

Critique of the SHC submission

Although the majority of the supplemental information contained in the report regarding the CLL4 study methods agrees with that outlined in the published study protocol, [35] the additional outcome data presented within the submission are not in the public domain and, therefore, cannot be verified externally. However, the outcome data are largely in accordance with early results of CLL4 presented in the Catovsky *et al* abstract.[5] Nonetheless, follow-up is ongoing and, until this study is fully published and the complete data made available for evaluation, these results must be interpreted with caution.

4.2.3 Eichhorst 2006 [3]

Trial summary

This trial compared the effects of fludarabine alone (F) and FC (fludarabine in combination with cyclophosphamide) on overall survival (OS) and clinical response rates in patients younger than 66 years with previously untreated advanced CLL. A structured critical appraisal of this trial is presented in Appendix 4.

The published trial data showed that the complete remission rate was significantly higher with FC (23.8%) than F (6.7%). The overall (complete remission plus partial remission) rate was also significantly higher with FC (94.5%) than F (82.9%). OS did not differ significantly between the two groups. However, the median follow-up duration was 22 months, too short for this end point to be reached. FC treatment also resulted in longer median PFS (48 vs. 20 months; P =.001) and longer treatment-free survival (37 vs. 25 months; P < .001).

Important trial points

Key trial points are outlined below:

- The dose of fludarabine used in the F arm was the licensed dose (25 mg/m^2 intravenously for five days repeated every 28 days for a maximum of six courses). The FC regimen comprised fludarabine 30 mg/m^2 plus cyclophosphamide 250 mg/m^2 (both iv daily for three days). The F dose is higher than that used in the FC combination regimen in the CLL4 study. The Fludara® SPC does not state doses for the fludarabine and cyclophosphamide combination regimen.

- Overall, the trial design was appropriate and all participants appear to have been followed up and data collected in the same way. The baseline characteristics of the patients in the two arms of this trial look similar, and the authors state that comparison indicated no significant difference regarding the main clinical features

and risk categories. However, no details of statistical comparison of the two groups are presented and no details of the randomisation method used are given.

- Of the 375 patients originally randomised, 13 were excluded and of the remaining 362, 182 were assigned to F and 180 to FC. Eleven were lost to follow-up, leaving 351. Survival data were available for 351 patients, response data for 328 and toxicity data for 346. Response data were available for 164 patients from each treatment arm and response rates are presented as percentages of these 164, rather than as percentages of the 182 and 180 randomised to F and FC, respectively. The authors state that response rates were calculated for all patients who received at least one cycle of therapy, which may account for this.

- The complete remission rate was significantly higher with FC (23.8%) than F (6.7%; $p < 0.001$), as was the overall (complete remission plus partial response) rate (FC 94.5% with FC vs. 82.9% with F ($p = 0.001$)). OS in the two groups did not differ significantly. However, the median follow-up duration was 22 months, too short for this end point to be reached. Progression-free survival was significantly longer with FC than F (48 vs. 20 months; $p = 0.001$), as was treatment-free survival (37 vs. 25 months; $p < 0.001$). It is not clear which of these outcomes is/are primary or secondary end points.

- The incidence of grade 3 and 4 toxicities, particularly myelotoxicity, was significantly higher with FC than F.

Critique of the SHC submission

No detailed structured appraisal of this trial is presented in the submission. The trial data are presented in Table 13, in the form of answers to questions 46 to 53. The majority of the answers provided in the submission are accurate, with the following slight differences:

- The submission states that a secure randomisation method was used, where the randomisation sequence was kept away from the clinical area and administered by staff not directly involved with clinical care. This might have been the case, but it is not possible to be sure from the information provided in the paper. The authors state that randomisation was performed by the Institute of Medical Statistics and Epidemiology, Technical University, Munich, Germany, but give no details of the randomisation method used. The authors do not specifically state that the

randomisation sequence was kept away from the clinical area and administered by staff not directly involved in patient care.

- The trial was conducted in Germany. The submission states that there is no clinical difference in clinical practice or patients compared to intended population in the UK.

- With respect to how the subjects included in the trial compare with patients who are likely to receive the drug in the UK, the submission states: no difference in patients compared to intended population in UK. However, the patients included in this study were younger patients (range 42 – 64 years) and, therefore, may not be an accurate representation of the intended UK population.

- The dosage regimens used are discussed above (Important trial points).

Summary

The Eichhorst *et al* 2006 study was generally well conducted and its inclusion in the submission was justified. The FC regimen resulted in superior OR, CR and longer PFS and TFS, but OS did not differ with the F and FC regimens. However, the F dose used in FC regimen was higher than that used in the combination regimen in the CLL4 study, on which the economic evaluation is based and which reflects current practice.[13]

4.2.4 Eichhorst 2005b [6]

Trial summary

This trial compared the efficacy and toxicity of fludarabine, when administered to younger (median age 59 years) and elderly patients (median age 71 years). It was published as an oral session and was a meta-analysis of two phase III trials of the German CLL Study Group (GCLLSG). It appears that the meta-analysis was done on abstracts of the phase III trials. A structured critical appraisal of this trial is presented in Appendix 5.

Important trial points

Key trial points are outlined below:

- Data are presented as an abstract of an oral session only and details of the trial protocol are not available.
- The meta-analysis was performed on abstracts of trials, as opposed to full published data. It is questionable how these trials can be compared when only limited information is available.

- The dosage of fludarabine administered in both trials was 30mg/m^2, which is higher than the licensed dose.

- The OR rate was similar in both arms, 82.9% in the younger group and 85.7% in the elderly. The CR rate was 6.7% in the younger patients and 10.4% in the elderly (p= 0.3).

- After a follow-up time of 24 months (mo), PFS was significantly shorter in the elderly group, with 18.7 mo compared to 19.8 mo in the younger group after 22 mo observation time (p=0.03).

- OS was significantly impaired in elderly patients (29 mo versus median not reached, p<0.001).

Critique of the SHC submission

The majority of the submission was accurate according to the data in the published abstract, and was a fair interpretation of the trial. Specific points include;

- The submission refers to this trial (on page 63) involving the EORTC-QoL-C30 questionnaire, but there appears to be no mention of it in the trial. As this was a meta-analysis, it is possible that this questionnaire was referred to in the two individual arms but, as they were presented as abstracts only, this is unclear.

- On page 64, the submission states that fludarabine is equally effective in younger and in older patients, but this is not strictly true. The trial showed overall response rates were similar, 82.9% in the younger group and 85.7% in the elderly. After 24 months follow up, the progression-free survival was significantly shorter in the elderly group (18.7 months) compared to 19.8 months in the younger group after 22 months observation time (p=0.03). Overall survival was lower in elderly patients as well (29 months versus median not reached, p<0.001). It is difficult to tell whether the figures in the table are correct as some of the data are not quoted in the abstract (Chl – n=99 Result NR).
- No reference is made to this meta-analysis in section 2.6 of the submission – Meta Analysis.

Summary

There is very little useful information that can be gleaned from this trial, mainly due to the fact that it is presented as an oral session and consists of a meta-analysis of unpublished trials.

4.2.5 Flinn 2004 [7]

Trial summary

This trial compared the efficacy and toxicity of fludarabine and cyclophosphamide (FC) treatment versus fludarabine monotherapy in patients with previously untreated CLL. A structured critical appraisal of this trial is presented in Appendix 6.

The interim analysis of this intergroup trial showed that FC treatment significantly increases both CR and OR compared to F in untreated CLL patients. Preliminary estimates of the median PFS were also significantly increased following FC treatment compared with F. There was no significant difference in the incidence of non-haematologic toxicity or infections.

Important trial points

Key trial points are outlined below:

- Data are presented in abstract format only and details of the trial protocol are not available.

- No comparison is made to the current recommended first-line treatment chlorambucil.

- The dose of cyclophosphamide is higher than the SPC recommendation and doses used in other trials (600mg/m^2 vs 250 mg/m^2).

- Not all patients are clearly accounted for at the conclusion of the trial.

- At the planned interim analysis (76% information), CR and OR were superior for patients receiving FC compared with patients receiving F treatment (22.4% vs 5.8% p=0.0002, and 70.4% vs 49.6% p=0.001).

- PFS 41.0 months for FC compared to 17.7 months for F (p<0.001).

- Non-haematologic toxicity and infections were more frequent but not significantly different in patients receiving FC compared with patients receiving F treatment (17% vs 13% p=0.48, and 17% vs 11% p=0.21).

Critique of the SHC submission

The majority of the submission was accurate according to the data in the published abstract, and was a fair interpretation of the trial. Specific points include:

- The submission states on page 14 that this trial 'confirms that fludarabine, and particularly FC, offer significantly superior outcomes compared to chlorambucil'. This is incorrect as this trial makes no comparison to chlorambucil treatment. This statement is repeated on page 16.

- The submission states on page 64 that 'the lower overall response rate in the Flinn study is postulated by Eichhorst *et al* (1) to be due to a higher proportion of elderly or high-risk patients in the study'. This statement is too strong as the population characteristics in the Flinn study are very similar to those used in the Eichhorst study (median age 62 vs 59, Rai stage 0-2 56% vs 56.6%, and Rai stage 3-4 44% vs 41.0%, respectively).

Summary

The intergroup E2997 trial is a relevant trial to be included in the submission, despite some submission inaccuracies and the abstract format. The trial showed significant improvement in CR and OR following FC treatment compared to F treatment without signs of increased toxicity.

4.2.6 Karlsson 2004 [8]

Trial summary

This trial was designed to compare the efficacy of cladribine (CdA, another purine analogue) and fludarabine (F) as first-line treatment of symptomatic B-CLL, with high dose chlorambucil (Chl) as a control. A structured critical appraisal of this trial is presented in Appendix 7.

This interim analysis published in abstract form only showed no significant differences in response rates, haematological toxicity or serious infection between the treatment groups. Autoimmune haemolysis was seen only with purine analogues and was similar for the two therapies.

Important trial points

Key trial points are outlined below:

- The dose of Chl used in this study is higher than the UK standard (100mg/m^2 compared to 70mg/m^2), this is due to patients receiving Chl for 10 days instead of the UK standard of 7. F was given as per SPC.

- Overall responses were documented in 57%, 67% & 74% of Chl, F & CdA patients, respectively (Chl vs. CdA, p=0.06). Although no significant difference in response rates was observed between the therapies, the data suggest that CdA may be as least as effective as F.

- CR, nPR and PR for F were 4.2%, 6.4% and 47.0%; for Chl 0.0%, 4.4% and 62.2%; for CdA 4.2%, 6.4% and 64%, respectively. PFS, OS and duration of follow-up are not reported.

- Grade 3 & 4 infections were seen in 25%, 28% & 30% of Chl, F & CdA patients, respectively.

- Autoimmune haemolysis occurred in 4 patients during F, and 3 during CdA treatment.

Critique of the SHC submission

Although the submission makes little reference to this trial, the data presented therein were accurate according to the data in the published abstract, and were a fair interpretation of the trial. The submission included additional information derived from the study protocol.[34] Specific points include:

- Table 13 states ‘where?’ as the location of the study. The published abstract shows patients to be from Scandinavia, Australia and the UK.

Summary

The Karlsson trial is a relevant study to include in the submission; however, further patient level data are necessary to accurately interpret this study. The trial showed equivalent response rates for CdA, F and high-dose Chl. Toxicity and infections did not differ between arms.

4.2.7 Rai et al study [4]

Trial summary

This trial was designed to compare the efficacy of fludarabine (F) with that of chlorambucil (Chl) and a combination regimen of fludarabine plus chlorambucil (FChl) in patients with previously untreated chronic lymphocytic leukaemia (CLL). A structured critical appraisal of this trial is presented in Appendix 8.

This fully published trial showed that, when used as initial treatment for CLL, F is associated with a significantly longer PFS and higher response rates than those treated with Chl alone. However, there was no significant advantage to combination treatment over F alone in terms of response, and there was no statistically significant difference in OS among the three treatment groups. The benefits in PFS and response appear to be offset by the significantly higher overall incidence of grade 3 or 4 side effects, especially neutropenia and infections in the F group compared to the Chl group.

Important trial points

Key aspects of the trial are summarised in the following points:

- The dose of Chl used in this study is lower than the UK standard (40mg/m^2 compared to 70mg/m^2).

- The study was generally well conducted. Follow-up was adequate and there was no imbalance among the three treatment groups with respect to baseline characteristics, clinical features or risk categories.

- The study was open to detection bias through lack of blinding. However, the endpoints were clearly measurable and centralised review was required for all specimens from all patients who had a complete remission.

- Assignment to the F plus Chl group was stopped after a planned interim analysis revealed excessive toxicity and a non-superior response rate to F alone.

- There was a significantly longer median time to disease progression among the patients treated with F compared to those receiving Chl (20 months vs. 14 months, respectively, $p<0.001$).

- OR was significantly higher in those patients treated with F (63% for those receiving F alone and 61% for those receiving F plus Chl) than those treated with Chl alone (37%, p<0.001 for both comparisons). CR was F 20%, Chl 4% and FChl 20%; PR was F 43%, Chl 33% and FChl 41%. There was no significant advantage to combination treatment over F alone in terms of response.

- The study has limited power to detect a statistically significant difference in OS among the three treatment groups (F 66 months, Chl 56 months and FChl 55 months, p=0.21), or between the F and Chl groups (p=0.10).

- Some outcomes are missing from patients initially randomised to receive FChl.

- The median number of F cycles administered and the median number of cycles needed to induce a CR are not stated.

- Non-responding patients or those relapsing within <6 months were allowed to crossover; therefore, there is the possibility of a carry-over effect.

- No information is available regarding the ratio of benefit to harm in non-responders.

Critique of the SHC submission

The majority of the data included in the submission was a fair and accurate representation of the data presented in the published study.

- The typical inclusion and exclusion criteria outlined in Table 9 do not accurately reflect those applied to this study. For transparency, the inclusion and exclusion criteria should be outlined separately for each study.

- Table 11 states transfusion requirements, incidence of infection and performance status as quality of life outcomes. No QoL data are reported in the published study.

Summary

The Rai *et al* study was generally well conducted and its inclusion in the submission was justified. The study showed that treatment with F is associated with improvements in PFS

and response rates compared to Chl alone. In terms of response FChl was not superior to F alone.

4.2.8 Spriano 2000 [9]

Trial summary

This study compared the response rate and safety of fludarabine versus standard therapy with chlorambucil and prednisone in patients with previously untreated active B-CLL. A structured critical appraisal of this trial is presented in Appendix 9.

The published abstract data showed that response rates (CR + PR) were similar in patients treated with either fludarabine or with chlorambucil and prednisone. However, response duration was significantly prolonged in patients who received fludarabine compared to standard therapy. Toxicity was comparable between the two treatment groups.

Important trial points

Key trial points are outlined below:

- Data are presented in abstract format only and details of the trial protocol are not available.

- No information is given as to why chlorambucil and prednisone was chosen as standard therapy and not chlorambucil alone.

- No information regarding treatment group characteristics and thus prognostic indicators is given.

- The response rate (CR + PR) was 71% in the fludarabine arm and 71% in the chlorambucil and prednisone arm.

- Response duration was longer following fludarabine treatment (28 vs. 21 months, p=0.007).

- Toxicity is stated as comparable between the treatment groups but no data on how toxicity was measured, frequency of toxicity and which symptoms were reported is given in the abstract.

Critique of the SHC submission

Although the submission has little reference to this trial, the majority was accurate according to the data in the published abstract, and was a fair interpretation of the trial. Specific points include:

- It is only highlighted once throughout the submission (page 63) that data from this trial (and indeed others) are only in abstract format and should be interpreted with caution. This is not sufficient considering the lack of detailed trial information in the abstract.

- The submission does not attempt to explain why this trial shows that fludarabine has no additional benefit in response rate compared with chlorambucil and prednisone. As this is the decision problem, so this issue should have been addressed in the submission.

Summary

The Spriano trial is a relevant study to include in the submission; however, further details regarding trial protocol and patient characteristics are necessary to accurately interpret the data. The trial showed equivalent response rates following both fludarabine treatment and chlorambucil and prednisone treatment. Response duration was prolonged following fludarabine treatment; however, further investigation is needed to confirm these results.

4.3 Studies excluded from the submission

The manufacturer's submission stated that their search identified two trials which compared fludarabine with CHOP and CAP.[36, 37] The reason given for exclusion of these trials is that these regimens are not relevant to the manufacturer's decision problem. A further study was excluded because it did not report response rates according to NCI criteria.[38] In this study response was defined according to total tumour mass (TTM) reduction, an evaluation not used in any other.

4.3.1 Johnson et al *[36]*

Trial summary

This trial analysed how fludarabine compared to cyclophosphamide, doxorubicin and prednisolone (CAP) with regards to efficacy and adverse effects in patients with advanced CLL. This study did show that fludarabine gave a statistically significant improvement in overall response rate compared to CAP in patients with advanced CLL; however, the trial was carried out 16 years ago, and whether CAP is a relevant comparator today may affect

the applicability and relevance of the results. In untreated patients, the ORR was not statistically significant, but the trial was not adequately powered to show a benefit in this sub-group.

Important trial points

Key trial points are outlined below:

- 100 patients were previously untreated, but 96 had received prior therapy.
- A statistically significant response rate was only seen in pre-treated cases.
- In previously pre-treated patients, remission duration and survival did not differ between treatment groups. In previously untreated patients, fludarabine induced significantly longer remission than CAP.
- This trial was published 10 years ago and it could be questioned whether CAP is a relevant comparator today.

Critique of the SHC submission

This trial was excluded from the submission. The explanation for the omission of this trial was stated as being: 'The search identified RCTs comparing fludarabine with other interventions e.g. CHOP, CAP. Since these are not relevant to our decision problem, these studies were excluded at this stage'.

This explanation rests on the fundamental decision on the 'appropriateness' of CAP as a comparator and, if it is not routinely used in CLL patients in the UK, then it was appropriate for the trial to not be considered. Expert advice given to the ERG confirms that CAP is not widely used in the UK as first-line treatment for CLL. Its exclusion from the submission was, therefore, deemed to be appropriate, although a full consideration of this treatment and its relevance to the decision problem was judged to be outside the scope of the ERG report.

4.3.2 Leporrier et al *[37]*

Trial summary

This trial compared the efficacies of two anthracycline-containing regimens (CHOP and CAP) and fludarabine as first-line treatments for patients aged less than 75 years with previously untreated Binet stage B or C CLL. Nine-hundred-and-thirty-eight patients were randomised

to CHOP (n = 357, of whom 346 received it), CAP (n = 240, of whom 232 received it) or fludarabine (n = 341, of whom 336 received it).
The published trial data showed no significant differences among the overall survival rates (primary end point) of the three treatment arms. The clinical and overall remission rates (secondary end points) in the in the CAP group were lower than those in the CHOP and fludarabine groups, which were similar. Time to disease progression (secondary end point) showed no significant differences among the three groups.

Important trial points

Key trial points are outlined below:

- The dose of fludarabine used was the licensed dose (25 mg/m^2 intravenously for five days repeated monthly for a maximum of six courses). The CHOP regimen comprised: vincristine intravenously (iv) 1 mg/m^2 and doxorubicin iv 25mg/mg^2 on day 1 plus cyclophosphamide orally (po) 300 mg/m^2 and prednisone po 40 mg/m^2 from day 1 to 5. The CAP regimen comprised: cyclophosphamide iv 750 mg/m^2 from day 1 to 5, doxorubicin iv 50 mg/m^2 on day 1 plus prednisone po 40 mg/m^2 from day 1 to 5 or F iv 25 mg/m^2 daily for five days. CHOP and CAP were also repeated monthly for a total of six courses.

- Recruitment to the CAP arm was stopped early because the first interim analysis showed that survival and response rates were lower with CAP than the other two regimens.

- The median follow-up duration was 70 months. The respective overall survival durations in the CHOP, CAP and fludarabine groups were 67, 70 and 69 months (no significant differences among the groups).

- The CAP group showed the lowest clinical remission rate (15.2%) compared with 29.6% with CHOP and 40.1% with fludarabine ($p = 0.003$). The CAP group also showed the lowest overall remission rate (58.2%) compared with 71.5% with CHOP and 71.1% with CAP ($p < 0.001$ for each).

- Comparison of adverse effects of fludarabine with those of CHOP and CAP together showed that the rates of grade 3 myelotoxicity were higher with F than CAP and CHOP, whereas the converse applied to alopecia and grade 3 nausea and vomiting.

Critique of the SHC submission

This trial was excluded from the submission. The explanation for the omission of this trial was stated as being; 'The search identified RCTs comparing fludarabine with other interventions e.g. CHOP, CAP. Since these are not relevant to our decision problem, these studies were excluded at this stage. This resulted in two studies being excluded.'

This explanation rests on the fundamental decision on the 'appropriateness' of CAP and CHOP as a comparator and if it is not routinely used in CLL patients in the UK, then it was appropriate for the trial to not be considered. Expert advice given to the ERG confirms that CAP and CHOP are not widely used in the UK as first-line treatments for CLL. Its exclusion from the submission was, therefore, deemed to be appropriate, although a full consideration of this treatment and its relevance to the decision problem was judged to be outside the scope of the ERG report.

4.3.3 Jaksic [38]

Trial summary

This trial compared the efficacy of fludarabine with high-dose chlorambucil in previously untreated advanced B-CLL patients. The treatment period was 18 weeks, which corresponds to six 3-week cycles of fludarabine.

The data, which are only available in three abstracts at median follow-up of 32, 33 and 48 months, do not indicate that fludarabine is significantly different to high-dose chlorambucil with respect to the objectively measured response rate based on total tumour mass (p=0.295 at 33 months). The analysis does not appear to have been carried out on an intention-to-treat basis with the trial population consisting of 88 randomized patients plus 25 non-randomized control-group patients. Analysis was carried out on 98 patients but no further details as to the make up of that group are provided.

- At 48 months (n=98) CR was 47% and 33% and PR was 41% and 42% in Chl and F groups respectively. No p values are provided. In the non-randomized Chl group CR was 57% and PR was 29%. It is not stated whether the overall results include the non-randomized group results.

- Overall survival was 59% and median progression-free survival was 2 years.

- Haematological toxicity was significantly higher in the Chl group, due to the toxicity tailored schedule, without translating into a higher clinical toxicity.

- Infections occurred significantly more often in the F group vs. the Chl group (46% vs. 23%, respectively: p=0.05).

Critique of the SHC submission

This trial was excluded from the submission on the basis that response was measured by TTM reduction. As this outcome measure was not used in any other trial it would not have been possible to compare the outcome with any of the included studies. Its exclusion from the submission was, therefore, deemed to be appropriate by the ERG. However, toxicity data from this study provides additional data on the comparative safety profiles of F and Chl.

4.4 Other relevant studies

No other relevant studies were identified by the ERG during a comprehensive literature search. See Appendix 2 for search strategy.

4.5 Relevant ongoing studies

The following databases were searched for current research: Current Controlled Trials register (searched across multiple registers, including, ISRCTN, MRC NHS, and the National Institutes of Health registers), IFPMA, proceedings of the American Society for Clinical Oncology, National Research Register and the National Cancer Institute, British Society for Haematology, Leukaemia Research Fund, Scirus and a general web search using Google.

Other than more complete and fully published results of the studies included in the submission (principally CLL4 & CLL5), the following studies were identified as relevant ongoing trials that are likely to provide significant additional evidence within the next 6-12 months.

- CLL6: A randomized phase III study in previously untreated patients with biological high-risk CLL: fludarabine and cyclophosphamide (FC) versus FC and low-dose alemtuzumab. **ISRCTN25180151**.

- CAM307: Phase III study to evaluate the efficacy and safety of front-line therapy with alemtuzumab (Campath®) vs. chlorambucil in patients with progressive B-cell chronic lymphocytic leukaemia (B-CLL). **NCT00046683**.

4.6 Meta-analyses

Of the seven studies included in the submission, only two were fully published [3, 4] and the remaining five studies were available in abstract form only.[5-9] These abstracts are unlikely to have been subject to peer-review and there are insufficient data in terms of the methods and results presented to allow for their inclusion in a robust meta-analysis. Of the two fully published studies one compares F with FC [3] whilst the other compares F with Chl.[4] Therefore, pooling of data would not add further insight to the decision problem. A meta-analysis has been undertaken by Zhu *et al* comparing F with alkylator-based regimens.[39] This study concluded that F as an induction agent for patients with CLL yields a better clinical response with acceptable toxicity when compared with alkylator-based combination therapy (Chl with or without prednisone, and CHOP), but without a survival benefit by 5-6 years follow-up. However, of the five studies chosen for inclusion, three were abstracts and there were notable inconsistencies in the approach to the studies selected to analyse each outcome.[40]

4.7 Review of international guidelines submitted in part C of the SHC submission

Current guidelines issued on behalf of the British Committee for Standards in Haematology (BCSH) [15] suggest that, for the majority of patients, who are ineligible for a transplant procedure and in whom there is no contraindication to fludarabine, entry into the MRC CLL4 study should be offered. Both fludarabine and chlorambucil are options for patients who do not wish to enter the study. However, this study has now closed to recruitment and it is anticipated that revised guidelines will be issued to reflect this.

The details of these guidelines are attached in Appendix 10.

Chapter 5
Economic Evaluation

5.1 Introduction

This section provides a structured critique of the cost-effectiveness evidence submitted by Schering Health Care Ltd ('the manufacturer') by the ERG. As part of the STA process, manufacturers are expected to perform a systematic review of existing cost-effectiveness evidence for the health care technology or process being assessed. Where there is no existing evidence or the existing evidence is insufficient, manufacturers may perform their own cost-effectiveness analysis.

The manufacturer's economic submission to NICE includes:

(i) a narrative description of the cost effective literature relating to the use of fludarabine (F and FC) in patients with CLL (2 previous studies identified, p79-81 of the manufacturer's submission);

(ii) a short report on the quality of life literature available to inform the decision problem (13 references identified, Table 37 p109);

(iii) a report on a resource and cost audit of the CLL4 trial undertaken to provide data on resource use to inform the decision problem (Appendix 9);

(iv) a report on the economic evaluation undertaken and presented specifically for the NICE STA process (in particular Figure 3, p91 the schematic of the model and Tables 33,34 and 35 p93-94 providing information on the inputs used in the base-case and their sources); and

(v) an Excel-based model comprising the manufacturer's economic model.

Following an initial list of questions posed by the ERG to the manufacturers, an addendum was submitted with a revised analysis. The manufacturer's addendum ("Additional analysis to support QB10 and B12") includes:

(i) an explanation of the errors in the original submission and the amendments that have been made to the model based on queries raised by the ERG (Response B10, p3);

(ii) base-case costs and effectiveness results from the revised model (Tables 1 and 2, p4);

(iii) probabilistic sensitivity analysis results from the revised model (Table 3, p4 and Figures 1-3, p5-6);

(iv) sub-group analysis from the revised model (Tables 4 to 7 p6-7);

(v) results from the deterministic one-way sensitivity analysis on the new model (Table 8, p8-9); and

(vi) Tornado diagrams showing the cost per QALY results from the one-way sensitivity analysis (p10-11).

This section focuses on the economic evidence (including the main submission and the addendum) submitted by the manufacturer. The submission is critiqued on the basis of the manufacturer's report and by direct examination of the electronic model. The critical appraisal is conducted with the aid of a checklist for assessing the quality of economic evaluations [41] and a narrative review to highlight key assumptions and possible limitations. A range of key uncertainties is highlighted and additional work undertaken by the ERG to address several of these uncertainties and to explore the robustness of the submitted model is presented in Chapter 6.

5.2 Existing cost-effectiveness evidence

This section provides a narrative overview of existing published cost-effectiveness evidence reported in the manufacturer's submission. The searches undertaken by the manufacturers were replicated by the ERG in order to validate the evidence base considered by the manufacturers.

5.2.1 Literature searches

The search strategy used to identify cost-effectiveness studies is reported in the manufacturer's submission in Appendix 2, p7. In order to validate this component the ERG has undertaken its own searches detailed in Appendix 11 of this report. The searches undertaken by the ERG did not identify any additional studies which were not already considered in the manufacturer's submission.

5.2.2 Description of identified studies

Two papers were identified in both the manufacturer's submission and the ERG searches which reported on the cost-effectiveness of fludarabine monotherapy in comparison to chlorambucil in the management of CLL in previously untreated patients. These were a West Midlands HTA Collaboration (WMHTAC) report by Hancock *et al* (2002)[12] and an earlier Development and Evaluation Committee (DEC) report (1995).[11]

A brief overview of these studies is provided below.

5.2.3 The WMHTAC report

The report by Hancock *et al* (2002)[12] comprised a systematic review of effectiveness and cost-effectiveness studies and a decision-analytic model to assess the cost-effectiveness of fludabarine (F) compared to chlorambucil (Chl). At the time of this review only one trial was identified which compared F and Chl. No previously published economic evaluations were identified for the use of fludarabine in the first-line treatment of CLL.

In the absence of existing evidence on the cost-effectiveness of fludarabine, a simple decision analytic model was developed comprising three health states (death, survival with disease progression, and progression-free survival) to compare F and Chl over a three year time frame. It was estimated that treatment with F would result in a gain of 0.08 QALYs at an extra cost of £3,830, when compared to Chl, resulting in an incremental cost per QALY of £48,000. One-way deterministic sensitivity analysis was undertaken and this found that the cost-effectiveness estimates were not particularly robust to alternative assumptions.

The authors of the report recommended that the use of F as a first-line therapy for CLL was borderline because of the high cost per QALY. However, for the purpose of informing this STA, there are a number of key limitations of this study, including: (i) the information on effectiveness and costs was limited, e.g. effectiveness evidence was only derived from a single trial and the cost analysis did not consider costs arising from adverse events in chlorambucil treated patients; (ii) the cost estimates for fludarabine in the first-line setting are based on the use of fludarabine at second-line and may, therefore, not be an accurate reflection of the true cost; and (iii) the utility values used are not based on an accepted preference-based instrument. Finally, the choice of comparators was constrained to fludabarine monotherapy and chlorambucil and did not include other relevant comparators pertinent to this STA.

5.2.4 Wessex DEC report

This report, published in 1995 by The Development and Evaluation Committee (DEC) of the South and West Regional Health Authority[11], considered the economic impact of the use of fludarabine as a first-line treatment for CLL. The report compares the use of fludarabine monotherapy (F) to the use of chlorambucil plus prednisone (Chl + P).

The single effect of treatment incorporated into the QALY calculations was an increase in time free of progressive disease after first-line treatment (32 months for fludarabine compared to 24 months for chlorambucil plus prednisone). Based on utility scores of 0.96 for remission state and 0.81 for progressive disease state, and estimated percentages of

patients achieving remission of 74% for F and 77% for Chl + P, QALY gains of 0.29 were suggested for F and 0.23 for Chl + P compared to no treatment. The results are summarised in Table 5.1.

Table 5.1 Cost-effectiveness results of Wessex DEC report

	F compared to no treatment	Chl+P compared to no treatment	F compared to Chl+P (calculated by ERG)
Cost of Therapy	£6810	£410	£6400
Health gain	0.29 QALYs	0.23 QALYs	0.06 QALYs
Cost per QALY	£23,480	£1,780	£106,667

The report concluded that it could not be proven that fludarabine was more cost effective than first-line therapy with chlorambucil plus prednisone. However, there were several key weaknesses of the report, including: (i) the effectiveness data were drawn from case studies or was derived using data from single arms of studies against a range of different comparators; (ii) the basis of utility estimates is unclear and may not be preference-based; (iii) the study did not report the incremental cost-effectiveness of fludarabine compared to chlorambucil plus prednisone (however the ERG has made the calculation itself and provided it in the table above); (iv) a standard duration of treatment was used and no account was taken to response (i.e. in clinical practice patients receive treatment until the best response is achieved and this varies among patients); and (v) the choice of comparators was again limited.

5.2.5. Conclusion of previous economic evaluations

The ERG concurs with the conclusions reached by the manufacturer in their appraisal of the existing cost-effectiveness evidence. The current literature on the cost-effectiveness of fludarabine as a first-line treatment for CLL is very limited. Although both studies were constructed from an appropriate UK perspective, the applicability of their results is limited due to the scant clinical and economic evidence on which the studies were based (mainly due to the limited evidence available at the time these studies were undertaken) and the restricted range of comparators considered. In particular, no previously published studies have examined the cost-effectiveness of fludarabine combined with cyclophosphamide as a first-line treatment for CLL. Consequently the '*de novo*' submission by the manufacturer comprises the most relevant evidence to consider for the purposes of this STA.

5.3 Overview of manufacturer's economic evaluation

The manufacturer's submission is based on a '*de-novo'* decision analytic Markov model to estimate the cost-effectiveness of treatment with (i) fludarabine monotherapy (F), (ii) fludarabine in combination with cyclophosphamide (FC) and (iii) chlorambucil (Chl). A brief overview of the key assumptions used in the analysis, alongside a narrative description of the main approach used is reported below. This is followed by a more detailed critique of the model structure, assumptions and data inputs applied in the model.

The key assumptions used in the model include:

(i) that initial response rates from the CLL4 study are appropriate from a UK perspective and that they provide the best available evidence, such that other potentially relevant evidence on the use of these treatments as first-line therapy were excluded (p83 of the manufacturer's submission);

(ii) that patients must progress through all lines of therapy before CLL mortality is possible (p94);

(iii) that previous estimates of re-treatment response rates from different studies are appropriate for re-treatment with F and Chl, and that re-treatment with FC has the same response rate as initial treatment (p94);

(iv) that duration of re-treatment for responders is the same as for initial duration from the CLL4 trial (p93);

(v) that F is the most appropriate second-line therapy for non-responders initially treated with Chl, that CHOP is most appropriate for those initially treated with FC, and that FC is the most appropriate second-line for treatment for non-responders initially treated with F (p103-104);

(vi) that previous estimates of these re-treatment response rates are applicable to the decision problem (p118); and

(vii) that overall survival is the same for all treatment arms (p94).

The results for the economic evaluation are presented for the base-case, and thereafter for several other scenarios through sensitivity analysis (results of which are provided in Table 54, p133-134). A probabilistic sensitivity analysis has also been undertaken (details of distributions and their sources are provided in Table 40, p118, and results are provided in Table 43, p121 and Figures 6, 7 and 8 on p122-123).

5.3.1 Natural history

The model uses patient level clinical data from CLL4 to model first-line treatment including the response rates applied in the model. A detailed summary and appraisal of the CLL4

study is described in Sections 4.2.1 and 4.2.2 of this report. In summary, this study included 783 patients with an age range from 35 to 86 years. A total of 194 patients (24.8%) were treated with F, 387 (49.4%) with Chl, and 196 (25%) with FC, leaving 6 (0.77%) patients untreated. Data for second-line and subsequent treatment rates have been taken from a variety of published sources (details of which are listed in Table 34, p93).

The analysis is based on a Markov model with 260 cycles (with each cycle representing a 28 day period) giving a time horizon of approximately 20 years. There are 16 health states in the model, which can be separated into 5 different treatment states, 5 treatment response states, 5 disease progression states and a death state (representing both CLL and non CLL mortality). Patients enter the model on initiation of first-line treatment and remain in the initial state for the period of time for which their first-line treatment continues. Patients are then divided between those who have a response of 12 months or more ("responders") and those who do not ("non-responders"), where a responder is a patient who has a response of 12 or more months. In subsequent cycles of the model, responders to first-line treatment remain in the "response" state or they experience disease progression and move into a "progression" state for a period of time before receiving their second-line chemotherapy. In accordance with CLL4 protocol, responders to first-line treatment are assumed to be re-treated with the same agent as first-line when their disease progresses. These patients then remain in the "re-treat" state while their treatment continues at which time they move to either the "response" state or directly to the "progression" state. Those patients that achieve a response to re-treatment will remain in the "response" state until they move into the "progression" state. Following a period in the "progression" state these patients then move into the "salvage" state where third-line therapy is initiated. Patients remain in this state for a number of cycles before moving into either the "progression" or "response" states as in a similar manner to that assumed for second-line treatment. Patients who respond to salvage therapy remain in the response state until they ultimately move to the final "progression" state. Once patients enter the "progression" state following third-line therapy they are assumed to be at a constant risk of death from CLL.

Patients who do not achieve an initial 12 month duration of response to first-line treatment (non-responders) follow a similar path to the "responders" but the second-line treatment is not a repeat of the first-line therapy given (details of which second-line treatment is given are detailed in Table 32, p92). While mortality from CLL is only allowed once patients have progressed through the complete sequence of treatments (first-line, second-line/re-treatment and salvage), patients are allowed to make a transition to death due to non-CLL mortality from any state in the model.

5.3.2 Treatment effectiveness

The treatment effectiveness data for first-line treatments in the model is derived directly from the CLL4 patient-level data. Treatment effectiveness comprises two main elements: (i) the initial response to treatment and (ii) the duration of the response. The clinical effectiveness data from the CLL4 trial demonstrates a higher response rate with FC compared to both F and Chl ([redacted] vs [redacted] and [redacted] respectively). The data also showed a longer time in response state for patients receiving FC, with F having the next longest time ([redacted] months for FC, [redacted] months for F and [redacted] months for Chl).

The model uses the individual patient data from CLL4 and puts this directly into the model until a patient is censored. An individual is censored when they enter second-line treatment, are lost to follow up or the follow up ended while still in the response to therapy or progression states. Once a patient becomes censored the model uses transition probabilities based on non-censored patients from the CLL4 trial, and on other studies (Table 32, p92), to estimate progression through the model. Patients who received second-line treatment were censored from the CLL4 study patient level data provided and hence response rates for second-line and salvage treatments were not available from the CLL4 study patient level data provided. As a result data from other studies have been used to estimate subsequent response rates. Since no previous data was reported to be available on the re-treatment response rate for patients treated with FC, the manufacturers have assumed a response rate equal to that observed at first-line and this appears a very strong assumption (given that other data used for re-treatment with Chl and F were lower than the estimates applied for first-line treatment). Furthermore, due to a lack of external data on the duration of response of re-treatment with the initial therapy, the model assumes that this duration is equal to the initial duration of response.

Due to the limited follow-up currently available from the CLL4 trial, the existing survival data were not considered to be sufficiently mature to demonstrate differences in overall survival between the treatments. The manufacturers have, therefore, assumed that overall survival was the same for all groups. This is implemented in the model by making the time from first progression to death shorter in patients who had received F or FC than it was for those who received Chl as first-line therapy. Consequently, the model assumes that any gain in median progression-free survival associated with F or FC was offset by an equivalent decrease in the median survival after final progression.

5.3.3 Health related quality of life

The manufacturer's submission presents results in terms of cost per QALY gained. In the absence of a preference-based measure of quality of life instrument used in the CLL4 study, utility estimates applied in the model were derived from external sources. These estimates were identified through a systematic literature search, details of which can be found in Appendix 2 of the manufacturer's submission. Utility values were obtained for three health states: (i) receiving treatment; (ii) response and (iii) progression.

The potential differential quality of life impact from side effects from the alternative treatments was not considered in the base-case analysis. However, an additional sensitivity analysis was undertaken to examine the robustness of the base-case results to an additional decrement applied to patients treated with F and FC.

5.3.4 Resource utilisation and costs

The manufacturer's submission includes data on resource utilisation including: chemotherapy, monitoring, medications for prophylaxis, surgery, management of serious adverse events related to chemotherapy and longer-term resource utilisation after the initial treatment was completed. The medical resource use data were based on data from an audit of 113 UK patients from the CLL4 study. The resource use data were collected from the patient's entry into the study until the earliest of (i) death, (ii) initiation of second-line therapy, or (iii) the date of data collection. Data were abstracted from computerised databases or extracted manually from patients' charts and transferred to case record forms. Regression approaches were then used to adjust for differences between the audit population and the trial population in the final model, based on differences in the number of cycles of treatment received.

Unit costs were derived from relevant national sources. The average unit cost for outpatient attendance in a haematology department was chosen as the cost for out patient visits. Unit costs for tests and hospitalisation were taken from NHS Reference costs. Day case costs were based on applying the day case cost of £344 to each cycle in which one or more day case visits were reported. The inpatient unit cost was estimated as the average of the two HRG groups "Malignant disorders of Lymphatic or Haematological Systems with complications" and "Malignant disorders of Lymphatic or Haematological Systems without complications". Monitoring costs were split into four groups; (i) haematology, (ii) biochemistry, (iii) histology and (iv) radiology. Medication costs were taken from the prescription cost analysis database.

The costs of treatment for second-line patients were based on the NICE technology appraisal for fludarabine and CHOP in second-line treatment for CLL, while the second-line use of FC was assumed equal to those treated with it at first-line. The cost of re-treatment with the same agent was assumed to be equal to the cost of initial treatment. The costs of salvage treatment were based on an audit of patients' notes in low-grade Non-Hodgkin's Lymphoma.[42]

Only costs and resource use of the NHS were considered and no account has been taken of costs to PSS.

5.3.5 Discounting

The manufacturer's submission has used a constant discount rate of 3.5% for future costs and QALYs.

5.3.6 Sensitivity analyses

The manufacturer's submission includes simple one-way deterministic survival analysis, probabilistic sensitivity analysis and scenario analyses.

5.3.7 Model validation

The submission reports that the structure and key assumptions in the model have been validated by two experts in the treatment of CLL: Professor D Catovsky (Lead Investigator of the CLL4 study) and Professor P Hillmen (Chairman of the CLL Trials sub-group of the NCRI). They also report that numeric values in the model have been checked by an experienced modeller not involved in the construction of the model or the subsequent analyses.

5.4 Critique of manufacturer's economic model

The ERG has considered the methods applied in the manufacturer's economic evaluation in the context of the critical appraisal questions listed in Table 5.2 which are drawn from common checklists for economic evaluation methods.[41]

Table 5.2: Critical appraisal checklist

Item	Critical Appraisal	Reviewer Comment
Is there a well defined question?	Yes	
Is there a clear description of alternatives?	Yes	Fludarabine vs fludarabine plus cyclophosphamide vs chlorambucil
Has the correct patient group/ population of interest been clearly stated?	?	The CLL4 study includes asymptomatic stage B patients and these individuals would not be eligible for treatment based on the current license for fludarabine.
Is the correct comparator used?	?	Chlorumbacil is the most common first-line treatment in the NHS but choice of second-line therapies may not reflect optimal practise.
Is the study type reasonable?	Yes	CEA model used.
Is the perspective of the analysis clearly stated?	Yes	Perspective stated as costs to NHS and health benefits to patients.
Is the perspective employed appropriate?	Costs- Yes Outcomes-Yes	Submission adopts a UK NHS perspective for costs, although they fail to take account of costs to PSS, so only partially consistent with the NICE reference case. Perspective on outcomes is that of the patient.
Is effectiveness of the intervention established?	?	CEA is based on clinical effectiveness data from the CLL-4 trial (this also includes Stage B non progressive patients). The effectiveness of the interventions is established for first-line response rate but not for other parameters such as response to second-line treatments, median progression-free survival and overall survival as the data is not sufficiently mature enough yet.
Has a lifetime horizon been used for analysis (has a shorter horizon been justified)?	Yes	CEA used a 20 year horizon (which according to submission the majority of people have died by)
Are the costs and consequences consistent with the perspective employed?	?	Costs are consistent with NHS perspective but limited info on collection of cost data. Consequences measured in QALYs but not all based on preference-based utility measures.
Is differential timing considered?	Yes	
Is incremental analysis performed?	Yes	
Is sensitivity analysis undertaken and presented clearly?	Yes	Sensitivity analysis is taken and presented clearly but could be considered limited since primarily based on one-way deterministic analysis.

Table 5.3 below compares the manufacturer's submission to that of the NICE reference case.

Table 5.3: NICE reference case checklist

Attribute	Reference Case	Included in submission	Comment on whether de Novo evaluation meets requirements of NICE reference case
Comparator(s)	Alternative therapies including those routinely used in NHS	Yes	Chlorambucil is the most commonly used first-line therapy for CLL in the NHS with over 60% of patients receiving the drug
Perspective costs	NHS and PSS	?	NHS costs have been taken into account but no consideration of PSS costs has been undertaken
Perspective benefits	All health effects on individuals	Yes	QALY benefits to treated individuals are considered
Time horizon	Sufficient to capture differences in costs and outcomes	Yes	The 20 year time horizon is sufficient to capture the lifetime of the vast majority of patients
Synthesis of evidence	Systematic review	?	The base-case uses the results from the single study which could be considered the most relevant. A systematic review of other studies was undertaken to populate a sensitivity analysis although some of the methods for evidence synthesis are very crude
Outcome measure	QALYs	?	Non preference-based utilities utilised.
Health states for QALY measurement	Described using a standardised and validated instrument	?	The base line utility value is based on a measurement using the EQ-5D on elderly patients suffering from a lymphoproliferative malignancy with a similar EORTC QLQ-C30 score to those in the CLL4 trial. However utility values for the progressive disease and progression-free disease have been taken from a previous economic evaluation which used the EORTC QLQ-C30 which is not a validated utility instrument. Values then used in the sensitivity analysis from other studies also used this non-preference-based instrument.
Benefit valuation	Time Trade Off or Standard Gamble	?	
Source of preference data	Sample of public	?	
Discount rate	Health benefits and costs	Yes	Benefits and costs have both been discounted at 3.5%
Equity	No special weighting	Yes	No special weighting was undertaken
Sensitivity analysis	Probabilistic sensitivity analysis	Yes	Probabilistic sensitivity analysis has been undertaken.

5.5 Detailed critique of modelling methods

A critical review of modelling methods has been undertaken. The review has used the previous checklists and the framework for good practice in modelling presented by Phillips el al (2004) [43] as a guide for addressing the modelling methods employed by the manufacturers.

5.5.1 Modelling approach / Model Structure

The decision problem and objective are clearly stated within the manufacturer's submission, including details of:

1) Disease - Chronic Lymphocytic Leukaemia.
2) Patient group- those with B-cell CLL with "sufficient bone marrow reserves" and first-line treatment should only be initiated in those with (i) Binet stage C or (ii) Binet stage A/B with disease related symptoms or evidence of progression.
3) Possible treatment pathways- F followed by FC for non-responders, FC followed by CHOP for non-responders and Chl followed by F for non-responders.

Although the decision problem is clearly stated, the patient group subsequently considered by the economic model differs from that stated in the original decision problem. In particular, the economic model is based directly on the CLL4 trial data and, as such, (1) incorporated patients with Binet stage B without progressive symptoms; and (2) did not specify that patients should have sufficient bone marrow reserves. Comments from the manufacturer and our own clinical advisor suggest that these differences are unlikely to significantly affect the results presented here.

The choice of second-line treatment has been modelled in a very rigid manner which may not reflect the variation in the use of second-line treatments in routine clinical practice. Indeed, patients on the CLL4 trial requiring second-line treatment were actually randomised to either treatment guided by the results of the DiSC assay or to treatment guided by protocol guidelines, which could result in any of the other 2 treatments or CHOP being used. As such it is clear that there is a number of alternative second-line treatment strategies that the manufacturers could have considered. Additional sensitivity analyses have been undertaken to examine two alternative sequences. An additional analysis ("FCR") considered the impact of assuming that patients who do not respond to FC at first-line receive a second-line of chemotherapy with fludarabine, chlorambucil and rituximab (FCR) before proceeding to salvage treatment. A second analysis ("C-FC") considered the use of FC instead of F as second-line therapy after patients fail to respond to chlorambucil monotherapy.

The model is a Markov model and would appear to be an appropriate modelling approach to the decision problem, although there are some concerns over the implementation of the data into the model (see later sections). The model structure and the choice of treatments are stated in the submission to have been developed from current guidelines, consultation with clinical experts and the design of the CLL4 study. Although this is contrary to the Phillips *et al* guidelines [43] which state that structure should not be dictated by current practice, this is not a major concern here as the current guidelines are to follow the trial. However, there is some concern that the treatment strategies followed in the trial may not contain the optimal strategy (e.g. second-line therapies may not be optimal) and that the limited range of sequences considered in the model may exclude alternative sequences which may be more efficient than those that were included. The choice of second-line treatments could have implications for the overall cost-effectiveness of the first-line strategies.

The perspective on costs taken is that of the NHS, this differs from the NICE reference case as no account was taken of the costs relating to PSS. However, the exclusion of PSS costs is considered by the ERG to be unlikely to significantly affect the current results and may be potentially conservative towards the fludarabine regimens.

The perspective on benefits is quality of life benefits to treated individuals and this is consistent with the NICE Reference Case which requires estimation of all health effects to individuals. The model produces QALY estimates and NHS costs, although there is some concern from the ERG about the utility values used to inform the QALY estimates.

The model uses a 20 year time horizon which would appear to be appropriate for the condition as median survival for CLL is 10 years. However, survival in the model is between [redacted] and [redacted] in the treatment arms at 20 years so there could still be differing costs and benefits after 20 years (especially if the conservative assumption of equalising survival duration does not hold).

5.5.2 Structural assumptions

The submitted economic evaluation assumes that all the important factors related to the disease and its treatment can be captured in the sixteen Markov states (see model schematic in the manufacturer's submission, p91). In particular the possibility of more than three lines of treatment, the exclusion of a specific state for individuals who suffer side effects and the choice of specific second-line treatments for non-responders in each drug could affect the results presented here.

The ERG noted two important structural assumptions related to the current model structure which may potentially affect the cost-effectiveness estimates presented here:

1. The model assumes that patients must progress through all lines of therapy before dying as a consequence of their CLL. This will have differential effects on both the costs and benefits between treatments as the time spent in each state will vary according to the probabilities assigned to the particular therapy lines. This assumption may result in patients incurring additional costs and quality of life decrements that they would not have actually experienced.

2. A separate structural issue was identified relating to the modelling of second-line treatment (either re-treatment with same initial agent or treatment with an alternative agent). Within the current structure the model assumes there is only the possibility either being a responder or going straight to progression. This means that it does not allow for people to respond for less than a year and hence the utility and cost implications of response times of the period of any partial response are not considered (i.e. the approach is inconsistent with the way that non-response to first-line treatment is modelled). It would have been more appropriate to include an extra state to which non-responders would move (similar to initial first-line treatment) that would then allow them to move to the relapse state, but at a quicker rate than those who were responders. This would allow response duration of less than 12 months to have been incorporated into the analysis. Failure to include this state will lead to quicker progression to the relapse state and thus to lower estimates of QALYs (and potentially higher costs) for these patients. Omission of this state will potentially bias the cost-effectiveness results towards interventions with the higher initial response rate.

5.5.3 Data Inputs

Patient group

The analysis uses patients from the CLL4 trial. They have justified using only one trial as it is the only RCT to compare all three treatments being considered in the economic analysis. Despite the lack of head-to-head trial evidence comparing all treatment simultaneously, additional trial evidence has reported on various pair-wise comparisons of these treatments. There must therefore be some concern that they have excluded relevant data from the other trials they identified which may have improved the precision of their estimates (for examples see Table 10, p45 in manufacturer's submission).

The CLL4 study inclusion criteria permitted recruitment for all patients with B-cell CLL, previously untreated, diagnosed by a persistent lymphocytosis (greater than 10x109/l) and bone marrow infiltration of at least 40%, who require treatment, with stage A progressive, stage B or stage C disease using the International (Binet) staging system. Sub-group analyses of patients aged above or below 65 years of age and of the grade of disease at baseline were also conducted. The CLL4 study includes individuals with Binet stage B with non-progressive symptoms (although it has not been possible to ascertain how many) who would not be eligible based on the current license for fludarabine. The CLL4 study does not include individuals with hepatic impairment who might be included in the licensed indication and this may result in the results of the study not being fully applicable to the decision problem.

Clinical effectiveness - first-line treatment

The model uses data from the CLL4 study on initial response rates, duration of response and time between progression and re-treatment to estimate the relevant transition probabilities following first-line treatment. This assumes that the data from the CLL4 study is able to fully reflect the transitions of the patients that the decision problem is focused on. Furthermore, the model also assumes that transition probabilities between all states are constant over time. Given the aging of the cohort and the 20 year time horizon this assumption may not hold. Although the submission has varied the absolute transition probabilities in the sensitivity analysis it has not considered varying the transition probabilities over time and it is not clear what effect this would have on the results. The ERG identified two potential concerns regarding these aspects:

1. Although the CLL4 trial is the only study to have considered a comparison of all three treatments in the same study, a number of separate studies have examined the effectiveness of FC versus F alone and F versus Chl. This wider-set of evidence constitutes potentially relevant evidence which could have been considered using indirect approaches to evidence synthesis. It is unclear what impact including these additional studies would have had on the final cost-effectiveness estimates. However, the results of the individual trials reported in the previous chapter do not appear to be inconsistent with the results from the CLL4 trial.
2. The current approach to extrapolation assumes that the risk of particular transitions follow an exponential distribution (i.e. that the hazard is constant with respect to time). No supporting evidence is provided to justify this assumption. However, since the manufacturers had access to the patient level data then statistical approaches using

survival analysis could have been used to formally test this assumption. The disparity between the cost-effectiveness results at 5-years and at longer periods suggests that the main cost-effectiveness advantage is conferred in the period of extrapolation. Hence the approaches to extrapolation are likely to be central to the validity of the subsequent estimates of cost-effectiveness.

In the manufacturer's submission first-line treatment transition probabilities have been estimated directly from the CLL4 patient level data. Separate transition probabilities were estimated for patients classified as responders (i.e. a response lasting 12-months or more) and non-responders. The ERG identified a number of important concerns regarding the approach used by the manufacturers to estimate the transition probabilities according to response status to initial treatment and the way they are subsequently applied in the model.

Individual patient data is employed in the model to estimate particular transitions until patients are censored or die. Once a patient in censored they enter into the model at the same state they were censored from. Subsequent transitions for censored patients are then estimated based on the data from non-censored patients. For example, if the patient was still in the treatment response state at the time of censoring they are entered into this state in the model and the speed at which they exit this state is subsequently determined by the transition probability calculated from the observed trial data. In reviewing the manufacturer's calculations the ERG identified a number of potential sources of concern.

- The transition probabilities have been estimated by simply calculating the total number of exits of the state (not including the count of the number of patients censored) divided by the total amount of time spent in the state (including the time that those who were censored spent in the state). However, as the time spent in the state already includes the time experienced by censored patients, subsequent transition probabilities for this group need to be estimated conditional upon being censored. Since the appropriate conditional probabilities were not estimated, the subsequent transition probabilities for the censored/unobserved group are potentially underestimated. This would not necessarily cause bias in favour of any one treatment if the error impacted on all three treatments equally but since patients treated with F or FC remain in response to their first-line treatment longer than Chl there is potential for bias in favour of these treatments.

- Another key concern exists with the approach taken to estimating transition probabilities from the CLL4 data, such that the method used to estimate the particular

risks appear flawed. Two potential problems were identified by the ERG. Firstly, the transition probabilities themselves are based on time at risk for all possible transitions, rather than estimating time at risk for each possible transition separately. Secondly the probabilities of three transitions have been summed to estimate a single probability estimate for leaving the response state. In the manufacturer's model this single probability is assumed to represent the probability of moving to the "progression" state but in fact this transition also includes the probabilities of dying and of patients moving to second-line treatment. This is in part due to the structural assumptions of the model which only allows patients to die from CLL after they have progressed through all possible lines of treatment. It would have been more appropriate to model the possible transitions from this state separately (i.e. to progression, to second-line treatment and to death) which would also allow the time patients are at risk for each possible transition to have been incorporated.

An example of the calculations employed by the manufacturers is shown in Table 5.4 below which shows the calculation of the transition probabilities for responders to fludarabine, highlighting the issues raised above.

Table 5.4 Calculation of transition probabilities for fludarabine responders

	Responders	
Months observed	[redacted]	
Number of transitions	Censored	[redacted]
	Dead	[redacted]
	Progression	[redacted]
	Second-line treatment	[redacted]
% monthly risk of transition (MRT)	Censored	[redacted]
	Dead	[redacted]
	Progression	[redacted]
	Second-line treatment	[redacted]
% risk of exit state	(Sum of MRT for dead + MRT Progression +MRT 2nd line calculations)	[redacted]

The ERG has a number of serious concerns about the approaches used in the manufacturer's submission for the estimation of transition probabilities. Some of these relate to the structural assumptions in their model (e.g. patients can only make a transition from response to therapy to progression and not to any other states) and some are due to the statistical approaches used (e.g. not making the transition probabilities conditional upon being censored and not adjusting estimates for particular transitions to allow for time at risk of each particular event). To adequately address these concerns formal survival analysis methods could be employed (which would allow for the censoring issue to be dealt with

correctly). However, then to incorporate the results into the model would require a major restructuring of the current model and hence this is beyond the scope of the ERG report.

Additional analyses were undertaken using formal survival analysis approaches by the ERG using the patient-level data from the CLL4 trial to more formally test the manufacturer's assumption of a constant hazard for particular transitions. Full details of these additional analyses are reported in Chapter 6 of this report.

Clinical effectiveness - second-line and salvage treatment

In the absence of relevant external data on the response rate for re-treatment with FC, this has been assumed to be equal to the initial response rate (i.e. the rates based on those who responded for more than one year). In contrast, the response rates for re-treatment with chlorambucil or fludarabine have been taken from alternative published sources (details of which can be found on p118). These response rates have been multiplied by the probability of exiting the treatment state in each period (calculated from the median number of treatment cycles using a standard exponential approach) to obtain the transition probability for exiting the treatment state to the treatment response state. The probability of exiting the treatment state to the disease progression state is equal to the probability of exiting the treatment state minus the probability of exiting the treatment state to the treatment response state. This means that those who do not respond to second line treatment do not spend any time in a response state and would move directly from the treatment state to the progression state which has lower utility (unlike at first line where non-responders still respond for a period of time and thus spend time in the higher utility state).

The assumption that the re-treatment response rate for FC is the same as for the initial response rate is very strong and may potentially over-state the cost-effectiveness of this treatment in the model. This assumption has been made as no other evidence about re-treatment with FC was reported to be available. It is worth noting that, due to the assumption of an equal re-treatment response rate for FC, which is larger than re-treatment rates for F and Chl, then this could bias the results in favour of FC as more individuals will move to the response state and experience a higher quality of life than individuals who move to subsequent treatment and/or progression.

Although the manufacturers conducted a sensitivity analysis on the re-treatment response rates, the choice of range in sensitivity analysis for this parameter is potentially conservative as it is based on a 95% confidence interval from the bootstrap of the original response rate data. This is potentially inappropriate as it is still based on the same assumption of an equal

re-treatment rate. This issue is discussed further in the sensitivity analysis section of this report and additional analyses have been undertaken by the ERG on this particular aspect in Chapter 6.

The model also assumes that the duration of response for re-treatment with the same agent is equal to the duration of response to the initial treatment. This assumption has again been made as no further evidence was found during the manufacturer's search. Given the disparity in the duration to first response observed in the trial between the treatments ([redacted] weeks for FC, [redacted] weeks for F and [redacted] weeks for Chl) this assumption could be important as it is effectively double counts any initial treatment benefit, yet no evidence is provided to support such a claim and no attempt has been made to address the uncertainty surrounding this assumption.

Transition probabilities for second and subsequent lines of treatment have been estimated from published median values (see Table 40, p118). They have assumed a constant relative risk of transition out of a state, and have used a standard exponential approach to calculate the cyclical transition probability. The ERG has several concerns about the approach used to implement data from other studies. In particular, the methods used to synthesise data from several sources for one parameter are a cause for concern. The submission has simply pooled the data from several sources for the response rates for second-line treatment with FC after F as first-line treatment and for F after Chl as first-line treatment. By taking absolute estimates from the studies, the benefits of randomisation are lost and the differences observed may simply be due to the different patient characteristics from the different studies. The ERG is, therefore, concerned that this approach could potentially affect the absolute cost-effectiveness estimates for both the fludarabine and chlorambucil first-line treatments and this may then impact upon the relative cost-effectiveness estimates for all treatments.

5.5.4 Survival

In the base-case of the manufacturer's model it is assumed that overall survival is the same for all treatments. The manufacturer's submission argues that, due to the limited follow-up data available, current CLL4-trial data are not mature enough to be able to demonstrate any mortality benefits with individual treatments. Consequently, the manufacturers appear to have taken a conservative approach of equalising survival across all treatments. They attempted to achieve this by "assuming that any gain in median progression-free survival associated with fludarabine or fludarabine with cyclophosphamide was offset by an equal decrease in median survival after final progression" (p94). The ERG identified a number of potential concerns regarding this assumption:

1. This “conservative” assumption focuses on equalising median survival rather than mean survival. Within the current modelling approach, differences will still exist between treatments in terms of mean survival estimates (with estimates of mean survival highest for treatment with FC). As a result, any differences in mean survival will be reflected in the subsequent estimates of the ICER. (NB: Following correspondence with the manufacturer an addendum was submitted which presented the results of the analysis based on equalising mean survival).

2. There are also concerns that, by forcing people in the other treatment arms to spend longer in the final progression period (which has a low utility but still incurs a cost), the model may be potentially biasing the results in favour of the intervention with the longest original progression-free survival time.

3. The data from the CLL4 trial results actually show a higher mortality (although not statistically significant) in the F and FC treatment arms. Hence an analysis based on extrapolation of the CLL4 trial data itself could have altered the current cost-effectiveness estimates. Until more mature survival data are reported from the CLL4 trial it is unclear whether the current approach is actually conservative or not.

5.5.5 Health-related quality of life

Due to a lack of a preference-based measure in the CLL4 trial external sources were used. These were found using a systematic search the details of which can be found in Appendix 2 of the manufacturer’s. In the submission the main health benefit assessed was quality-adjusted life-years (QALYs). These were calculated using preference scores for the three states; (i) receiving treatment, (ii) in response, and (iii) progressive disease.

In the base-case analysis, utility values for patients receiving initial treatment were taken from a single study by Doordujin *et al* (2005).[44] The patient population considered in this study were elderly patients with aggressive non-Hodgkin’s lymphoma, of which CLL patients comprised only a proportion of the group. Utility estimates in this study were obtained using the EuroQol (EQ-5D). Utility values representing the “response” and “progressive disease” states were referenced to the previous cost-effectiveness study by Hancock *et al* (2002).[12] Closer examination of this study reveals that the utility values were themselves actually derived from an earlier study by Holzner *et al* (2001).[45] These values were obtained using disease-specific quality of life instruments (EORTC QLQ-C30 and the FACT-G) and, as such, these values do not constitute preference-based measures of health utility. They are,

therefore, not directly comparable to the values taken from the Doordujin *et al* study [44] and their appropriateness needs to be questioned in the subsequent calculation of QALYs. Table 5.5 summarises the utility values applied in the manufacturer's submission.

Table 5.5 Utility values applied in the submission

Source	Model base-case
Base line (Treatment)	0.74
Utility value response	0.80
Utility value, progressive / active disease	0.60

In the base-case of the model it is assumed that patients have a small decrease in their quality of life while receiving chemotherapy, perhaps reflecting toxicity or other side effects. The quality of life data from the CLL4 study (captured by the EORTC QLQ-C30 instrument which have problems as discussed above) found reductions in the average quality of life for the first three months with F and FC but not with Chl. However, these differences were small and not statistically significant. The manufacturer's submission, therefore, includes a sensitivity analysis whereby patients on the chlorambucil therapy do not experience the decrement in quality of life while on treatment.

5.5.6 Resource use and costs

Resource use has been based on data from a sample of 113 UK patients from the CLL4 study. Only limited details were reported in the original submission about how this sample was selected. An important issue is that this sample may no longer constitute a randomised sample since it is unclear how the patients were selected for this audit.

Following queries raised by the ERG, the manufacturers responded that "participating centres were selected at a meeting of the study managers so as to give a reasonable geographic spread and variety of hospital types whilst ensuring data collection was possible within budget and time constraints.

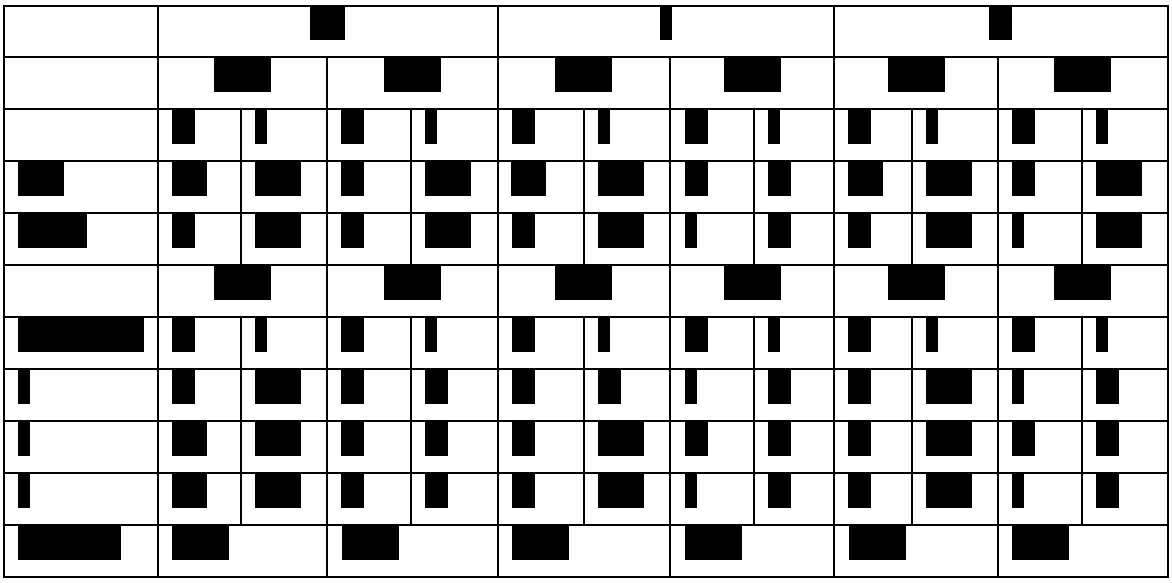
[redacted]. The total possible patient pool from the centres was confirmed by the manufacturers to be 126 and the sample obtained represented 90% of potential patients from the selected centres.

[redacted] Consequently it is not possible for the ERG to assess the potential

bias that may have been introduced with the sampling approach used to estimate resource use and cost data.

Although the submission claims that the patients in the audit were broadly similar to those in the trial in terms of age, gender and Binet staging there were differences as can be observed from the Table 5.6 below:

Table 5.6 Characteristics of trial and audit populations



In particular it is worth noting the gender variation in fludarabine and FC, and the large variations in Binet staging. It is not clear what effect this will have on the reliability of their estimates or the applicability of their resource use data to the population of interest. However, it is worth noting that a higher proportion of patients were reported to be in Binet stage B in the fludabarine regimens in the audit compared to the model. In the absence of data reporting on the proportions of patients in stage B with progressive symptoms it is difficult to assess the potential bias that this may introduce. The inclusion of asymptomatic patients with Stage B could, however, result in lower overall costs as these patients maybe less likely to require costly treatment.

Of more relevance to the validity of the model results is that the mean number of cycles in the audit differed from the modelled dataset. The mean number of cycles in the audit in the Chl, F and FC arms were [redacted] cycles and in the whole modelled dataset were [redacted] respectively. In an attempt to correct for this difference the manufacturer has conducted a regression analysis on the audit dataset to estimate per patient cost as a function of the number of cycles of the specified chemotherapy received. However, a

potential limitation of this approach is that they have used a standard ordinary least squares procedure which was estimated assuming the cost of treatment (excluding adverse events) was a function of a common intercept term and the number of cycles of the specific chemotherapy the patient was on. The reliability of this result is questionable due to the very possible exclusion of other relevant independent variables (e.g. Binet stage, age etc) which would lead to a biased estimate of per cycle cost. The choice of an ordinary least squares approach may also be inappropriate due to the skewed nature of cost data.

The submission uses per cycle cost estimates from the regression analysis (the coefficients on the number of cycles) and has inflated them to 2004/2005 prices. The subsequent use of these costs in the model excludes the intercept term from the regression analysis (the intercept term was [redacted]). This would seem appropriate for first-line treatment as all individuals will incur the intercept cost (representing a cost which is independent of the number of cycles received) and, therefore, the cost will be common for all treatment arms. However, due to the differing mortality and response rates between the treatment arms, these costs will not be common for re-treatment and therefore should be included in the estimates for the cost of re-treatment. Failure to include these could lead to subsequent biases in the cost-effectiveness estimates, since a higher proportion of patients in the model will be retreated with FC for whom this intercept term has not been included in the cost estimates of re-treatment.

The regression analysis also excludes the costs for adverse events. [redacted]. Due to the small number of patients considered in the audit, including such costs may have biased the results in favour of FC [redacted]. By excluding them the submission has implicitly assumed that the number and type of adverse events are common to all treatments and thus can be excluded from the analysis. However due to differing mortality between treatment arms these costs will not be common to all patients since, for example, re-treatment rates with initial therapy will differ. Therefore the exclusion of such costs could result in biases in the results. Furthermore, it may have been more appropriate to use the audit data to estimate the resource use associated with particular side-effects (regardless of initial treatment allocation) and to relate these to the trial-specific estimates of the incidence of side-effects for each treatment, due to the small number of patients evaluated in the audit and the relative infrequency in which particular side-effects were reported. Consequently there is a potential disparity between the incidence of side-effects reported in the trial and the cost estimates applied in the model.

All costs relating to the treatment period have been put into one component for each treatment and all costs relating to periods in response or progression have been considered equal. There is also a lack of transparency about the nature of the costs included in the model. The patient level data from the audit was not provided with the submission and it is therefore impossible to verify the costs which have been included. In particular, we are unable to verify how the estimated follow-up costs have been included into the analysis.

The audit is something of a 'black box' and it is difficult for ERG to draw any conclusions about what biases may exist, if any, and which way such biases may influence the results. For example in Appendix 9 in Tables 17a to 17c of the manufacturer's submission we are given details of the breakdown of the costs for patients during the treatment period. However, it is then not entirely clear which of these costs have then been included into the model.

5.6 Consistency

The submitted Excel model has been examined for internal and external consistency and accuracy.

5.6.1 Internal consistency

Random checking has been carried out on some of the key equations in the model, for example on Excel worksheets titled "Processing", "Control Panel", "F" and "Summary". A comprehensive checking process against all cells in the model has not been undertaken. The submitted model is fully executable and input changes in the "Control Panel" or "Processing" sheet produce immediate changes in the deterministic results on the "Summary" sheet. The model is shown using the baseline scenario but the user is able to look at the other sub-group scenarios (i.e. by age or Binet stage) or at the probabilistic model by using the appropriate cells on the "Control Panel" sheet. However, there appears to be no easy way to check the deterministic sensitivity analysis other than to track down the corresponding cell on either the "Processing" or "Control Panel" sheets and enter the new figure manually (for example, to change the second-line treatment for FC from CHOP to FCR the results from CHOP must be replaced with those from FCR on the Processing sheet or the destination cells on sheet FC must be changed to use the FCR probabilities in the model).

One potential inconsistency in the model is the way in which QALYs have been calculated. The cycle length in the model is 28 days. To convert these cycles and their corresponding utility scores into QALYs the submitted model simply divides the summation of the product of the utility scores for each state by the number of cycles spent in each state by 12. As the

cycles are only 28 days this is an incorrect method and a more appropriate approximation would be given by dividing the summation by 13. The failure to do this has biased the results in the favour of the most effective intervention although the bias is only small. Table 5.7 below shows the difference in the ICERs calculated for the baseline case when this correction has been made.

Table 5.7 Incremental cost-effectiveness ratios with 13 cycle correction

	ICER by dividing by 12	ICER by dividing by 13
	C/E ratio	C/E ratio
F vs Chl	£ 19,613	£ 21,247
FC vs Chl	£ 2,602	£ 2,819
FC vs F	-£ 8,452	-£ 9,157

The submission assumes that overall survival durations with F, FC or Chl were the same. The manufacturers have implemented this in the model by assuming that any gain in median progression-free survival associated with F or FC was offset by an equal decrease in median survival after final progression. If overall survival is equalised then the number of lifer years should be equal across all treatment arms. The ERG has tested this by setting the utility scores for all non-death states equal to 1. If they have implemented their assumption of equal survival the resulting outcome measures on the summary sheet would be equalised. However, as Table 5.8 shows this is not the case.

Table 5.8 Estimates of life years from manufacturer's original model per cohort of 1000 patients

Comparator	Outcomes
Chl	7380
F	7515
FC	7633

This disparity is due to the manufacturers equalising median, as opposed to mean, survival. To equalise mean overall survival it is the mean gain in progression-free survival that needs to be offset by an equal decrease in mean survival after final progression. Since it is the mean survival (and QALY) estimates which are being used in the ICER calculation, an approach based on equalising mean survival would appear more appropriate. To address this particular concern the manufacturers submitted an addendum based on this alternative approach.

Parameter inputs have been checked for corrective predictive validity (i.e. Independent sensitivity analyses have been undertaken and results were consistent with those expected, e.g. increasing the response rate to chlorambucil increases its effectiveness).

5.6.2 External consistency

The model uses the CLL4 study data but, due to the nature of the model (i.e. individuals only enter the model once they are censored in the trial data), it is very difficult to validate the model against the observed data. One interesting comparison which has been undertaken is to look at the 5 year mortality data for both observed and unobserved (those who are censored and enter the model) patients. This is shown in Figure 5.1 below.

Figure 5.1: Survival curves for observed and censored patients

While FC has the highest observed mortality rate, it has the lowest unobserved mortality rate. This suggests that the observed mortality data may not have been used correctly in the model. This relates back to the problem that the mortality data are only used along with the progression and second-line treatment to inform one probability, that of exiting the response state to the progress state. As discussed before this assumption could have a large effect on both the cost and effectiveness of an intervention and thus could be influencing the subsequent cost-effectiveness results.

5.7 Results

5.7.1 Summary

The results of the model are presented in the manufacturer's submission from p120 to 134. In particular it is worth noting that the submission includes: (i) the base-case results showing ICERs of £19,613 per additional QALY for F vs Chl, £2,062 for FC vs Chl and FC dominating F (a summary of which can be found in Tables 41 and 42, p120); (ii) results from the probabilistic sensitivity analyses (Table 43 and Figures 6 to 8, p121-123 and a summary of the distributions used in this analysis is provided in table 40, p118); (iii) results from the sub-group analyses of patient age (above or below 65 years of age) and grade of disease at baseline (Binet stage A+, B or C) (Tables 44 to 47, p124-125); and (iv) results from the one-way sensitivity analyses conducted in the company's submission (details of which can be found from p127 to 131 and results on Table 54, p133-134).

The results from the addendum are presented in “Additional analysis to support QB10 and B12” from p4 to 12. The addendum includes (i) base-case costs and effectiveness results from the revised model (Tables 1 and 2, p4); (ii) probabilistic sensitivity analysis results from the revised model (Table 3, p4 and Figures 1-3, p5-6); (iii) sub-group analysis from the revised model (Tables 4 to 7 p6-7); (iv) results from the deterministic one-way sensitivity analysis (Table 8, p8-9); and (v) Tornado diagrams showing the cost per QALY results from the one-way sensitivity analysis (p10-11).

5.7.2 Base-case analysis

The results from the base-case of the original submission are presented below in Tables 5.9 and 5.10. Costs and outcomes are presented per 1000 patients and results are based on a deterministic analysis.

Table 5.9 Base-case results

1st line Treatment strategy	Costs	Outcomes (QALYs)
Chl	£11,659,803	5,096
F	£17,590,562	5,399
FC	£13,657,485	5,864

Table 5.10 Incremental results (Base-Case)

Incremental	Costs	Outcomes (QALYs)	C/E ratio
F vs Chl	£ 5,930,759	302	£ 19,613
FC vs Chl	£ 1,997,683	768	£ 2,602
FC vs F	-£ 3,933,077	465	Dominant

In the base-case analysis, treatment with F instead of Chl increased costs by approximately £5.931m and QALYs by 302 per 1000 patients, giving an ICER of £19,613 per additional QALY. Treatment with FC instead of Chl increases costs per 1000 patients by roughly £1.997m, and increases QALYs by 768, resulting in an incremental cost per QALY gained of £2,602. In the base-case analysis, FC strictly dominates F with both lower costs and improved outcomes.

The results from the base-case of the manufacturer's amended model are presented in Table 5.11 and 5.12 below. These results are based on the approach in which mean survival is equalised across the three strategies.

Table 5.11 Revised base-case results presented in addendum

Treatment strategy	Costs	Outcomes (QALYs)
Chl 1st line	£ 11,920,074	5248
F 1st line	£ 17,712,428	5469
FC 1st line	£ 13,919,492	5864

Table 5.12 Incremental results from addendum (Base-Case)

Incremental	Costs	Outcomes (QALYs)	C/E ratio
F 1st line vs Chl 1st line	£ 5,792,354	222	£ 26,105
FC 1st line vs Chl 1st line	£ 1,999,418	616	£ 3,244
FC 1st line vs F 1st line	-£ 3,792,936	394	Dominant

The changes made to the model have resulted in extra costs for all three treatment lines and extra QALYs for the chlorambucil and fludarabine lines. There are also additional costs in the FC arm due to a calculation error which the manufacturers identified in their original model which increased the cost by nearly 2%. The changes in costs and QALYs for all 3 treatment strategies are a result of the recalculated transition probabilities which have resulted in changing the average amount of time spent in states. The amendments made to the model have resulted in an increase in the ICER of F compared to Chl; however, F is still dominated by FC. There is also a slight increase in the ICER of FC compared to Chl (£3,244 vs £2,602 in the original submission).

The ERG considers that the approach used in the addendum based on equalising mean survival is more appropriate than the original one based on equalising median survival. Furthermore, the results reported in the addendum also incorporate the correction to the error identified in the cost calculations. As a result, all subsequent discussion of the manufacturer's results will focus on the results of the sub-group and sensitivity analyses reported in the manufacturer's addendum.

5.7.3 Sub-group analyses

Sub-group analyses were conducted based on age (patients aged above or below 65) and on grade of disease at baseline (Binet stage A, B or C). The results of the age sub-groups are shown in the Tables 5.13 and 5.14 below. All sub-group analyses were based on results from the deterministic analysis.

Table 5.13 Sub-group analyses for Fludarabine patients

	F		Chl		ICER
	Costs	**QALYs**	**Costs**	**QALYs**	
Age >=65	£15,363,339	4633	£10,060,707	4411	£23,923
Age<=64	£20,289,703	6495	£14,105,843	6279	£28,666

Table 5.14 Sub-group analyses for FC patients

	FC		Chl		ICER
	Costs	**QALYs**	**Costs**	**QALYs**	
Age >=65	£11,542,847	4828	£10,060,707	4411	£3,556
Age<=64	£16,750,918	7164	£14,105,843	6279	£2,989

In their submission the manufacturers note the alternative directions in which the ICERs move for F vs Chl and FC vs Chl (i.e. F is more cost effective for over 65 year olds than for under 65 year olds, while FC is more cost effective for under 65 year olds than over 65 years olds). The manufacturers state that these differences may simply be due to the sample size of the groups.

Due to disease state at baseline being a key prognostic indicator of overall survival, a sub-group analyses by Binet Stage was also undertaken. The results are shown in Tables 5.15 and 5.16.

Table 5.15 Sub-group analyses by Binet stage- Fludarabine vs Chlorambucil

	F		Chl		ICER
	Costs	**QALYs**	**Costs**	**QALYs**	
Binet Stage A+	£17,766,739	5254	£11,395,804	5086	£37,933
Binet Stage B	£18,933,937	5982	£12,631,558	5621	£17,444
Binet Stage C	£15,747,525	4883	£11,337,116	4848	£128,561

Table 5.16 Sub-group Analysis by Binet Stage- FC vs Chlorambucil

	FC		Chl		ICER
	Costs	QALYs	Costs	QALYs	
Binet Stage A+	£12,959,858	5508	£11,395,804	5086	£3,709
Binet Stage B	£14,986,995	6381	£12,631,558	5621	£3,099
Binet Stage C	£13,108,177	5423	£11,337,116	4848	£3,085

Analysis by Binet stage showed lower costs and QALYs gained for stage C patients, as would be expected due to their being more severely ill and with lower overall survival than other stages. However, the estimates of mean QALYs and costs were both higher in stage B patients than patients in either stage A+ or stage C. This result appears potentially counter-intuitive since one would expect these to lie between those for stage A+ and stage C patients. This may be explained by the inclusion of non-progressive patients in the CLL4 trial data, such that the prognosis of the stage B patients (including both progressive and non-progressive patients) appears better than either of the other two stages considered. However, these results may also simply be due to chance findings caused by the smaller sample sizes considered in the sub-group analysis. Compared to the base-case findings, F became more cost-effective (versus Chl) for those in stage B but less cost effective for stage A+. FC compared to Chl became less cost effective in stage A+ but was relatively unchanged for stage B. FC continued to dominate F in the sub-group analyses.

5.7.4 One-way sensitivity analyses

The results of the one-way sensitivity analysis contained in the manufacturer's submission are presented in Table 5.17. In summary the ICERs remained largely unaffected by the majority of these analyses. The results were most sensitive to the time-horizon of the model and the assumptions made related to re-treatment response rates for patients treated with the same agent used as part of first-line treatment. Both these analyses are discussed in more detail below.

Table 5.17 One-way deterministic sensitivity analysis results

Sensitivity analysis	F - Chl	FC-Chl
	C/E ratio	C/E ratio
Base-case	£26105	£3244
FC followed by FCR	£26105	£6659
CLB followed by FC	£37470	£4162
Re-treat if response ≥ 6 months	£23853	£3411
Response rates at 1st line from literature	£11313	£2268
F re-treat: response rate upper limit	£19036	£3244
F re-treat: response rate lower limit	£52331	£3244
CLB re-treat: response rate upper limit	£58164	£3874
CLB re-treat: response rate lower limit	£18229	£2895
FC re-treat: response rate upper limit	£26105	£3055
FC re-treat: response rate lower limit	£26105	£3466
FC after F: response rate upper limit	£22107	£3244
FC after F: response rate lower limit	£32207	£3244
F after CLB: response rate upper limit	£27775	£3298
F after CLB: response rate lower limit	£24481	£3187
CHOP after FC: response rate upper limit	£26105	£3225
CHOP after FC: response rate lower limit	£26105	£3261
Salvage: response rate upper limit	£26965	£3321
Salvage: response rate lower limit	£25420	£3183
Salvage data from pooled studies	£28831	£3486
Response rates on re-treatment equal first-line	£86770	£4185
Utility values from Doorduijn	£19567	£2380
Utility values from Wessex	£33795	£4253
No utility decrement during chlorambucil treatment	£34610	£3559
Removing rapid ‘progression to death time’ for fludarabine and making equal to chlorambucil	£9610	£2212
Time horizon 5 years	£310663	-£119663
Time horizon 10 years	£42516	£5726
Time horizon 15 years	£28178	£3616
Discount rates: cost 0%; outcomes 6%	£34351	£5124
Discount rates: cost 0%; outcomes 0%	£20223	£2633
Discount rates: cost 6%; outcomes 6%	£31121	£3860
Discount rates: cost 6%; outcomes 0%	£18322	£1983

Of particular interest is the difference in ICER depending on time horizon with the ICER for 5 years for both F and FC compared to Chl being markedly higher, or in the case of FC vs chlorambucil being negative (meaning that chlorambucil dominates FC). These findings may be attributed to the relatively slow progression times experienced by patients, such that at even 5-years follow-up the differences between treatments in terms of improving progression-free survival is not sufficient to demonstrate cost-effectiveness. This difference may also be being driven by the higher observed death rate in the F and FC arms of the trial and the reason that this is not continued for a longer time horizon (since the death rate has not been used to inform mortality for censored patients, instead this transition probability was combined

with several other transitions to estimate a single transition to the progression state). It is unclear what effect the inclusion of the mortality data into the model in an appropriate way would have had on results. Regardless, it is clear that the approaches used for extrapolation are central to the cost-effectiveness estimates.

Deterministic sensitivity analysis of the response rate assumed for patients who were re-treated with the same therapy used as part of first-line treatment had a substantial impact on the ICER of F versus Chl (£86,770 vs £26,105 in the base-case), while the ICER for FC versus Chl was only marginally affected (£4,185 vs £3,244). One explanation for these findings may be due to the approach used in the sensitivity analysis for the re-treatment response rate with FC. In the absence of published data on this parameter, the manufacturers varied the base-case estimate by applying the lower bound of the 95% confidence interval of the initial response rate to FC derived from bootstrapping the CLL4 trial data. The ERG does not consider that this is sufficiently rigorous since this is still based on data which assumes an equal re-treatment response rate.

Additional work undertaken by the ERG related to both these areas is presented in Chapter 6 of this report.

5.7.5 Probabilistic sensitivity analysis

The results from the probabilistic sensitivity analysis contained in the manufacturer's submission are presented in Table 5.18.

Table 5.18 Probabilistic sensitivity analysis results

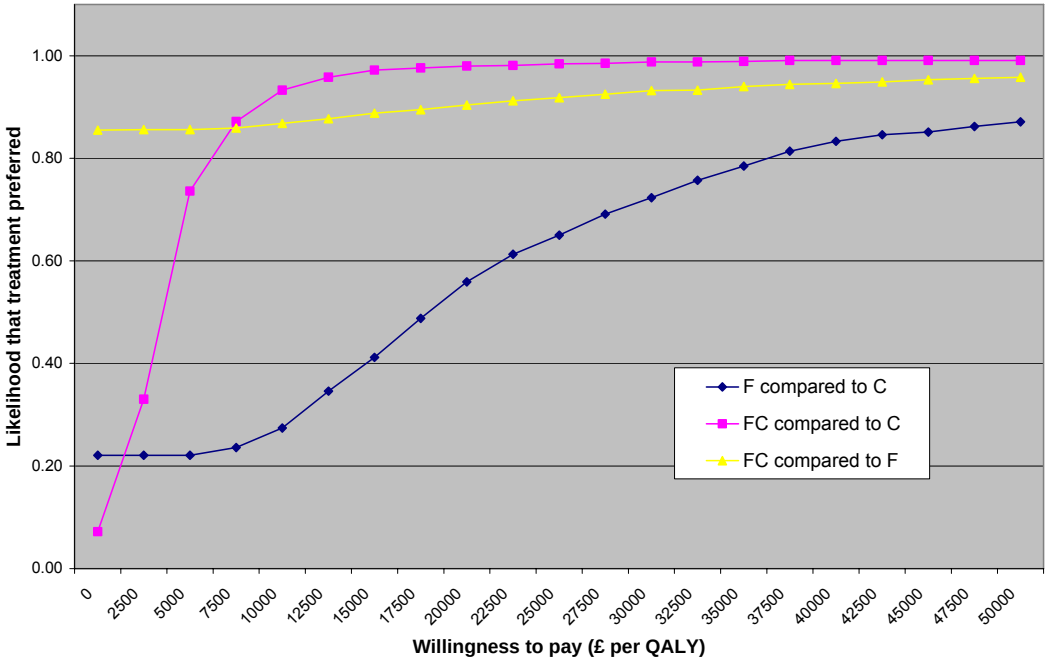
	Costs	**Outcomes**	
Chl (95% CI)	£11,920 (£10400, £13683)	5.25 (4.84, 6.49)	
F (95% CI)	£17,712 (£15587, £20532)	5.47 (4.95, 6.78)	
FC (95% CI)	£13,919 (£12301, £15760)	5.86 (5.26, 7.35)	
Incremental	**Costs**	**Outcomes**	**C/E ratio**
F vs Chl (95% CI)	£5,792 (£3250, £8909)	0.22 (-0.31, 0.73)	£26,105 (£-210608, £206922)
FC vs Chl (95% CI)	£1,999 (£-303, £4372)	0.62 (-0.09, 1.46)	£3,244 (£-9872, £14598)
FC vs F (95% CI)	-£3,793 (£-6786, £-1267)	0.39 (-0.36, 1.28)	-£9,616 (£-141972, £62966)

The central estimates of the cost-effectiveness ratio are identical to those presented in the manufacturer's deterministic analysis. It is clear that the probabilistic sensitivity analysis was, therefore, only used to estimate the confidence intervals around the mean estimates, as

opposed also to deriving the expected value itself using the probabilistic analysis. For non-linear models (i.e. Markov models) the correct estimate of the mean can only be obtained by taking the expected value of the outcomes across all the iterations. Furthermore, since the 95% confidence intervals presented in the table above also contain negative ICERs, the subsequent interpretation of these results is problematic since a negative ICER could mean that the intervention dominates the comparator (or vice-versa). To address these potential issues, the ERG have undertaken additional analyses to assist in the presentation of the probabilistic results. This work is reported in Chapter 6.

The submission also presents cost-effectiveness acceptability curves for the treatments. These are presented in the Figure 5.2 below:

Figure 5.2 Cost-effectiveness acceptability curves



The cost-effectiveness acceptability curves presented above represent pair-wise comparisons of the interventions (i.e. they show the probability of one drug being cost effective compared to another drug for different willingness to pay values). This is a potentially misleading way of presenting the results as, for any value of willingness to pay (i.e. cost-effectiveness threshold), only one intervention can be cost effective and thus the sum of the probabilities of each intervention being cost effective should sum to one. It would, therefore, be correct to show the cost-effectiveness acceptability curves based on a comparison of all treatments

simultaneously, such that the vertical sum of the curves at any value of willingness to pay is 1. An assessment of the decision uncertainty surrounding the simultaneous comparison of all treatments by the ERG is presented in the next chapter.

5.8 Summary of uncertainties and issues

In general, the ERG considered the manufacturer's economic submission to be of reasonable quality. The economic model structure (including the comparators) was considered appropriate for the decision problem, and the data sources used to inform the model were deemed appropriate from a UK NHS perspective. A range of sub-groups was considered and uncertainty in parameter estimates was addressed using probabilistic approaches. The manufacturer's submission contains a good description of the data sources and justification for the assumptions used for the electronic model. An electronic copy of the model was provided, which allowed a more detailed assessment of the model structure and the approaches used to populate individual parameters. However, a number of issues may compromise the validity of the model results, including:

- structural issues related to modelling of non-response to second-line treatments;
- the lack of existing evidence on the re-treatment response rate for FC;
- the lack of existing evidence on duration of response to re-treatment with FC, F and Chl;
- the potential for the choice of second-line treatment (which was modelled in a fixed way and the approaches used to synthesise evidence was considered crude) to influence the cost-effectiveness of the decisions related to first-line treatment;
- the lack of transparency in some of the costings and the potential differences between the audit population and the population in the CLL4 trial;
- the exclusion of adverse event costs from the analysis and the exclusion of the intercept term from the re-treatment costs;
- the use of non-preference-based quality of life values for certain parameters;
- the approach used to estimate transition probabilities from the CLL4 trial and the assumption that these were constant over the time horizon of the model; and
- the presentation of the results from the probabilistic sensitivity analysis.

Chapter 6
Additional work undertaken by the ERG

6.1 Overview

The ERG has undertaken additional work to address several of the issues and uncertainties identified during the structured critique of the manufacturer's submission in the previous chapter. This additional work has been undertaken to examine the potential robustness of the base-case results to several of the assumptions made in the manufacturer's model, and also to identify possible sources of bias. This work has been performed on the revised model presented in the manufacturer's addendum and can be separated into three main areas: (i) additional one-way sensitivity analyses to examine the robustness of the base-case ICER to alternative assumptions related to the response rate for re-treatment and the duration of this response; (ii) a more appropriate presentation of the probabilistic sensitivity analysis results from the submission; and (iii) formal survival analysis of the individual patient data from the CLL4 trial to explore the appropriateness of assuming constant transition probabilities to extrapolate over a lifetime time horizon.

It should be noted that these analyses are selective, and that the revised economic analyses have been undertaken to examine the robustness of the manufacturer's own model to alternative assumptions. These analyses are clearly subject to the same potential limitations outlined in previous sections regarding the structural assumptions, the general approach used to estimate transition probabilities and issues related to the modelling of second-line treatments. The results should, therefore, be taken as indicative of the potential impact on the potential cost-effectiveness estimates.

6.2 Additional sensitivity analyses undertaken by the ERG

A key assumption made in the manufacturer's submission was that the re-treatment response rate for FC was the same as the initial treatment response rate. This assumption was made on the basis that no other evidence was identified to inform this parameter estimate. While such an approach may be considered justifiable (in the absence of contradictory evidence from the literature), it does appear a strong assumption given that the evidence for the re-treatment response rates for F and Chl reported in the literature are both lower than the estimates used for first-line treatment. The ERG was also concerned that the choice of values used in the sensitivity analysis undertaken in the submission for FC was not sufficiently rigorous to test the

robustness of the model results, since it was based on the 95% confidence interval from the bootstrap of the initial treatment rate. The ERG has varied the potential probability of response to re-treatment with FC from 0.1 to 0.9 (compared to the base-case assumption of a probability of [redacted]) to determine how sensitive the cost-effectiveness results are to this parameter. The results are provided in Table 6.1.

Table 6.1 Results of the ERG's sensitivity analysis on the re-treatment response rate for FC

	Re-treatment response rate for FC								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
FC vs Chl	Dominated	Dominated	Dominated	£17,522	£8,083	£5,464	£4,235	£3,521	£3,055
FC vs F	£9,408	£12,134	£17,400	£31,859	£244,583	Dominates	Dominates	Dominates	Dominates

The results of the additional sensitivity analysis demonstrate that the re-treatment response rate would have to be significantly lower than that assumed for first-line treatment before this might result in a change to the decision related to the cost-effectiveness of FC. Indeed, the results suggest that the re-treatment response rate would have to fall to somewhere between 0.3 and 0.4 before FC no longer appears cost-effective in comparison to Chl.

The manufacturer's submission also assumes that the duration of response for re-treatment is equal to the duration of response for initial treatment. Again, this assumption was made based on the lack of evidence from current trials. However, the manufacturer's submission did not assess the robustness of their results to this assumption as part of their sensitivity analysis. The ERG has, therefore, undertaken a series of additional analyses to explore the robustness of the base-case results to this assumption. Following clinical advice, it was decided that only shorter periods of duration of response would be considered. In addition, since the ERG could not distinguish whether there were differences between treatments in terms of the duration of response to re-treatment, the same approach has been applied to all treatments. The additional analyses were undertaken by increasing the transition probability of moving to progression from the re-treatment response states. These probabilities were scaled between 1 and 2 (with 1 representing the base-case analysis and 2 representing a doubling in the transition probability, such that the median duration of survival is halved). The results for this analysis are presented in Table 6.2.

Table 6.2 Results of the ERG's sensitivity analysis on the re-treatment response duration

Scale increase in re-treatment response duration transition probability						
	1	1.2	1.4	1.6	1.8	2
F vs C	£26,105	£31,796	£38,544	£46,636	£56,482	£68,698
FC vs C	£3,244	£3,550	£3,883	£4,244	£4,635	£5,058
FC vs F	-£9,616	-£10,332	-£11,162	-£12,085	-£13,089	-£14,164

The impact of varying these transition probabilities by a constant factor do not appear to have a significant impact on the ICER for FC compared to Chl, although it does increase the ICER of F compared to Chl markedly. However, FC still continues to dominate F.

6.3 Probabilistic sensitivity analysis

The previous chapter outlined the limitations in the manufacturer's presentation of the results from the probabilistic sensitivity analysis. In particular, the ERG noted that the estimates of the ICER were based on the mean results from the deterministic analysis which is not appropriate for non-linear models. The ERG has, therefore, calculated the correct estimate of the ICER based on the expectation of the mean costs and QALYs based on 1000 iterations of the model. The results are presented in Table 6.3.

Table 6.3 Probabilistic sensitivity analysis results

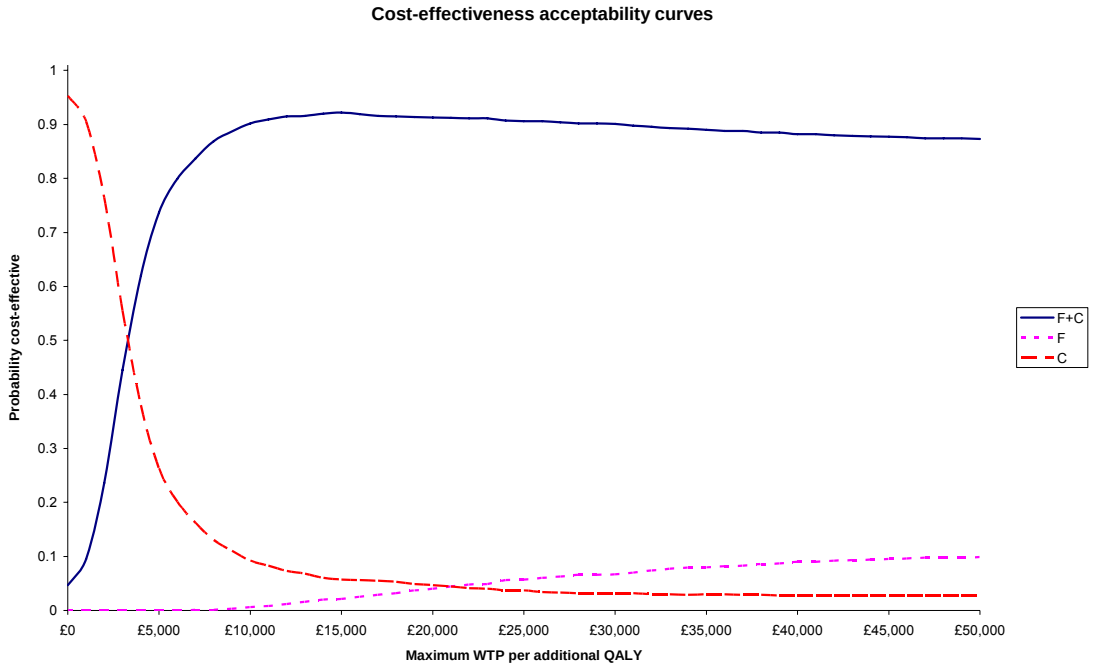
Comparator	Mean Costs	Mean QALYs	ICER (compared to Chl)	Probability cost-effective at WTP		
				£20,000	£30,000	£40,000
Chl	£11,836	5.48	-	0.047	0.032	0.028
F	£17,840	5.70	Dominated by FC	0.04	0.067	0.09
FC	£13,291	6.13	£3,213	0.913	0.901	0.882

As the table shows, there are only small changes in the results compared to those presented in the manufacturer's submission and re-estimating the ICER based on the expected values does not appear to significantly affect the cost-effectiveness estimates.

The ERG also noted that the CEACs presented in the manufacturer's submission are based

on pair wise comparisons of the three treatments. As only one intervention can be cost effective for a given willingness to pay threshold, it is more appropriate to compare all 3 interventions simultaneously such that the sum of the probabilities of each of them being cost effective for a given willingness to pay is equal to one. This analysis is presented in Figure 6.1. The probabilities that each treatment is cost-effective for a range of thresholds are also presented in Table 6.3. At a willingness to pay of £30,000 per QALY the probability that FC is cost-effective is 0.901.

Figure 6.1 Cost-effectiveness acceptability curves- revised by the ERG



6.4 Survival analysis

A central criticism of the manufacturer's submission by the ERG relates to the approaches used to estimate relevant transition probabilities based on the individual patient data from the CLL4 trial. A number of specific issues were identified by the ERG; however, due to the current structure of the electronic model, it was not considered possible to explore the robustness of the manufacturer's results to alternative approaches to estimating the transition probabilities. However, a key assumption within the manufacturer's submission is that the transition probabilities remain constant over the time horizon of the model. This implies a constant hazard (i.e. following an exponential distribution), which is a strong assumption and could be influencing the cost-effectiveness results. The ERG noted the divergence between the cost-effectiveness results presented for alternative time horizons and outlined that the extrapolation

approach is likely to be a key determinant of the final cost-effectiveness estimates. In order to assess the validity of the assumption of constant hazards, the ERG has undertaken additional analysis on the individual patient data contained in the manufacturer's submission using survival analytic techniques. This analysis was undertaken to formally test the underlying assumption of a constant hazard and to deal with censored data.

6.4.1 Results of the survival analysis

The ERG has separated the data into two groups: responders and non-responders. Two separate transitions were considered representing the transitions to progression and death following first-line treatment. A separate Weibull distribution has been fitted to both groups, and to each separate transition, to test the assumption of a constant hazard assumed in the submission. The Weibull distribution has the following probability density function:

$$f(t) = \lambda\gamma t^{\gamma-1} \exp\{-\lambda t\gamma\}$$

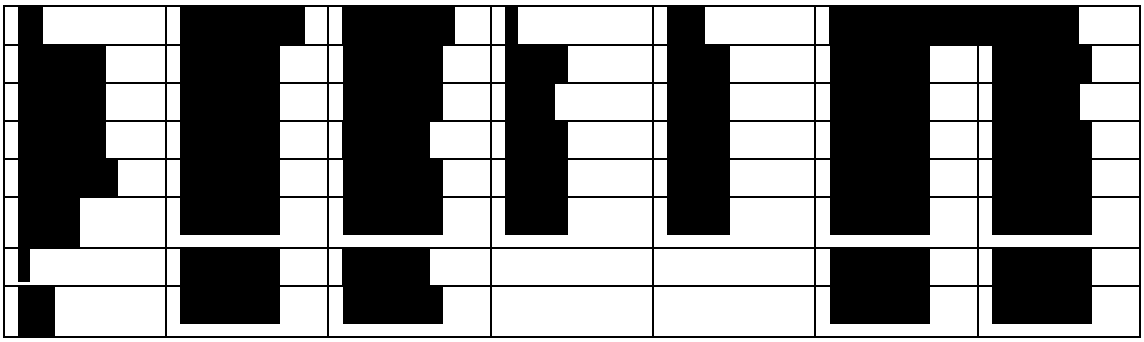
which is characterised by the two parameters λ and γ and time t. This gives a hazard function of:

$$h(t) = \lambda\gamma t^{\gamma-1}$$

When γ=1 the Weibull distribution reduces to the exponential distribution and hence would indicate that the assumption of constant transition probabilities would be appropriate. Therefore, the ERG has tested the value of the parameter γ for the two groups using survival analysis undertaken using the statistical package STATA.

Table 6.4 reports the results for the responders group for the transition to progression.

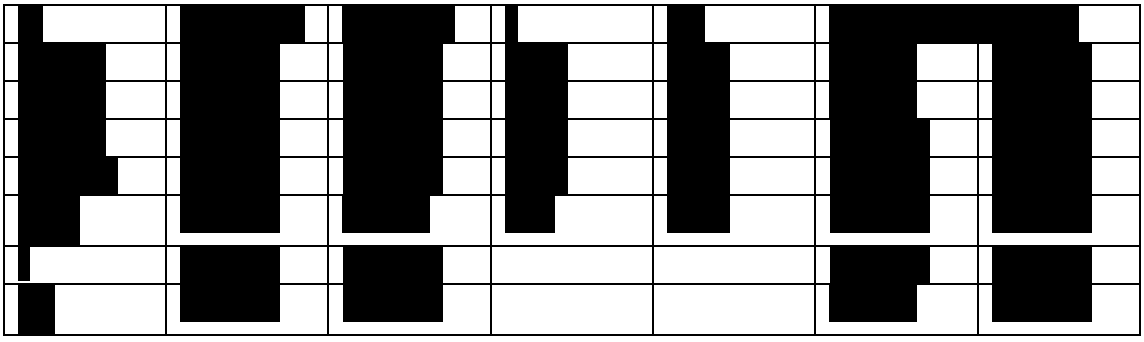
Table 6.4 Survival analysis results for progression (responders) undertaken by the ERG



The t test on ln(γ) is the test for the exponential distribution (i.e. a null hypothesis of γ=1) . Clearly here we have to reject the null hypothesis exponential model at 0.0000 significance (as ln(γ) does not equal zero and hence γ does not equal 1) and, therefore, we can reject the assumption of a constant hazard. As γ is greater than one, it implies that the hazard of progression increases with time.

Table 6.5 reports the results based on data from non-responders (those who responded for less than 12 months to initial therapy). These results imply that the assumption of exponential distribution can be rejected at the 0.1% significance level. Again as the value for p is greater than one it implies that the hazard of progression increases with time.

Table 6.5 Survival analysis results for progression (non-responders)



A similar analysis was undertaken for the transition to death. Again, the data have been split into groups of responders and non-responders and Weibull distributions have been fitted using STATA. The results for responders are presented in Table 6.6 and for non-responders in Table 6.7.

Table 6.6 Survival analysis results for death (responders)

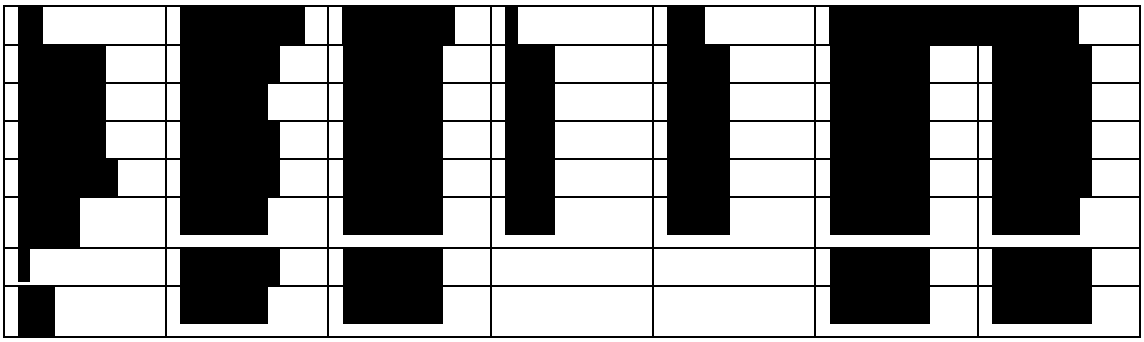
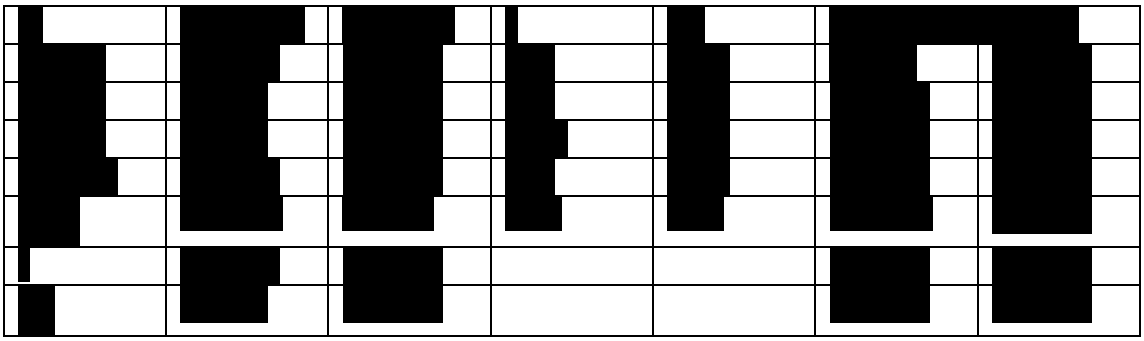


Table 6.7 Survival analysis results for death (non-responders)



For the responders the exponential model is rejected at a 0.001% significance level and, therefore, we can again reject a constant hazard and thus transition probability. As γ is greater than one, it again implies that the hazard of death increases with time. For the non-responders, the exponential distribution cannot be rejected at a 5% significance level but at a 10% significance level it would be rejected.

The results of the survival analysis undertaken by the ERG suggest that the assumption of constant transition probabilities do not appear to be appropriate based on the individual patient level data from the CLL4 trial. Since the assumption of constant transition probabilities underpins the manufacturers approach to extrapolation, then the subsequent findings must be interpreted accordingly. Given that FC has the highest initial response rate (and hence a larger proportion of this group continues to follow the transitions based on the CLL4 trial data), it is clear that the impact of assuming constant transition probabilities may be acting as a possible source of bias towards this group.

Chapter 7
Discussion and conclusions

7.1 Summary of clinical effectiveness issues

The manufacturer's submission was considered to comprise the most relevant clinical effectiveness evidence for the purpose of this STA. Seven studies were included in the submission, of which two were fully published and five were available in abstract form only. Fludarabine and fludarabine plus cyclophosphamide were compared with chlorambucil (Chl) in five studies and two compared fludarabine with fludarabine plus cyclophosphamide. The CLL4 study was presented as the pivotal source of evidence as this was the only study to compare all three regimens in the same patient population. Supplemental 'academic in confidence' individual patient data were presented by the manufacturer alongside the published abstract of the CLL4 trial.

The majority of studies showed an improvement in overall response in those patients receiving fludarabine or fludarabine plus cyclophosphamide compared to those receiving the comparator chlorambucil. In the single study comparing median progression-free survival between the fludarabine and chlorambucil regimens there was a significantly longer duration of response in the fludarabine arm. A significantly longer duration of response was also seen when fludarabine plus cyclophosphamide was compared to fludarabine alone. At present, the follow-up period of the studies included in the submission are too short to demonstrate any significant improvement in overall survival. Therefore, fully matured survival data are necessary to ascertain whether any improvement in progression-free survival translates in to an increase in overall survival.

The fludarabine and fludarabine plus cyclophosphamide regimens were associated with a significantly higher incidence of neutropenia and thrombocytopenia and non-haematological toxicity compared to chlorambucil. However, limited quality of life data from the CLL4 study showed a similar quality of life for each treatment group at 12 months. It is anticipated that the results of further analyses are likely to become available within the next year.

Because five of the studies included in the submission are not fully published and report only preliminary results in abstract form the ERG felt they provided insufficient data fully to assess the validity of these studies with regards to the decision problem. Although the ERG deemed the unpublished CLL4 study to be the most relevant source of evidence for the decision problem, the supplemental individual patient data [10] provided by the manufacturer

are not in the public domain and, therefore, cannot be verified externally. Therefore, until the CLL4 and other ongoing studies are fully published and the complete data made available for evaluation, the results must be interpreted with caution.

7.2 Summary of cost-effectiveness issues

Previously published studies were not considered particularly relevant by the ERG due to the limited clinical and economic evidence on which the studies were based and the restricted range of comparators considered. The submission by the manufacturer was considered to comprise the most relevant evidence to consider for the purposes of this STA.

The manufacturers submission included on a '*de-novo'* decision analytic Markov model to estimate the cost-effectiveness of treatment with (i) fludarabine monotherapy (F), (ii) fludarabine in combination with cyclophosphamide (FC) and (iii) chlorambucil (Chl). The economic model structure (including the comparators) was considered appropriate for the decision problem, and the data sources used to inform the model were deemed appropriate from a UK NHS perspective. The results from the manufacturers demonstrated that FC was cost-effective compared to Chl under a broad range of different assumptions. FC was shown to dominate F. Sub-group analyses by age and Binet stage did not alter these conclusions.

The ERG identified a number of potential issues related to the manufacturer's economic submission. One issue related to the approach used to equalise survival across the different regimens was subsequently addressed by the manufacturer as part of an addendum. However, the ERG identified a number of sources of possible bias which may over-state the cost-effectiveness of the FC regimen. While the manufacturer's results appeared robust to several of these issues as part of the additional sensitivity analyses undertaken by the ERG, there remained a number of issues which the ERG could not adequately explore due to the current model structure. In particular, the ERG was concerned with the approach the manufacturer used to estimate a number of key probabilities derived from the CLL4 trial data which could affect the validity of the assumptions underpinning the extrapolation of data over a lifetime time horizon. In particular, the assumption of constant transition probabilities assumed by the manufacturers did not appear to be supported by the data. Given the sensitivity of the cost-effectiveness results to the time-horizon, additional sensitivity analyses on the transition probabilities would help to determine the robustness of the current model to this issue.

7.3 Implications for future research

In order to allow an accurate assessment of the clinical and cost-effectiveness of fludarabine as first-line treatment of chronic lymphocytic leukaemia, there is clearly a need for further evidence to clarify those areas of uncertainty outlined in the report. Until more mature data from the CLL4 trial are available, the ERG is unable to suggest a great deal about what future research would be valuable in helping to inform the decision problem. Currently the major uncertainties are focused around overall survival, second-line treatment response rates (in particular the re-treatment with FC response rate) and the duration of response to second-line treatments. Further evidence on the incidence, severity and duration of treatment-related adverse events is required in patients given prophylaxis to assess the impact of fludarabine treatment on quality of life, especially in older patients.

The lack of direct utility measurements in the population of interest from an appropriate preference-based instrument is also an area of uncertainty. More evidence on utility values for the states, and any differences in utility values of states across treatments, would help to inform the decision problem. There is also a wider evidence base available to inform the decision than just the CLL4 trial. Many other pair-wise comparisons of treatments from trials could be synthesised with the CLL4 trial data using mixed treatment comparison methods to further inform the decision problem.

References

Endnote 9

1. Schering Health Care Ltd, *Fludarabine phosphate for the first-line treatment of Chronic Lymphocytic Leukaemia*. 2006.
2. Schering Health Care Ltd, *Fludara Summary of Product Characteristics.* 2005, Electronic Medicines Compendium.
3. Eichhorst, B.F., et al., *Fludarabine plus cyclophosphamide versus fludarabine alone in first-line therapy of younger patients with chronic lymphocytic leukemia.* Blood, 2006. **107**(3): p. 885-91.
4. Rai, K.R., et al., *Fludarabine compared with chlorambucil as primary therapy for chronic lymphocytic leukemia.* N Engl J Med, 2000. **343**(24): p. 1750-7.
5. Catovsky, D., S. Richrads, and P. Hillmen, *Early results from LRF CLL4: A UK Multicentre Randomised Trial.* ASH Annual Meeting Abstracts 2005. **106**(11): p. 716.
6. Eichhorst, B.F., et al., *Comparison of the Efficacy and Toxicity of Fludarabine (F) in First Line Therapy of YoungerVersus Elderly Patients (Pts) with Advanced Chronic Lymphocytic Leukemia (CLL): Results of aMeta-Analysis of Two Phase III Trials of the German CLL Study Group (GCLLSG). Abstract #717.* Blood 2005. **106**(11).
7. Flinn, I.W., et al., *Fludarabine and Cyclophosphamide Produces a Higher Complete Response Rate and More Durable Remissions than Fludarabine in Patients with Previously Untreated CLL: Intergroup Trial E2997.* ASH Annual Meeting Abstracts 2004. **104**(11): p. 475.
8. Karlsson, K., et al., *Cladribine (CdA) or Fludarabine (F) or High-Dose Intermittent Chlorambucil (Chl) as First-Line Treatment of Symptomatic ChronicLymphocytic Leukemia? First Interim Analysis of Data from the InternationalRandomized Phase III Trial. Abstract #3470.* Blood, 2004. **104**(11).
9. Spriano, M., et al., *Muliticentre prospective randomized trial of fludarabine versus chlorambucil and prednisone in previously untreated patients with active B-cell CLL: final report. Abstract #086.* Haematology and Cell Therapy, 2000. **42**(1): p. 93.
10. CLL4 patient-level dataset 2006, *Data points from first randomisation to second-line randomisation (or last visit if still in remission, or death, if sooner) by January 2006.* 2006.
11. Best, L., *Fludarabine in the treatment of chronic lymphocytic leukaemia*. 1995, South and West Regional Health Authority: Bristol.
12. Hancock, S., B. Wake, and C. Hyde, *Fludarabine as first line therapy for chronic lymphocytic leukaemia*. 2002, West Midlands Health Technology Assessment Collaboration: Birmingham.
13. Summerfield, G.P., *Teleconference with Dr Geoffrey Summerfield, Consultant Haematologist. 27th June 2006*. 2006.
14. Centre for Cancer Education. University of Newcastle upon Tyne, *CancerWEB online medical dictionary. http://cancerweb.ncl.uk/cgi-bin/omd Accessed 24th May 2006.*
15. Oscier, D., et al., *Guidelines on the diagnosis and management of chronic lymphocytic leukaemia.* Br J Haematol, 2004. **125**(3): p. 294-317.
16. Warell D.A, Cox T.M, and Firth J.D, *Oxford textbook of medicine, 4th Edidtion.* 2003, Oxford.
17. Hallek, M., R.A. Stahel, and R. Greil, *ESMO Minimum Clinical Recommendations for diagnosis, treatment and follow-up of chronic lymphocytic leukemia.* Ann Oncol, 2005. **16 Suppl 1**: p. i50-1.
18. Welsh Cancer Intelligence and Surveillance Unit, *Cancer incidence 2003 report.*
19. Office for National Statistics, *Cancer registrations in England 2003.*

20. Office for National Statistics, *Components of population change (Constituent countries of the United Kingdom): Population Trends 123, report D9255.*
21. Shanafelt, T.D. and T.G. Call, *Current approach to diagnosis and management of chronic lymphocytic leukemia.* Mayo Clin Proc, 2004. **79**(3): p. 388-98.
22. Rai, K.R., et al., *Clinical staging of chronic lymphocytic leukemia.* Blood, 1975. **46**(2): p. 219-34.
23. *Chronic lymphocytic leukaemia: proposals for a revised prognostic staging system. Report from the International Workshop on CLL.* Br J Haematol, 1981. **48**(3): p. 365-7.
24. Binet, J.L., et al., *A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis.* Cancer, 1981. **48**(1): p. 198-206.
25. Gentile, M., et al., *New developments in the diagnosis, prognosis and treatment of chronic lymphocytic leukemia.* Curr Opin Oncol, 2005. **17**(6): p. 597-604.
26. Rummel, M.J., et al., *Fludarabine and epirubicin in the treatment of chronic lymphocytic leukaemia: a German multicenter phase II study.* Ann Oncol, 1999. **10**(2): p. 183-8.
27. Abbott, B.L., *Chronic lymphocytic leukemia: recent advances in diagnosis and treatment.* Oncologist, 2006. **11**(1): p. 21-30.
28. Baxter Healthcare Ltd, *Endoxana Summary of Product Characteristics*. 2004, Medicines.org.uk.
29. Phramacia Ltd, *Cyclophosphamide Summary of Product Characteristics*. 2005, Electronic Medicines Compendium.
30. Mehta D.K. Ed, *British National Formulary*. Vol. 51. 2006, London: British Medical Association/Royal Pharmaceutical Society of Great Britain.
31. *IMS Oncology Analyzer*. 2005, IMS Health.
32. Hillmen P, T.T. Skotnicki A, and e. al., *Preliminary Phase III Efficacy and Safety of Alemtuzumab vs Chlorambucil as Front-Line Therapy for Patients with Progressive B-Cell Chronic Lymphocytic Leukemia (BCLL).* . Proceedings from the 42nd annual meeting of the American Society of Clinical Oncology. Atlanta, Ga June, 2006. Abstract # 6511.
33. Genzyme Europe BV, *Campath Summary of Product Characteristics*. 2005, Electronic Medicines Compendium.
34. *International Phase III-Trial in B-cell CLL - Protocol. http://www.cck.ki.se/CLL-study/index.html (Accessed 06/07/06).*
35. Catovsky, D., *CLL4. Chronic Lymphocytic Leukaemia Trial 4 A LRF sponsored trial from the MRC Adult Leukaemia Working Party. MREC (1) 98/101. Protocol Web 1999.*
36. Johnson, S., et al., *Multicentre prospective randomised trial of fludarabine versus cyclophosphamide, doxorubicin, and prednisone (CAP) for treatment of advanced-stage chronic lymphocytic leukaemia. The French Cooperative Group on CLL.* Lancet, 1996. **347**(9013): p. 1432-8.
37. Leporrier, M., et al., *Randomized comparison of fludarabine, CAP, and ChOP in 938 previously untreated stage B and C chronic lymphocytic leukemia patients.* Blood, 2001. **98**(8): p. 2319-25.
38. Jaksic, B., et al., *Fludarabine versus high-dose continuous chlorambucil in untreated patients with B-CLL1 EORTC randomised trial.* Hematology & Cell Therapy, 2000. **42**(1): p. 97.
39. Zhu, Q., et al., *Fludarabine in comparison to alkylator-based regimen as induction therapy for chronic lymphocytic leukemia: a systematic review and meta-analysis.* Leuk Lymphoma, 2004. **45**(11): p. 2239-45.

40. Richards, S., *Fludarabine increases complete response but not survival compared with conventional alkylator-based regimens for previously untreated chronic lymphocytic leukaemia.* Cancer Treat Rev, 2005. **31**(4): p. 332-5.
41. Drummond, M.F., et al., *Methods for the Economic Evaluation of Health Care Programmes*. 2005: Oxford Medical Publications.
42. Hutchinson, J., et al., *Costs of chemotherapy for indolent follicular non-Hodgkin's lymphoma in the UK: an observational study.* Acta Haematol, 2006. **115**((1-2)): p. 28-34.
43. Phillips, Z., et al., *Review of guidelines for good practise in decision analytic modelling in health technology assessment.* Health Technology Assessment, 2004. **9**(36).
44. Doorduijn, J., et al., *Self-reported quality of life in elderly patients with aggresive non-Hodgkin's lymphoma treated with CHOP chemotherapy.* Eur J Haematol, 2005. **75**(2): p. 116-123.
45. Holzner, B. and e. al, *Quality of Life measurement in oncology - a matter of the measurement instrument?* European Journal of Cancer, 2001. **37**: p. 2349-2356.
46. Eccles, M. and J. Mason, *(2001) How to develop cost-conscious guidelines.Health Technology Assessment 5 (16): 1–69.*

Appendix 1: Table displaying the Binet and Rai staging system for CLL. [22, 24]

Staging System	Features	% of patients
Binet stage		
A	<3 lymphoid areas	[redacted]
B	>3 lymphoid areas	
C	Haemoglobin <10g/dl or platelets <100 x 10^9/l	[redacted]
Rai stage		
0 (low level risk)	Lymphocytosis only	[redacted]
I (intermediate level risk)	Lymphadenopathy	
II (intermediate level risk)	Hepato or splenomegaly, with or without lymphadenopathy	[redacted]
III (high level risk)	Haemoglobin <11g/dl	[redacted]
IV (high level risk)	Platelet <100 x 10^9/l	

Appendix 2: Search strategy undertaken by ERG for fludarabine STA for the clinical effectiveness literature review.

Inclusion criteria:

Participants: Chronic Lymphocytic Leukaemia

Interventions*:* Fludarabine

Comparator: Any

Outcomes: No restrictions applied (outcomes included; OS, PFS, OR, CR, PR, ADRs & QoL)

Design: RCT

Exclusion criteria:

Participants: Previously treated patients

Intervention*:* None

Study selection: Peer review panel

Database: PUBMed (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi)

Host: NHS Trust Internet

Date search run: 03/07/2006

Date span of search: MEDLINE (1966-January 2006), OLDMEDLINE (1950 through 1965)

Database: Embase (EMZZ)

Host: Dialog DataStar

Date search run: 05/07/2006

Date span of search: 1974 to January 2006

Search strategy:

#1. Search “Leukemia, B-Cell” [MeSH]
#2. Search cll[Title/Abstract] OR b-cll[Title/Abstract]
#3. Search chronic AND lymphocytic AND leukemia[Title/Abstract]
#4. Search chronic AND lymphocytic AND leukaemia[Title/Abstract]
#5. Search #1 OR #2 OR #3 OR #4
#6. Search “Study Characteristics”[Publication Type]
#7. Search “Single-Blind Method”[MeSH]
#8. Search “Double-Blind Method”[MeSH]
#9. Search “Cross-Over Studies”[MeSH]
#10. Search “Follow-Up Studies”[MeSH]
#11. Search “Evaluation Studies”[MeSH]
#12. Search “Epidemiologic Study Characteristics”[MeSH]
#13. Search “Prospective Studies”[MeSH]
#14. Search #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13
#15. Search Study[Title/Abstract] OR Trial[Title/Abstract]
#16. Search #14 OR #15
#17. Search #5 AND #16 Limits: Humans
#18. Search naive Field: Title/Abstract, Limits: Humans
#19. Search untreated Field: Title/Abstract, Limits: Humans
#20. Search first AND line Field: Title/Abstract, Limits: Humans
#21. Search first-line Field: Title/Abstract, Limits: Humans
#22. Search initial Field: Title/Abstract, Limits: Humans
#23. Search #18 OR #19 OR #20 OR #21 OR #22 Field: Title/Abstract, Limits: Humans
#24. Search #23 AND #17 Field: Title/Abstract, Limits: Humans
#25. Search fludarabine OR cyclophosphamide OR chlorambucil Field: Title/Abstract, Limits: Humans
#26. Search #24 AND #25 (214 hits)
#27. Search "Randomized Controlled Trials"[MeSH] Limits: Humans
#28. Search "Randomized Controlled Trial"[Publication Type] Limits: Humans
#29. Search "Controlled Clinical Trials"[MeSH] Limits: Humans
#30. Search randomized OR randomised Field: Title/Abstract, Limits: Humans
#31. Search #27 OR #28 OR #29 OR #30
#32. Search #26 AND #31 Field: Title/Abstract, Limits: Humans

Appendix 3: Structured critical appraisal of Catovsky 2005 [5]

CRITICAL APPRAISAL

Name of Trial: Early results from LRF CLL4: A UK Multicenter Randomized Trial

Reference: Catovsky D, Richards S, Hillmen P. Blood (ASH Annual Meeting Abstracts) 2005;106:Abstract 716.

Question: Is the combination of fludarabine and cyclophosphamide (FC) associated with a better response in adults with chronic lymphocytic leukaemia to that with either fludarabine (F) alone or chlorambucil (Chl) alone.

Summary: Whilst early results seem promising until the results are fully published the magnitude of response is unclear as are those patients groups who might benefit most.

Did the study ask a clearly focussed question?

Can't tell
This is presented as written abstract from an oral conference session presenting early results although the trial protocol is available separately[1]. Details relating to the population studied, the interventions and the primary outcome considered are limited.

Was the study design appropriate?

Yes
It would be appropriate to carry out this research as a RCT. No detail is given in the abstract as to the choice of control arm (chlorambucil) or the choice of combination cyclophosphamide) although this is discussed in the protocol.

Were participants appropriately allocated to intervention and control groups?

Can't tell
The abstract states that 783 patients were randomised with 6 exclusions but does not state the method of randomisation or the reason for exclusion. No detail is given around the balancing of groups with respect to population characteristics. Overall the male: female ratio was 2.8:1 and the distribution by Binet stages was A progressive 25%, B 45% and C 30%. One third of cases were aged <60 years and one third 70 years or over. No mention is made of the remaining patients ages.

Were participants, staff and study personnel 'blind' to participants study group?

No mention is made of blinding within the abstract however the protocol states that patients and clinicians were not blinded[1]. Responses were made by bone marrow biopsy and it is unclear from the abstract whether interpretation of results was carried out by blinded staff.

Were all of the participants who entered into the trial accounted for at its conclusion?

No
At the point of reporting these results, data from only 661 patients was available. No information is given as to the reasons for lack of data from the remaining 116 patients or the impact this may have had on results. A significantly higher proportion of patients in the chlorambucil arm (78/387 or 20%) did not have data available compared to F alone (18/194, 9%) or the FC combination (20/196, 10%). This may have the effect of under estimating the effectiveness of the control arm.
No information is given with respect to analysis using an intent to treat basis.

Were the participants in all groups followed up and data collected in the same way?

Can't tell
A lack of detail in the abstract prevents further comment

Was the study large enough?

Can't tell
No detail is given of any power calculation in the abstract however this is included in the protocol. 500 patients (250 allocated to Chl and 250 to F based treatment) would provide more than 90% power to detect an absolute difference of 15%, from 40% to 55% in survival at 5 years using a 2-sided p-value. There would be 65% power to detect a difference of 10%. This would also be the power in detecting a difference between the FC and F arms.

How are the results presented and what is the main result?

Analysis of progression-free survival showed fewer events with FC (O/E) 0.5) than F (O/E 1.1) and chlorambucil (O/E1.3). F + FC v Chl $p<0.00005$; FC vs F $p<0.0005$. PFS at 3 years was reported to be 23% Chl, 31% F and 62% FC. No difference is reported between the groups for overall survival. Furthermore younger patients appear to benefit more from treatment across all treatment groups.

How safe were the regimens?
More neutropenia was reported with FC (55%) than F (40%) and Chl (29%) with an increased number of hospitalisations in the fludarabine-containing groups. Conversely, more haemolytic anaemias were reported in the Chl group (13%) compared with F (10%) and FC (4%). More nausea/vomiting and alopecia were reported with FC than other regimens although precise figures are not quoted.

How precise are the results?

Confidence intervals are not reported although p values are for all fludarabine containing combinations vs Chl and FC vs F. It is difficult to draw firm conclusions from the data provided.

Can the results be applied to the local population?

The population recruited into this trial appear to reflect the population in which the treatments would be used in the UK. 88% of patients were from the UK and the male:female and age balance would seem appropriate. It is not clear though whether the groups were balanced once outcomes were analysed and until this data is available it is not certain the results can be applied to the general population of patients with CLL.

Appendix 4: Structured critical appraisal of Eichhorst 2006 [3]

CRITICAL APPRAISAL

Name of Trial: Fludarabine plus cyclophosphamide versus fludarabine alone in first-line therapy of younger patients with chronic lymphocytic leukemia

Reference: Eichhorst BF, Busch R, Hopfinger G et al. Blood 2006; 107:885-891

Question: Is fludarabine plus cyclophosphamide superior to fludarabine alone as first-line therapy for patients younger than 66 years with previously untreated advanced chronic lymphocytic leukaemia (CLL)?

Summary: In patients aged 65 years or younger with previously untreated advanced CLL, complete remission and partial response rates were significantly higher (both $p \leq 0.001$) with fludarabine plus cyclophosphamide (23.8% and 94.5%, respectively) than fludarabine alone (6.7% and 82.9%, respectively). The combination resulted in longer progression-free survival (by 28 months) and treatment-free survival (by 12 months), but at the expense of higher incidences of grades 3 and 4 toxicity, particularly myelotoxicity. Overall survival with the two regimens did not differ significantly, but the follow-up duration was too short (median 22 months) for this end point to be reached.

Did the study ask a clearly focussed question?

Yes – This trial was designed to compare the effectiveness of fludarabine alone (F) and in combination with cyclophosphamide (FC) in previously untreated patients with predominantly advanced chronic lymphocytic leukaemia (CLL). The population studied, interventions given and outcomes considered are clearly stated.

Population: Patients aged 18-65 years with chronic lymphocytic leukaemia (CLL), at Binet stages C, B if they had rapid disease progression or symptoms or severe B symptoms, and A if they had B symptoms, who had not received previous treatment for CLL, had a life expectancy of at least six months and an Eastern Cooperative Oncology Group performance status of 0, 1 or 2. The majority (about 90%) had advanced disease (Binet stage B or C).
Intervention: F alone (25 mg/m^2 intravenously (iv) daily over 30 minutes for five days) or FC (F 30 mg/m^2 plus cyclophosphamide 250 mg/m^2 both iv daily over 30 minutes for three days). Both regimens were repeated every 28 days for a maximum of six courses.
Outcomes considered: Clinical responses (defined according to the guidelines of the guidelines of the National Cancer Institute (NCI)-sponsored workshop[1]) were calculated for all patients who received at least one cycle of therapy, overall survival (OS, time from randomisation to death), progression-free survival (PFS, time from randomisation to time of disease progression or death) and treatment-free survival (TFS, time from end of therapy to time of second-line treatment or death). Clinical response appears to be the primary outcome, although the authors do not specify which of the outcomes are primary or secondary end points.

Was the study design appropriate?

Yes - A randomised controlled trial design was appropriate for this prospective, multicentre, phase 3, trial comparing the FC regimen with F monotherapy.

Were participants appropriately allocated to intervention and control groups?

Probably – Patients were randomised to receive F or FC by the Institute of Medical Statistics and Epidemiology, Technical University, Munich, Germany. No details of the randomisation method are given. The authors state that comparison of patients in the two groups indicated no significant difference regarding the main clinical features and the risk categories, but give no details of statistical analysis or p values. The baseline characteristics of the patients look similar.

Were participants, staff and study personnel 'blind' to participants study group?

Can't tell - Blinding is not mentioned. As the criteria for the outcomes are clearly defined, observer bias is unlikely.

Were all of the participants who entered into the trial accounted for at its conclusion?

Yes - Three-hundred-and-seventy-five patients were randomly assigned to therapy. Thirteen were excluded due to violations of inclusion criteria, leaving 362, of whom 182 were assigned to F and 180 to FC. Eleven patients were lost to follow-up, survival data were available for 351 patients, response data for 328 and toxicity data for 346. Thirty-seven patients died (20 and 17 in the FC and F arms, respectively). A median of six courses was administered in both treatment arms, with 70.7% and 64% of patients completing six courses in the F and FC arms, respectively. Fifty-one patients (28.0%) withdrew from the F arm due to non-response (33%), autoimmune haemolysis (23%) and toxicity (14%) and 63 patients stopped FC early due to toxicity (30%, partial or complete response (13%) or non-response (9%).

Were the participants in all groups followed up and data collected in the same way?

Yes - After every course of therapy, patients were evaluated by clinical examination and blood count. After three and six courses of chemotherapy, responses were assessed by clinical examination, blood count, serum chemistry profile, ultrasound (US) examination or computed tomography (CT) and bone marrow biopsy to confirm a complete response. During follow-up, response was assessed every three months by clinical examination, blood count and, if clinically indicated, US examination.

Was the study large enough?

Can't tell - The authors state that statistical analysis was performed on an intent-to-treat basis and included the eligible patients. No power calculation is presented. Time to event was estimated using the Kaplan-Meier method, and treatment comparison was tested with the log-rank test. Response rates were calculated for all patients who received at least one cycle of therapy. Treatment arms were compared by the X^2 test. All statistical tests were two-sided and statistical significance was defined as $p < 0.05$. Confidence intervals are not given.

How are the results presented and what is the main result?

Response rates: Response data were available for 164 patients from each treatment arm and are quoted as percentages of these 164 patients, rather than as percentages of the 182 and 180 randomised to F and FC, respectively. FC treatment induced more complete remissions than F alone (23.8% vs. 6.7%; $p < 0.001$) and a higher rate of overall responses (94.5% vs. 82.9%; $p = 0.001$).

Overall survival: Median follow-up duration was 22 months. OS in the two arms did not differ significantly (175 assigned to F and 176 to FC were evaluable for overall survival, with 17 and 20, respectively, dying). Deaths were CLL-related in 15/20 in the FC arm and 9/17 in the F arm ($p = 0.51$). Three-year survival rates for these patients were 80.7% in the F arm and 80.3% in the FC arm. Median OS was not reached in either arm.

Progression-free survival: Median PFS was significantly longer with FC than F (48 vs. 20 months, $p = 0.001$). Of the evaluable patients, 79/171 (46.2%) in the F arm and 53/168 (31.5%) in the FC arm had progressive disease.

Treatment-free survival: Median TFS was longer with FC (37 months, $n = 175$) than F (25 months, $n = 169$; $p < 0.001$).

Second-line treatment: Thirty-two FC-treated and 59 F-treated patients received second-line therapy for relapse.

How safe were the regimens?

Toxicity data were available for 173 patients in each group. Five patients died as a result of treatment, two (1.2%) in the FC arm (one each of severe autoimmune haemolytic anaemia and tumour lysis syndrome) and three (1.7%) in the F arm (one each of pneumonia with sepsis, cerebral

bleeding caused by thrombocytopenia and autoimmune haemolytic anaemia). Myelotoxicity of all grades was the major adverse effect in both arms. Myelotoxicity, particularly leucocytopenia, was significantly more frequent in patients given FC than F, although the severe and opportunistic infection rates of the two groups did not differ significantly. Gastrointestinal tract (GIT) side effects, such as nausea, vomiting, mucositis and gastritis were more common with FC than F. The severe adverse effects (classified according to the Common Toxicity Criteria (CTC 1.0) as grades 3 and 4 and NCI grades 3 and 4) experienced are shown below in Table 1.

. patients	3	3	
C grades 3 & 4			
	0%	6%	.001
elotoxicity	3%	2%	.001
ucocytopenia	0%	5%	.001
aemia	6%	%	.28
toimmune haemoly aemia	%	%	.37
rombocytopenia	7%	6%	.44
ection	%	%	.999
T side effects	%	%	.05
I grades 3 & 4			
aemia	%	%	.17
rombocytopenia	3%	9%	.02

Table 1 Patients with at least one grade 3 or 4 adverse effect

How precise are the results?

The FC regimen resulted in a significantly higher complete remission rate (23.8%) than F (6.7%, $p < 0.001$) and a significantly higher overall response rate (94.5%) than F (82.9%, $p = 0.001$). OS of the two groups did not differ significantly, but the median follow-up duration was short, only 22 months, which would account for this. Median PFS was 24 months longer with FC than F ($p = 0.001$), and TFS was 12 months longer with FC ($p <0.001$). However, the FC regimen was associated with significantly higher incidences of all CTC grades 3 and 4 toxicities, grades 3 and 4 myelotoxicity and leucocytopenia and borderline significantly ($p = 0.05$) more GIT grades 3 and 4 adverse effects. NCI grades 3 and 4 and thrombocytopenia occurred significantly more frequently with FC than F alone. There were no significant differences between the treatments with respect to other adverse effects, including the incidence of severe and opportunistic infections

Can the results be applied to the local population?

This study was carried out in Germany. Disease status was classified according to standard systems and disease response was determined according to NCI criteria, as in other trials in patients with CLL. Therefore, there are unlikely to be marked differences from the UK population with CLL that would mean the results can not be extrapolated to UK patients with CLL who meet the trial criteria.

Summary

In patients aged 65 years or younger with previously untreated advanced CLL, complete remission and partial response rates were significantly higher (both $p \leq 0.001$) with FC (23.8% and 94.5%, respectively) than F (6.7% and 82.9%, respectively). The combination resulted in longer progression-free survival (by 28 months) and treatment-free survival (by 12 months), but at the expense of higher incidences of grades 3 and 4 toxicity, particularly myelotoxicity. Overall survival with the two regimens did not differ significantly, but the follow-up duration was too short (median 22 months) for this end point to be reached.

Appendix 5: Structured critical appraisal of Eichhorst 2005b [6]

CRITICAL APPRAISAL

Name of Trial: Comparison of the Efficacy and Toxicity of Fludarabine in First Line Therapy of Younger Versus Elderly Patients with Advanced Chronic Lymphocytic Leukaemia (CLL): Results of a Meta-Analysis of Two Phase III Trials of the German CLL Study Group (GCLLSG). Session Type: Oral Session

Reference: Eichhorst B F, Busch R, Wendtner C M et al. ASH Annual Meeting Abstracts 2005; 106(11):717

Question: What is the efficacy and toxicity of fludarabine, when administered to elderly patients first line in advanced CLL compared to younger patients?

Summary: The meta-analysis studied fludarabine therapy in younger versus elderly patients and concluded response rates were similar in both groups, but progression-free survival and overall survival were significantly shorter in the elderly population. Due to the limitations mentioned, there is very little information that can be gleaned from this meta-analysis.

Did the study ask a clearly focussed question?

Yes - This study was designed to assess the efficacy and toxicity of fludarabine when administered to younger and elderly patients, within two phase III trials of the German CLL Study Group (GCLLSG).

Was the study design appropriate?

Can't tell - This study was published as an oral session, therefore did not provide details on study design. Patients (n=362, median age 59 years) were randomised to receive either fludarabine (n=182) or fludarabine plus cyclophosphamide (n=180) within the CLL4 trial. In the CLL5 protocol, 191 elderly patients (median age 71 years) received fludarabine (n=92) or chlorambucil (n=99). All patients were previously untreated and in advanced stage. For both trials, fludarabine dosage was 30mg/m^2/day for 5 days, every 28 days up to 6 cycles.

Were participants appropriately allocated to intervention and control groups?

Can't tell - Inclusion criteria were stated as being identical in both trials except for age limits, but no information was provided in terms of how patients were allocated, what the inclusion and exclusion criteria were or randomisation information.

Were participants, staff and study personnel 'blind' to participants study group?

No - No information is provided, but based on the different regimens for fludarabine, cyclophosphamide and chlorambucil, it may be considered unlikely that the patients were blind to the treatment they were receiving.

Were all of the participants who entered into the trial accounted for at its conclusion?

Can't tell - No information is provided with regards to any drop outs or patients that were not followed up, for any reason.

Were the participants in all groups followed up and data collected in the same way?

Can't tell - No information is provided.

Was the study large enough?

Can't tell - No information is given as to statistical analysis and numbers of patients needed to be included.

How are the results presented and what is the main result?

A mean number of 5.2 fludarabine courses was administered in the CLL4 trial and 4.9 courses in the CLL5 trial. The mean administered cumulative dose of fludarabine per patient was lower in the elderly patients (1076mg vs 1194mg, p=0.05). Overall response rates (ORRs) were similar in both arms, with 82.9% in the younger group and 85.7% in the elderly. The complete remission rate was 6.7% in the younger patients and 10.4% in the elderly (p=0.3). After 24 months follow up, the progression-free survival was significantly shorter in the elderly group (18.7 months) compared to 19.8 months in the younger group after 22 months observation time (p=0.03). Overall survival was lower in elderly patients as well (29 months versus median not reached, p<0.001).

How safe were the regimens?

Progressive disease was the main cause of death in both age groups. In each group, 3 treatment related deaths occurred due to infection or haemolysis. Side effect incidence was similar in both groups. Severe CTC grade 3 and 4, myelosuppression occurred in 39% of younger and 41% of elderly patients. No difference in the rate of leukocytopenia, thrombocytopenia or anaemia was observed. The incidence rate and severity of infections was similar in both groups (24% vs 32% all and 8.7% vs 6.9% CTC grade 3 and 4). The incidence of second neoplasia was significantly higher in elderly patients (2.2% vs 12.2%, p=0.001). The prevalence rate of neoplasia in the US population peaks at 11% in patients aged 70-79 years.

How precise are the results?

As no information was provided regarding statistical analysis, it is difficult to state how precise the results were, or even what the primary endpoint was. Overall response rates were similar. Progression-free survival was significantly shorter in the elderly group, but it is not known whether the trials were powered to show this.

Can the results be applied to the local population?

Detailed information was not given for the different populations so it is unknown how applicable the results are to the local population.

Appendix 6: Structured critical appraisal of Finn 2004 [7]

CRITICAL APPRAISAL

Name of Trial: Fludarabine and Cyclophosphamide produces a higher complete response rate and more durable remissions than fludarabine in patients with previously untreated CLL: Intergroup trial E2997

Reference: Flinn IW, Kumm E, Grever MR *et al*. ASH Annual Meetings Abstracts 2004; 104(11):475.

Question: What is the efficacy and toxicity of fludarabine and cyclophosphamide (FC) versus fludarabine (F) monotherapy in patients with previously untreated CLL?

Summary: Based on this small trial, FC treatment increases complete and overall response rates and progression-free survival times in untreated CLL when compared with F treatment alone with no significant increased risk of toxicity. However no comparison to current first-line treatment is made and patient groups included seem biased and too small for this trial to add to the decision problem evidence.

Did the study ask a clearly focussed question?

Yes – This was an interim analysis of a comparison between combination fludarabine and cyclophosphamide treatment (FC) and fludarabine monotherapy (F) in previously untreated patients with CLL. The intergroup trial E2997 compared FC regimens (cyclophosphamide 600mg/m^2 iv day 1 and fludarabine 20mg/m^2 iv days 1 through 5, followed by filgrastim 5mg/kg sc starting approx. day 8) with F regimens (fludarabine 25mg/m^2 iv days 1 through 5). The primary endpoint was differences in complete response (CR) rates which was determined at the planned review at 76% information.

Was the study design appropriate?

Can't tell/No – The intergroup trial E2997 was a phase III randomized controlled trial comparing the efficacy and toxicity of combination fludarabine and cyclophosphamide treatment with fludarabine monotherapy. In terms of toxicity the study design was appropriate and has shown similar levels of toxicity for both the combination and monotherapy. This is noteworthy as the combination of fludarabine with another agent may be expected to increase the risk of toxicity. In terms of efficacy the abstract makes no reference to any comparisons made with current first-line treatment chlorambucil. The use of this study as evidence in favour of the decision problem is therefore limited as no direct comparison with current therapy has been made.

Were participants appropriately allocated to intervention and control groups?

Can't tell – Although this study was a randomized trial no information is given regarding how patients were allocated to treatment groups. Furthermore there is no information given about any stratification processes used to take into account factors such as age of patient and staging of disease. This is interesting as it is highlighted that 70% of patients were male and only 30% female (this is however, noted as expected in CLL). There was also a higher proportion of patients with early disease stage. 56% of patients were Rai stages 0, 1 or 2 while 44% were in stage 3 or 4.

Were participants, staff and study personnel 'blind' to participants study group?

Can't tell – No reference to blinding of patients, staff or investigators is apparent in the abstract.

Were all of the participants who entered into the trial accounted for at its conclusion?

No – The intergroup trial E2997 enrolled 278 patients. Four patients declined to receive protocol treatment, including one who was later found to be ineligible. Five additional patients were also deemed ineligible. All patients with data were included in the analysis (intent to treat). However response data is only available on 246 of the 278 patients and toxicity data on 252 of the 278 patients.

Were the participants in all groups followed up and data collected in the same way?

Can't tell – The abstract suggests that all patients in the intergroup trial E2997 were not followed up. 278 patients were enrolled in the trial and only 9 patients who were enrolled are cited as being ineligible. However response data is only available on 246 of the 278 patients and toxicity data on 252 of the 278 patients. There is no reference to how the data was collected in the abstract or whether this process was applied uniformly. There is also no discussion of whether higher staged patients received more care from health workers.

Was the study large enough?

Can't tell – There is no reference to a power calculation. However the Data Monitoring Committee did plan a review at 76% information to determine if the null hypothesis of no difference in CR rates could be rejected. This would most likely have to be based on some form of power calculation.

How are the results presented and what is the main result?

Patient data was analysed as intent to treat. Following a planned review at 76% information the ECOG data monitoring committee gave permission for the submission of abstracts as a difference in CR rates was detected. Complete response rates were 22.4% (28/125 patients) and 5.8% (7/121 patients) in the FC and F treatment arms, respectively. Partial response rates were 48.0% (60/125 patients) in the FC treatment regimen and 43.8% (53/121 patients) in the F monotherapy regimen. Fisher exact tests for the difference in CR rates gave a p-value of 0.0002, while the test for difference in OR rate was 0.001. Preliminary estimates of the median progression-free survival time are 41.0 months for the FC arm and 17.7 months for the F alone arm ($p < 0.001$).

How safe were the regimens?

There was no significant difference in the incidence of non-hematologic toxicity and infections. There were two deaths due to infection with grade 3 or 4 neutropenia (one in each treatment arm). Grade 4 or higher non-hematologic toxicities were seen in 17% of CF patients and 13% in F regimen patients ($p = 0.48$). In terms of general infections 17% was seen in the FC regimen versus 11% in the F regimen ($p = 0.21$). There is no reference to any other forms of toxicity or adverse reactions.

How precise are the results?

Can't tell – Confidence limits are unavailable for all p values reported and patient numbers seem too small to make a decision ($n = 278$).

Can the results be applied to the local population?

The trial suggests that combination fludarabine and cyclophosphamide treatment compared with fludarabine alone provides a benefit in terms of increased complete and partial response rates and progression-free survival which outweighs any risks of increased toxicity. However patients included in the trial were predominately male with early stage disease. This could be different to local populations in ways that would produce different results. The outcomes of this trial are predominately toxicity and response rates and although preliminary estimates of progression-free survival rates are given there is no reference to health-related quality of life (HRQoL) which is one of the main outcomes that should be considered in CLL. The results of this trial will be of interest to policy makers and professionals as they indicate that combination fludarabine and cyclophosphamide treatment offers a clinical benefit in terms of survival times. If FC combination treatment is also cost effective the results of the intergroup trial E2997 will be of importance in the decision problem.

Question: What is the efficacy and toxicity of fludarabine and cyclophosphamide (FC) versus fludarabine (F) monotherapy in patients with previously untreated CLL?

Answer – There is no significant difference between FC treatment regimens and F treatment regimens in terms of non-hematologic toxicity and infections. FC treatment does however appear to significantly increase both complete and partial response rates in patients with previously untreated CLL in addition to increasing progression-free survival times.

Appendix 7: Structured critical appraisal of Karlsson 2004 [8]

CRITICAL APPRAISAL

Name of Trial: Cladribine or Fludarabine or High dose intermittent chlorambucil as first line treatment for symptomatic CLL: A first interim analysis of data from the international randomised phase III trial

Reference: Karlson K, Stromberg M, Jonsson V et al. ASH Annual meeting abstracts2004;104 (11): 3470

Question: Is Cladribine, High dose chlorambucil or Fludarabine best as first line treatment for CLL?

Summary: This was an interim analysis of a comparison between Cladribine (another purine analogue) Fludarabine and High dose chlorambucil. The primary endpoint is not clearly defined but it states that responses were assessed according to NCI criteria. The data thus far suggest similar levels of haematological toxicity and serious infection risk for fludarabine monotherapy compared with high dose chlorambucil. Cladribine was associated with a higher response rate compared with chlorambucil (p=0.06)

Did the study ask a clearly focussed question?

Yes - Patients aged 18-75 years from Scandinavia, Australia and UK. Randomisation was stratified according to age and stage.
At this time of reporting 150 patients had been recruited and 139 patients were evaluable, 6 had incomplete data and another 5 did not meet eligibility criteria. Doses used were Chlorambucil 10mg / m sq orally day 1-10. Fludarabine 25mg / m sq iv day (this is the recommended UK dose for monotherapy) 1-5 or Cladribine 5mg / m sq iv or sc day 1-5. These schedules were given monthly for up to six courses. The results were analysed as ITT. Responses are documented as NCI criteria but this acronym is neither defined nor explained. These criteria have however been used in the reporting of other similar studies.

Was the study design appropriate?

Yes - This was a randomised parallel group study with stratification to take account of known prognostic factors e.g. age of patient and staging of disease. It is not stated if it was blinded but this is unlikely given the nature of the study.

Were participants appropriately allocated to intervention and control groups?

Yes - Randomisation was stratified according to age and stage.

Were participants, staff and study personnel 'blind' to participants study group?

Cant tell – This appears to be an open label study from the study design. It would be almost impossible to run such a study double blind.

Were all of the participants who entered into the trial accounted for at its conclusion?

Yes - At this time of reporting 150 patients have been recruited and 139 patients were evaluable, 6 had incomplete data and another 5 did not meet eligibility criteria.

Were the participants in all groups followed up and data collected in the same way?

Can't tell

Was the study large enough?

Cant tell - No power calculation was reported within the abstract.

How are the results presented and what is the main result?

The results were analysed as ITT. Responses are documented as NCI criteria but this acronym is neither defined nor explained. These criteria have been used in the reporting of other similar studies. These responses are quoted as 57%, 67% and 74% for the chlorambucil / fludarabine / and cladribine groups respectively. A p-value of 0.06 is quoted for a Chi squared test between the chlorambucil and cladribine groups. There are results quoted for CR / nPR and PR (complete response, partial response and nodular partial response) and these were 1 / 3 / 22 for chlorambucil 0 / 2 / 28 for fludarabine and 2 / 3 / 30 for CdA. Grade III and IV haematological toxicity was seen in 25% and 20% for chlorambucil 24% and 5% for Fludarabine and 36% and 11% for Cladribine. Serious grade III-V infections were seen in 25% 28% and 30% in Chl, F and Cld groups.

How precise are the results?

In the absence of a power calculation and a well defined primary endpoint it is unclear whether the results are likely to simply be due to chance. As this is only an interim analysis it is likely that further data will be available in the future for patients who were subsequently recruited. The total number of patients planned to be recruited in the future is not stated.

Can the results be applied to the local population?

Doses of Fludarabine and chlorambucil were those licensed for use within the UK, but cladribine is unlicensed for this use. Patients groups appeared to be consistent with current treatment guidelines. This study does not add to the data comparing the combination of fludarabine and cyclophosphamide with chlorambucil or indeed other modalities of treatment. It does however compare a novel agent cladribine with both fludarabine and chlorambucil. As this is only an interim analysis the results should be interpreted with caution and a full report should be sought. The data thus far suggest similar levels of haematological toxicity and serious infection risk for fludarabine monotherapy compared with high dose chlorambucil and raises the possibility that cladribine may be an alternative agent with similar if not better efficacy, but possibly higher toxicity than fludarabine

Appendix 8: Structured critical appraisal of Rai 2000 [4]

CRITICAL APPRAISAL

Name of Trial: Fludarabine compared with chlorambucil as primary therapy for chronic lymphocytic leukaemia.

Reference: Rai KT, Peterson BL, Appelbaum FR et al. N Engl J Med 2000;343:1750-7.

Question: To compare the efficacy of fludarabine with that of chlorambucil, and a combination regimen of fludarabine plus chlorambucil in the primary treatment of chronic lymphocytic leukaemia.

Summary: This study showed that when used as initial treatment for CLL, F is associated with a significantly longer PFS and higher response rates than those treated with Chl alone. However, there was no significant advantage to combination treatment over F alone in terms of response and there was no statistically significant difference in OS among the three treatment groups. The benefits in PFS and response appear to be offset by the significantly higher overall incidence of grade three or four side effects, especially neutropenia and infections in the F group compared to the Chl group. The dose of Chl used in this study is lower than the UK standard.

Did the study ask a clearly focussed question?

Yes - The study was designed to compare the efficacy of fludarabine (F) with that of chlorambucil (Chl) and a combination regimen of fludarabine plus chlorambucil (FChl) in patients with previously untreated chronic lymphocytic leukaemia (CLL). 544 patients were randomly assigned to; fludarabine (25mg/m^2 BSA, administered IV daily for five days every 28 days as per SPC), chlorambucil (40mg/m^2, given orally every 28 days), or fludarabine (20mg/m^2 daily for five days every 28 days) plus chlorambucil (20mg/m^2, every 28 days). Patients also received oral allopurinol (300mg per day for nine days) from the day prior to treatment through to day eight of each 28-day treatment cycle for the first three cycles, and thereafter according to the judgement of their physicians. Patients with an additional response at each monthly evaluation continued to receive the assigned treatment for a maximum of 12 cycles. Switching from F to Chl or Chl to F was allowed if there was no partial response, if disease progression occurred, or if the patient relapsed within 6 months of stopping the initially allocated drug. Median duration of follow-up was 62 months. The primary endpoint was progression-free survival (PFS). Secondary endpoints were response according to stage, overall survival (OS), and safety.

Was the study design appropriate?

Yes - The study was a randomised, prospective multicentre controlled trial. Eligible patients had previously untreated CLL. The stage of disease was assessed according to National Cancer Institute (NCI) guidelines and the modified Rai staging system. All patients with high-risk (Rai stage III or IV) were eligible, as were Intermediate-risk Rai stage I or II patients if they had at least one of the following: weight loss, night sweats, extreme fatigue, lymphadenopathy, splenomegaly, hepatomegaly, >50% increase in lymphocytes over 2 months. Additional criteria were >18 years of age; ECOG performance status of 0, 1or 2; baseline values for liver or kidney function no greater than 1.5x the upper limit of normal and a negative direct antiglobulin (Coombs) test. Exclusion criteria were any previous treatment.

Were participants appropriately allocated to intervention and control groups?

Yes - A total of 544 patients were randomly assigned to F (195), Chl (200) or FChl (149). Thirty two patients (15, 7 and 10 respectively, in the three groups) were considered ineligible, and a further three patients (one in the F group and two in the FChl group) withdrew before receiving treatment. Patient assignment was balanced with respect to clinical features and risk categories. Randomisation was performed centrally. However, the method of randomisation was not reported. The analyses for PFS are based on an intent-to-treat principle. Baseline characteristics were well balanced between treatment groups.

Were participants, staff and study personnel 'blind' to participants study group?

No – There was no mention of blinding although allocation is likely to have been concealed. However, observer bias would be unlikely in this study as the endpoints were clearly measurable and centralised review was required for all specimens from all patients who had a complete remission.

Were all of the participants who entered into the trial accounted for at its conclusion?

Yes - Follow-up was adequate with survival data available for 507 of the 509 patients (>99%); 474 could be evaluated for a therapeutic response (93%); 477 could be evaluated for drug toxicity (94%); and 172 patients in the F group and 183 in the Chl group could be evaluated for progression-free survival.

Were the participants in all groups followed up and data collected in the same way?

No – However, all patients evaluated for a therapeutic response were followed up in the same way. All patients were evaluated monthly, before the next scheduled cycle of treatment, to assess the toxic affects of the treatment and clinical response. The exact parameters evaluated at each evaluation are not stated.
Patients who withdrew after starting therapy, who were withdrawn due to drug toxicity or a complicating disease, or who crossed over to the other treatment for reasons other than those defined in the study protocol were followed up for PFS. After 450 patients had been enrolled in the trial the FChl group was closed because of the high rate of life-threatening toxic side effects with the combined treatment. Patients in the Chl group were followed only to assess survival and the occurrence of a second cancer

Was the study large enough?

Can't tell – Initially the sample trial size was 450 patients this is stated as providing adequate statistical power to detect a difference in the rates of remission between the Chl group and either of the two groups assigned to receive F. However, a planned interim analysis showed that the response rate for the Chl group was significantly lower than the other two groups and therefore the protocol was modified to make PFS the primary endpoint; the target sample remained the same. After the interim analysis the overall median PFS was found to be longer than anticipated therefore an additional 94 patients were enrolled to increase statistical power making a revised target sample of 544 patients.

How are the results presented and what is the main result?

Progression-free survival Assignment to the F plus Chl group was stopped after a planned interim analysis revealed excessive toxicity and a non-superior response rate to F alone. Among the other two groups there was a significantly longer median time to disease progression among the patients treated with F compared to those receiving Chl (20 months vs. 14 months, respectively, $P<0.001$).

Clinical response Response rates for evaluated patients were significantly higher in those patients treated with F than those treated with Chl alone: Overall response (OR) F 63%, FChl 61% and Chl 37%; Complete response (CR) F 20%, FChl 20% and Chl 4%; Partial response (PR) F 43%, FChl 41% and Chl 33% ($p<0.001$ for all comparisons). There was no significant advantage to combination treatment over F alone in terms of response.

Overall survival There was no statistically significant difference in OS among the three treatment groups (F 66 months, Chl 56 months and FChl 55 months, $p=0.21$), or between the F and Chl groups ($p=0.10$).

How safe were the regimens?

All side effects were graded on a six point scale with zero defined as none, one as mild, two moderate, three severe, four life-threatening and five as lethal. Most recorded side effects were of grade one or two. Only one treatment related death was recorded. The incidence of grade three or four neutropenia and infections were higher in the F group compared to the Chl group (27% vs. 19%,

p=0.007, and 16% vs. 9%, p=0.01, respectively). Overall the incidence of all grade three or four side effects was significantly greater with F than Chl (44% vs. 55%, p<0.001).

How precise are the results?

This was a relatively small but generally well conducted trial. Follow-up was adequate and there was no imbalance among the three treatment groups with respect to baseline characteristics, clinical features and risk categories. A chi-square test was used to compare response rates between study groups. All time-to-event distributions were compared by the Kaplan-Meier method and log-rank test. In terms of the primary endpoint PFS the result was highly significant at p<0.001. The rates of complete and partial response were also highly significant at p<0.001. However, this study was not sufficiently powered to detect a difference in OS and some outcomes are missing from patients initially randomised to receive FChl. Furthermore, the median number of F cycles administered or the median number of cycles needed to induce a CR are not stated. Non-responding patients or those relapsing <6 months were allowed to crossover therefore there is the possibility of a carry-over effect. No information is available regarding the ratio of benefit to harm in non-responders. The study was open to detection bias through lack of blinding. However, specimens from patients who had complete remission were subject to centralised review.

Can the results be applied to the local population?

The study was carried out in the US with disease status assessed according to NCI guidelines and the modified Rai staging system. Therefore there are unlikely to be any significant differences from the UK population with CLL and age and gender balance seem appropriate. However, the dose of Chl used in this study is lower than the UK standard.

Appendix 9: Structured critical appraisal of Spriano 2000 [9]

CRITICAL APPRAISAL

Name of Trial: Multicentre prospective randomized trial of fludarabine versus chlorambucil and prednisone in previously untreated patients with active B-CLL: final report.

Reference: Spriano M, Chiurazzi F, Liso V *et al*. Haematology and Cell Therapy 2000; 42(1:93).

Question: What is the response rate and safety of fludarabine i.v. infusion versus standard therapy with chlorambucil in patients with previously untreated active B-CLL?

Summary: Data from this small study indicates that fludarabine treatment is as effective as standard therapy with chlorambucil and prednisone without any signs of increased toxicity. Response duration was significantly prolonged following fludarabine treatment however further investigation is needed to confirm this result.

Did the study ask a clearly focussed question?

Yes – This abstract describes a randomized prospective multi-centre study which aimed to compare the response rate and safety of fludarabine versus standard therapy with chlorambucil and prednisone. Previously untreated patient with active B-CLL were given either fludarabine 25 mg/m^2 by 30 minute i.v. infusion daily for 5 consecutive days every 4 weeks, or chlorambucil 30 mg/m^2 orally on days 1 and 15 plus prednisone 40 mg/m^2 i.m. on days 1-5 and 15-19 every 4 weeks.

Was the study design appropriate?

Yes – This study is a RCT, which would be an appropriate design when comparing response rates and toxicity. However no information is given in the abstract as to why chlorambucil and prednisone treatment is used as the standard therapy and not chlorambucil alone.

Were participants appropriately allocated to intervention and control groups?

Can't tell – Although this study was a randomized trial no information is given regarding how patients were allocated to treatment groups. Furthermore there is no information given about any stratification processes used to take into account factors such as age of patient and staging of disease.

Were participants, staff and study personnel 'blind' to participants study group?

Can't tell – No reference to blinding of patients, staff or investigators is apparent in the abstract.

Were all of the participants who entered into the trial accounted for at its conclusion?

Yes – One hundred and fifty patients were enrolled in this trial. Responses were not evaluable in 8 patients; however no reasoning as to why these patients were excluded is given. 69 patients were included in the fludarabine arm and 73 in the chlorambucil arm.

Were the participants in all groups followed up and data collected in the same way?

Can't tell – There is no reference to how the data was collected in the abstract or whether this process was applied uniformly. There is also no discussion of whether higher staged patients received more care from health workers.

Was the study large enough?

Can't tell – There is no reference to a power calculation in the abstract.

How are the results presented and what is the main result?

Patients with progressive (PD) or stable disease (SD) after 3 and 6 courses of chemotherapy, respectively, stopped treatment and were evaluated for survival. The response rate (CR + PR) was 71% (46 CR + 25 PR) in the fludarabine arm and 71% (37 CR + 34 PR) in the chlorambucil and prednisone arm. Refractory CLL was seen in 19% (10 SD + 9 PD) and 18% (11 SD + 7 PD) of patients respectively. Response duration was longer in the fludarabine arm (28 months versus 21 months; p = 0.007).

How safe were the regimens?

Toxicity was comparable in the two treatment groups; however no data is given and there is no reference to any specific forms of toxicity or adverse reactions.

How precise are the results?

Can't tell – Confidence limits are unavailable for all p values reported and patient numbers seem too small to make a decision (n = 142).

Can the results be applied to the local population?

Details relating to the population studied in this trial are limited and as such it is difficult to say whether they reflect the local population accurately. There is no reference to the male:female ratio or age characteristics. In addition it is not clear that groups were balanced once outcomes were analysed and until this data is available it is not certain the results can be applied to the general population of patients with CLL.

Question: What is the response rate and safety of fludarabine versus standard therapy with chlorambucil and prednisone in patients with previously untreated active B-CLL?

Answer – In terms of response rates, fludarabine is as effective as standard treatment with chlorambucil and prednisone and shows no signs of increased toxicity. Response duration was also longer following fludarabine treatment however further investigations are needed to confirm these results.

Appendix 10: Review of the Guidelines on the diagnosis and management of chronic lymphocytic leukaemia.

BCSH GUIDLINE SUMMARY

Guidelines issued on behalf of the British Committee for Standards in Haematology (BCSH) [15] aim to provide a rational approach to the diagnosis and management of patients with chronic lymphocytic leukaemia (CLL). Treatment recommendations were influenced by current and proposed clinical trials in the UK and guidance from the National Institute of Clinical Excellence (NICE).

Diagnosis

A definitive diagnosis of CLL is based on the combination of lymphocytosis (5 × 10^9/l.), lymphocyte morphology (small or medium sized lymphocytes with clumped chromatin, indistinct or absent nucleoli and scanty cytoplasm) and immunophenotyping (weak or absent expression of immunoglobulin, CD5, CD19, CD23, CD79B, CD22 and FMC7). Additional investigations such as direct antiglobulin tests, reticulocyte counts, renal and liver biochemistry, serum immunoglobulins, chest x-ray, bone marrow aspirate and trephine biopsy may also be helpful in diagnosis. Although not normally indicated in the diagnosis of typical CLL, lymph node biopsies, cytogenetic/fluorescence *in situ* hybridization analysis and computed tomography or ultrasound scanning can help if diagnosis is uncertain. Established factors such as clinical staging and serum markers and more recent tests such as CD38 expression, IgVH gene status, cytogenetic abnormalities and drug sensitivity testing appear to provide additional prognostic information.

Management

First-line

For the majority of patients who are ineligible for a transplant procedure and in whom there is no contraindication to fludarabine, entry into the MRC CLL4 study should be offered. This trial randomizes patients to chlorambucil, fludarabine or fludarabine and cyclophosphamide. Both fludarabine and chlorambucil are options for patients who do not wish to enter the study. Where a patient is considered suitable for entry into the MRC CLL5 trial or for allogenic transplantation, then an initial treatment, such as fludarabine or fludarabine and cyclophosphamide, which is likely to result in a

complete or very good partial remission, should be chosen (grade C recommendation, level IV evidence).

Patients in whom fludarabine is contraindicated and for whom a palliative approach has been adopted should be treated with chlorambucil. There is no survival advantage for including an anthracycline with chlorambucil in the initial treatment of advanced CLL (grade A recommendation, level 1A evidence). Further studies using standard response criteria are required before high-dose chlorambucil can be recommended as an initial treatment for CLL (grade C recommendation, level IV evidence). Alemtuzumab and rituximab monotherapy are not recommended for untreated CLL (grade B recommendation, level IIb evidence and grade C recommendation, level III evidence respectively). Rituximab combined with fludarabine (with or without cyclophosphamide) requires further evaluation in untreated CLL (grade B recommendation, level Ib evidence).

Second-line

Patients who relapse after an initial response to low dose chlorambucil may be treated with a further course of chlorambucil (grade B recommendation, level IIb evidence). Patients refractory to low dose chlorambucil should be treated with fludarabine. CHOP is an alternative treatment for patients unsuitable for fludarabine (grade B recommendation, level IIb evidence).

Patients who develop progressive disease more than 1 year after receiving fludarabine and whose CLL responded to fludarabine initially may be treated again with fludarabine alone (grade B recommendation, level IIb evidence). Patients who develop progressive disease within 1 year of previous fludarabine therapy may be treated with a combination of fludarabine and cyclophosphamide (grade B recommendation, level IIb evidence). Patients who are refractory or become resistant to fludarabine currently have a poor prognosis. In these patients high dose methylprednisolone is recommended in cases with bulky lymphadenopathy, alemtuzumab is recommended in cases without bulky lymphadenopathy. Rituximab combined with fludarabine (with or without cyclophosphamide) may be effective in refractory CLL and warrants further evaluation in this setting.

Autologous transplantation should be considered for patients in complete or good partial remission who are able to withstand high-dose chemotherapy and total body irradiation (TBI). Allogeneic transplant procedures should be considered for younger patients with good performance status who have been previously treated and have poor risk disease.

Complications

Increased susceptibility to infection is both intrinsic to the disease and therapy-related. Risk factors for infection include advanced age, number of previous treatments and ongoing treatment. Cycling antibiotics as infective prophylaxis is recommended in patients with recurrent chest infections or urinary tract infections. Prophylaxis against *Pneumocystis carinii* and herpes zoster/simplex should be considered for patients receiving purine analogues or alemtuzumab. Patients treated with high-dose methylprednisolone should receive prophylaxis against candidiasis with fluconazole. Patients with hypogammaglobulinaemia and recurrent bacterial infections, especially those in whom prophylactic antibiotics have proved ineffective, should be treated with prophylactic intravenous immunoglobulin (IVIG) (grade A recommendation, level Ib evidence). It is standard practice to recommend annual influenza vaccination for patients with CLL.

Hierarchy of strength of evidence used in grading recommendations in NICE clinical guidelines (strongest to weakest).

Ia evidence from systematic reviews or meta-analyses of randomised controlled trials.

Ib evidence from at least one randomised controlled trial.

IIa evidence from at least one controlled study without randomisation.

IIb evidence from at least one other type of quasi-experimental study.

III evidence from non-experimental descriptive studies, such as case-controlled studies.

IV evidence from expert committee reports or opinions or clinical experience of respected authorities.

Grading of recommendations.

A Based on hierarchy I evidence.

B Based on hierarchy II evidence or extrapolated from hierarchy I evidence.

C Based on hierarchy III evidence or extrapolated from hierarchy I or II evidence.

D Directly based on hierarchy IV evidence or extrapolated from hierarchy I, II or III evidence.

Adapted from [46]

Appendix 11: Search strategies used to identify published economic evaluations

Search in Medline (Silverplatter)

1 Leukemia, B-Cell/ (856)
2 (cll or b-cll).ti,ab. (6609)
3 (chronic and lymphocytic and leukemia).ti,ab. (8155)
4 (chronic and lymphocytic and leukaemia).ti,ab. (2237)
5 or/1-3 (11217)
6 naive.ti,ab. (24229)
7 limit 6 to humans [Limit not valid in: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations; records were retained] (12632)
8 untreated.ti,ab. (92408)
9 limit 8 to humans [Limit not valid in: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations; records were retained] (53420)
10 (first and line).ti,ab. (43197)
11 limit 10 to humans [Limit not valid in: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations; records were retained] (33881)
12 7 or 9 or 11 (98061)
13 5 and 12 (838)
14 "health care economics and organizations"/ (0)
15 economics/ (24921)
16 "costs and cost analysis"/ (35095)
17 Health Care Costs/ (14165)
18 Quality of Life/ (55584)
19 (costs or cost or costed or costly or costing).tw. (171459)
20 (economic$ or pharmacoeconomic$ or price$ or pricing).tw. (87141)
21 or/14-20 (328027)
22 13 and 21 (15)
23 from 22 keep 1-15 (15)

Search in Embase (Ovid)

1 cll.ti,ab. (5347)
2 "Cost and quality of life studies in first-line CLL".ti,ab. (0)
3 b-cll.ti,ab. (1808)
4 chronic lymphatic leukemia/ (8688)
5 b-cell leukemia/ (2763)
6 chronic lymphocytic leukaemia.ti,ab. (1622)
7 chronic lymphocytic leukemia.ti,ab. (5827)
8 or/1-7 (11850)
9 naive.ti,ab. (20604)
10 untreated.ti,ab. (73597)
11 first-line.ti,ab. (16286)
12 or/9-11 (109076)
13 8 and 12 (660)
14 Economic Aspect/ or Economic Evaluation/ or "Cost Benefit Analysis"/ (88248)
15 Health Economics/ (8543)
16 "COST EFFECTIVENESS ANALYSIS"/ or "COST"/ or "COST UTILITY ANALYSIS"/ (62707)
17 (cost or costs or costed or costly or costing).ti,ab. (135178)
18 Economics/ (4920)

19 Quality-of-life/ (65603)
20 Treatment-outcome/ (254995)
21 outcome-assessment/ (172924)
22 or/14-21 (518668)
23 13 and 22 (76)
24 from 23 keep 1-76 (76)

Search in NHS EED (NHS Economic Evaluation Database)

1. s leukemia b cell/kwo (1)
2. s cll (2)
3. s chronic(w)lymphatic(w)leukemia (3)
4. s chronic(w)lymphatic(w)leukaemia (3)
5. s s1 or s2 or s3 or s4 (6)
6. s naïve (28)
7. s untreated (205)
8. s first(w)line (225)
9. s s6 or s7 or s8 (441)
10. s s5 and s9 (1)