NHS England Evidence Review:

Sorafenib maintenance for FLT3-internal tandem duplication acute myeloid leukaemia undergoing allogeneic haematopoietic stem cell transplantation

NHS England URN: 2262

NHS England Evidence Review

Sorafenib maintenance for FLT3-internal tandem duplication acute myeloid leukaemia undergoing allogeneic haematopoietic stem cell transplantation

Drafted: January 2023

Prepared by Solutions for Public Health (SPH) on behalf of NHS England Specialised Commissioning

Contents

1. Introduction 3
2. Executive summary of the review 4
3. Methodology 8
4. Summary of included studies 9
5. Results 11
6. Discussion 19
7. Conclusion 21
Appendix A PICO Document 22
Appendix B Search strategy 25
Appendix C Evidence selection 26
Appendix D Excluded studies table 27
Appendix E Evidence Table 29
Appendix F Quality appraisal checklists 41
Appendix G GRADE profiles 42
Glossary 52
References 54

1. Introduction

This evidence review examines the clinical effectiveness, safety and cost effectiveness of sorafenib as maintenance therapy compared with standard of care for the treatment of patients with fms-like tyrosine kinase 3-internal tandem duplication (FLT3-ITD) acute myeloid leukaemia (AML) who have undergone allogeneic haematopoietic stem cell transplantation (allo-HSCT).

FLT3-ITD AML is an aggressive haematological malignancy. Allogeneic HSCT improves survival in these patients, however leukaemia relapse remains high. Maintenance therapy is the ongoing treatment of FLT3-ITD AML after the patient has received allo-HSCT.

Sorafenib is a tyrosine kinase inhibitor. It is taken orally at a starting dose of 200mg, twice daily, to a maximum of 400mg twice daily. Sorafenib is usually started 60-100 days after allo-HSCT for a maximum of 24 months or until toxicity occurs. Sorafenib is not licenced for use in FLT3-ITD AML. Patients may or may not have received a FLT3 tyrosine kinase inhibitor before HSCT.

There is no standard of care maintenance therapy for FLT3-ITD AML patients to prevent disease relapse after allo-HSCT. The comparators of interest are therefore no treatment and placebo.

In addition, the review scope included the identification of possible subgroups of patients within the included studies who might benefit from sorafenib more than others and the dose regimens of sorafenib used.

2. Executive summary of the review

This evidence review examines the clinical effectiveness, safety and cost effectiveness of sorafenib as maintenance therapy compared with standard of care for the treatment of patients with fms-like tyrosine kinase 3 (FLT3)-internal tandem duplication (FLT3-ITD) acute myeloid leukaemia (AML) who have undergone allogeneic haematopoietic stem cell transplantation (allo-HSCT). The searches for evidence published since November 2012 were conducted on 14th November 2022 and identified 202 potential references. These were screened using their titles and abstracts and 23 full text papers potentially relating to the use of sorafenib for FLT3-ITD AML after allo-HSCT were obtained and assessed for relevance.

Two randomised controlled trials (RCT) (published in three papers) were identified for inclusion. One RCT (SORMAIN, 15 centres in Germany and Austria) compared sorafenib (n=43) to placebo (n=40) in adults with FLT3-ITD AML in complete haematologic remission after allo-HSCT with median follow-up of 41.8 months (Burchert et al 2020). A second RCT (seven centres in China) compared sorafenib (n=100) to no maintenance therapy (n=102) in adults with FLT3-ITD AML in composite complete remission[1] before and after allo-HSCT. RCT results at a median follow-up of 21.3 months were published in Xuan et al (2020). Results at a median follow-up of 36.8 months were published in Xu et al (2022).

In terms of clinical effectiveness:

- **Relapse free survival (RFS) (critical outcome).**
  - *For sorafenib vs placebo:* One RCT provided high certainty evidence of statistically significantly higher RFS[2] at 24 months for sorafenib (85.0%) compared to placebo (53.3%) and moderate certainty evidence of statistically significantly fewer relapse events with sorafenib (23.3%) compared to placebo (47.5%) at a median follow-up of 42 months.
  - *For sorafenib vs no maintenance therapy:* One RCT provided high certainty evidence of statistically significantly higher RFS[3] with sorafenib compared to no maintenance therapy at two years (78.9% vs 56.6%) and three years (75.9% vs 52.5%). The same RCT also provided high certainty evidence of statistically significantly lower cumulative incidence of relapse with sorafenib compared to no maintenance therapy at one (7.0% vs 24.5%), two (11.9% vs 31.6%) and three (13.0% vs 34.8%) years.
- **Overall survival (critical outcome).**
  - *For sorafenib vs placebo:* One RCT provided high certainty evidence of statistically significantly higher overall survival[4] at 24 months for sorafenib (90.5%) compared to placebo (66.2%). In this RCT there was moderate certainty evidence that the difference in number of deaths between the groups was not statistically significant at a median follow-up of 55 months (25.6% vs 40.0%).

[1] Composite complete remission was complete remission, complete remission with incomplete platelet recovery or complete remission with incomplete haematological recovery
[2] Calculated as time from randomisation to the first occurrence of either AML relapse or death from any cause. Relapse was defined as loss of complete haematologic remission, according to the revised recommendations of the International Working Group (Cheson et al 2003)
[3] Calculated as time from transplantation until relapse or death from any cause. Relapse was defined as either reappearance of leukaemic blasts in the peripheral blood or at least 5% blasts in the bone marrow aspirate or biopsy specimen not attributable to any other causes, or reappearance or new appearance of extramedullary leukaemia
[4] Calculated as time from randomisation to death from any cause

- *For sorafenib vs no maintenance therapy:* One RCT provided moderate certainty evidence of statistically significantly higher overall survival[5] for sorafenib compared to no maintenance therapy at two (82.1% vs 68.0%) and three years (79.0% vs 61.4%).

- **Treatment adherence (important outcome).**
  - *For sorafenib vs placebo:* One RCT provided moderate certainty evidence that sorafenib patients had a shorter duration of therapy (34.6 weeks vs 54.4 weeks) and a higher proportion of dose reductions (49% vs 40%) than placebo patients at a median follow-up of 42 months. The groups were not statistically compared.
  - *For sorafenib vs no maintenance therapy:* One RCT reported moderate certainty evidence that patients received sorafenib for a median of 134 days with dose reductions and dose interruptions due to adverse events in 42% and 12% respectively.

- **Graft-versus-host disease (GVHD) (important outcome).**
  - *For sorafenib vs placebo:* One RCT provided moderate certainty evidence that a higher proportion of sorafenib patients had acute Grade ≥2 (24% vs 18%), mild/moderate chronic (43% vs 36%) and severe chronic (19% vs 10%) GVHD[6] than placebo patients at a median follow-up of 42 months. The groups were not statistically compared.
  - *For sorafenib vs no maintenance therapy:* One RCT provided moderate certainty evidence that sorafenib and no maintenance therapy patients had similar proportions of acute Grade ≥2 (23% vs 21%) and moderate/severe chronic (18% vs 17%) GVHD[7], assessed up to 210 days post-transplantation. The groups were not statistically compared.

- No evidence was identified for quality of life (critical outcome), hospitalisation (important outcome) or activities of daily living (important outcome).

In terms of safety:

- *For sorafenib vs placebo:* One RCT provided moderate certainty evidence of higher drug discontinuation due to toxicity with sorafenib (21%) than placebo (5%) at a median follow-up of 42 months. The groups were not statistically compared. The same RCT specified adverse events and drug-related adverse events of Grade ≥3[8] experienced by each group at a median follow-up of 42 months. However, the proportion of patients in each group with any adverse event of Grade ≥3 was not reported and the groups were not statistically compared. The most common Grade ≥3 adverse event for both groups was infections (26% vs 23%). However, there were 79% and 56% of Grade ≥3 adverse events recorded as 'other' (not further defined) in the sorafenib and placebo groups respectively. The most common drug-related Grade ≥3 adverse event was electrolyte alterations in the sorafenib group (7%) and gastrointestinal toxicity in the placebo group (8%). For drug-related Grade ≥3 adverse events, 19% and 10% were recorded as 'other' (not further defined) in the sorafenib and placebo groups respectively.

[5] Calculated as time from transplantation to death from any cause
[6] Acute GVHD categorised according to the Mount Sinai Acute GVHD International Consortium (Harris et al 2016). Chronic GVHD categorised according to the National Institutes of Health consensus criteria (Filipovich et al 2005)
[7] Acute and chronic GVHD were graded according to published guidelines (Prezpiorka et al 1995, Jagasia et al 2015)
[8] Grading criteria not specified

- *For sorafenib vs no maintenance therapy:* One RCT provided moderate certainty evidence that sorafenib and no maintenance therapy patients had similar proportions of deaths due to adverse events (4% vs 5%) and patients experiencing at least one Grade 3 or 4[9] adverse event (50% vs 46%), assessed up to 210 days post-transplantation. The groups were not statistically compared. The same RCT reported discontinuation of sorafenib due to adverse events in 5% of patients. The most common Grade 3-4 adverse event for both groups was infections (25% vs 24%). The most common Grade 3-4 treatment-related adverse events with sorafenib were skin-related (7%) or haematological (5%). No patients died from treatment-related adverse events.

In terms of cost effectiveness:

- No evidence was identified for cost effectiveness.

In terms of subgroups:

- Two RCTs conducted exploratory, post-hoc subgroup analysis for relapse free survival:
  - *For sorafenib vs placebo:* One RCT reported that RFS was statistically significantly higher for sorafenib vs placebo for patients with undetectable minimal residual disease (MRD) before allo-HSCT and patients with detectable MRD after allo-HSCT.
  - *For sorafenib vs no maintenance therapy:* One RCT reported that cumulative incidence of relapse at two years was statistically significantly lower for sorafenib vs no maintenance therapy for patients aged ≥35 years, but that there was no statistically significant difference for patients aged <35 years.

Dose regimens of sorafenib used:

- In the RCT by Burchert et al (2020), the starting dose of sorafenib was 2 x 200mg orally per day for two weeks (dose level 1). This was followed by 3 x 200mg orally per day for four weeks (dose level 2), then 4 x 200mg orally per day (dose level 3). Dose reductions were permitted. Treatment started between 60 and 100 days after allo-HSCT and continued for 24 months or until relapse or intolerable toxicity.
- In the RCT by Xuan et al (2020), the sorafenib dose was 2 x 400mg orally per day. Dose reductions or interruptions were allowed if adverse events of Grade ≥3 occurred. Dose reductions were to 200mg once or twice a day with return to 400mg twice a day after the resolution of adverse events. Treatment started between 30 and 60 days after allo-HSCT and continued up to 180 days post-transplantation.

Please see the results table (section 5) in the review for further details of outcomes.

Limitations:

The two RCTs were both well conducted with few risk of bias issues that were likely to impact the outcomes reported. However, statistical comparison between groups was not reported for treatment adherence, GVHD or safety outcomes and some outcomes were

[9] Non-haematological adverse events defined using the National Cancer Institute Common Terminology Criteria for Adverse Events (version 4.0) where Grade 1 = mild; Grade 2 = moderate; Grade 3 = severe or medically significant but not immediately life threatening; Grade 4 = life-threatening consequences; Grade 5 = death related to adverse event. Grade 3 haematological adverse events were defined as either an absolute neutrophil count <1.0 x 10^9 cells/L but ≥ 0.5 x 10^9 cells/L or a platelet count <30 x 10^9/L but ≥20 x 10^9/L. Grade 4 haematological adverse events were defined as either an absolute neutrophil count <0.5 x 10^9 cells/L or a platelet count <20 x 10^9/L

downgraded for imprecision due to wide confidence intervals around a hazard ratio. In the RCT by Xuan et al (2020), the comparator used (no maintenance therapy) limited the availability of comparative data for treatment adherence and some safety outcomes.

Conclusion:
This evidence review includes one RCT comparing sorafenib to placebo and a second RCT comparing sorafenib to no maintenance therapy. The populations of both studies were adults with FLT3-ITD AML after allo-HSCT. There was no evidence on cost effectiveness.

There were RCT data comparing sorafenib to placebo or no maintenance therapy for the critical outcomes of relapse free survival and overall survival and the important outcomes of treatment adherence, GVHD and safety. Both RCTs reported a statistically significant advantage for sorafenib for relapse free survival and overall survival up to at least two years. The groups were not statistically compared for treatment adherence, GVHD or safety outcomes in either RCT. This limits the interpretation of these outcomes, some of which appeared to favour placebo. However, GVHD and safety outcomes were more similar for sorafenib and no maintenance therapy.

No evidence was identified for the critical outcome of quality of life, or the important outcomes of hospitalisation and activities of daily living. The extent to which the improved relapse free survival and overall survival might improve or maintain patients' quality of life or allow patients to participate in and perform activities of daily living is therefore unclear.

Both RCTs reported relapse free survival for subgroups. Patients who may benefit more from sorafenib more than the wider population of interest included those with undetectable MRD before allo-HSCT, those with detectable MRD after allo-HSCT and patients aged 35 years or older.

The studies identified for this review therefore provide high to moderate evidence of better relapse free survival and overall survival with sorafenib compared to placebo or no maintenance therapy in adults with FLT3-ITD AML after allo-HSCT. The impact of sorafenib on outcomes that might relate to patients' quality of life is unclear.

3. Methodology

Review questions

The review questions for this evidence review are:

1. In patients who have FLT3-ITD AML and have undergone allogeneic HSCT, what is the clinical effectiveness of sorafenib compared with standard of care?

2. In patients who have FLT3-ITD AML and have undergone allogeneic HSCT, what is the safety of sorafenib compared with standard of care?

3. In patients who have FLT3-ITD AML and have undergone allogeneic HSCT, what is the cost effectiveness of sorafenib compared with standard of care?

4. From the evidence selected, are there any subgroups of patients that may benefit from sorafenib more than the wider population of interest?

5. From the evidence selected, what were the dose regimens of sorafenib used?

See Appendix A for the full review protocol.

Review process

The methodology to undertake this review is specified by NHS England in their 'Guidance on conducting evidence reviews for Specialised Services Commissioning Products' (2020).

The searches for evidence were informed by the PICO document and were conducted on 14 November 2022.

See Appendix B for details of the search strategy.

Results from the literature searches were screened using their titles and abstracts for relevance against the criteria in the PICO document. Full text references of potentially relevant evidence were obtained and reviewed to determine whether they met the inclusion criteria for this evidence review.

See Appendix C for evidence selection details and Appendix D for the list of studies excluded from the review and the reasons for their exclusion.

Relevant details and outcomes were extracted from the included studies and were critically appraised using a checklist appropriate to the study design. See Appendices E and F for individual study and checklist details.

The available evidence was assessed by outcome for certainty using modified GRADE. See Appendix G for GRADE Profiles.

4. Summary of included studies

Two RCTs (published in three papers) were identified for inclusion. One RCT compared sorafenib to placebo in adults with FLT3-ITD AML after allo-HSCT (Burchert et al 2020). A second RCT compared sorafenib to no maintenance therapy in adults with FLT3-ITD AML after allo-HSCT (Xuan et al 2020, Xu et al 2022).

Table 1 provides a summary of the included studies and full details are given in Appendix E.

No cost effectiveness studies were identified.

Table 1: Summary of included studies

Study	Population	Intervention and comparison	Outcomes reported
Burchert et al 2020 RCT (SORMAIN) 15 centres in Germany and Austria	83 adults with FLT3-ITD AML in complete haematologic remission after allo-HSCT Sorafenib: n=43 Placebo: n=40 The authors stated that the groups were well balanced in relation to potential prognostic factors such as cytogenetic and genetic risk category and time of transplantation (i.e. whether they were in first complete remission at transplantation). The proportion of males and the proportion of patients with an ECOG-PS of 0 were higher in the placebo group Subgroup analysis considered detectable minimal residual disease level before and after allo-HSCT	**Intervention** Sorafenib 2 x 200mg orally per day for 2 weeks, then 3 x 200mg orally per day for 4 weeks, then 4 x 200mg orally per day Treatment started between 60 and 100 days after allo-HSCT and continued for 24 months or until relapse or intolerable toxicity **Comparison** Placebo for up to 24 months **Concurrent treatments** Patients could be treated with TKIs (including sorafenib), chemotherapy or a second allo-HSCT for the treatment of relapse after study entry	**Critical outcomes** • Relapse free survival (reported at 24 months and median follow-up of 42 months) • Overall survival (reported at 24 months and median follow-up of 55 months) **Important outcomes** At median (IQR) follow-up of 41.8 months (24.1 to 42.5): • Treatment adherence • Duration of therapy • Dose reductions • GVHD • Acute GVHD • Chronic GVHD • Safety • Discontinuation due to toxicity • Adverse events Grade ≥3 • Drug-related adverse events Grade ≥3
Xuan et al 2020; Xu et al 2022 RCT 7 centres in China	202 adults with FLT3-ITD AML in composite complete remission before and after allo-HSCT Sorafenib: n=100 No maintenance therapy: n=102 The authors stated that prognostic factors were well balanced between groups. The groups were also well balanced for age and gender. A similar proportion of patients in each group had received	**Intervention** Sorafenib 400mg orally twice daily. Treatment continued up to 180 days post-HSCT **Comparison** No maintenance therapy with sorafenib or another FLT3 inhibitor **Concurrent treatments** GVHD and infection prophylaxis were permitted Patients could be treated with TKIs (including sorafenib), chemotherapy	Median (IQR) follow-up 21.3 months (15.0 to 37.0) (Xuan et al 2020) Median (IQR) follow-up 36.8 months (2.5 to 67.1) (Xu et al 2022) **Critical outcomes** • Relapse free survival (reported at 2 and 3 years) • Cumulative incidence of relapse (reported at 1, 2 and 3 years) • Overall survival (reported at 2 and 3 years)

Study	Population	Intervention and comparison	Outcomes reported
	sorafenib pre-transplantation Subgroup analysis considered age group (<35 years and ≥35 years)	or donor lymphocyte infusion after relapse	**Important outcomes** • Treatment adherence (reported at median follow-up of 21.3 months): • Duration of therapy • Dose reductions • Dose interruptions • GVHD (reported at up to 210 days post-transplantation): • Acute GVHD • Chronic GVHD • Safety (reported at up to 210 days post-transplantation): • Discontinuation due to adverse events • Deaths due to adverse events • Patients with ≥1 adverse events Grade 3-4 • Adverse events (Grades 1-4) • Treatment-related adverse events
Abbreviations Allo-HSCT: allogeneic haematopoietic stem cell transplantation; AML: acute myeloid leukaemia; ECOG-PS: Eastern Cooperative Oncology Group Performance Status; FLT3-ITD: fms-like tyrosine kinase 3-internal tandem duplication; GVHD: graft-versus-host-disease; HSCT: haematopoietic stem cell transplantation; IQR: interquartile range; mg: milligram; RCT: randomised controlled trial; TKI: tyrosine kinase inhibitor			

5. Results

In patients who have FLT3-ITD AML and have undergone allo-HSCT, what is the clinical effectiveness and safety of sorafenib compared with standard of care?

Outcome	Evidence statement
Clinical Effectiveness	
Critical outcomes	
Relapse free survival **Certainty of evidence:** High to moderate	This outcome is important to patients as it represents the time for which their disease is not progressing. Stable disease might represent longer survival and that patients experience less symptoms from the disease itself. In total, two RCTs provided evidence relating to relapse free survival (RFS) in patients with FTL3-ITD AML who have undergone allo-HSCT. RFS, calculated as time from randomisation to the first occurrence of either AML relapse[10] or death from any cause, was reported in one RCT comparing sorafenib to placebo. RFS, calculated as time from transplantation to relapse[11] or death from any cause, was reported in a second RCT[12] comparing sorafenib to no maintenance therapy. This second RCT also reported cumulative incidence of relapse. *Sorafenib vs placebo* **At median 42 months:** • One RCT (Burchert et al 2020) reported *statistically significantly fewer* relapse events for sorafenib (10/43, 23.3%) vs placebo (19/40, 47.5%) at median (IQR) follow-up of 41.8 months (24.1 to 42.5) (HR for relapse or death at median follow-up: 0.39 (95%CI 0.18 to 0.85), p=0.013). Median RFS was not reached for sorafenib and was 30.9 months (CI not reported) for placebo. **(MODERATE)** **At 21-24 months:** • One RCT (Burchert et al 2020) reported *statistically significantly higher* RFS at 24 months with sorafenib (85.0% (95%CI 70 to 93)) vs placebo (53.3% (95%CI 36 to 68)) (HR for relapse or death at 24 months: 0.26 (95%CI 0.10 to 0.65), p=0.002). **(HIGH)** *Sorafenib vs no maintenance therapy* **At 3 years:** • One RCT (Xu et al 2022) reported *statistically significantly higher* RFS at three years with sorafenib (75.9% (95%CI 66.2 to 83.1)) vs no maintenance therapy (52.5% (95%CI 42.2 to 61.7)) (HR: 0.41 (95%CI 0.25 to 0.67), p<0.001). **(HIGH)** • Xu et al also reported *statistically significantly lower cumulative incidence of relapse* at three years with sorafenib (13.0% (95%CI 7.3 to 20.4)) vs no maintenance therapy (34.8% (95%CI 25.5 to 44.2)) (HR: 0.31 (95%CI 0.16 to 0.58), p<0.001). **(HIGH)** **At 21-24 months:** • One RCT (Xuan et al 2020) reported *statistically significantly higher* RFS at two years with sorafenib (78.9% (95%CI 69.0 to 85.9)) vs no maintenance

[10] Relapse was defined as loss of complete haematologic remission, according to the revised recommendations of the International Working Group (Cheson et al 2003)
[11] Relapse was defined as either reappearance of leukaemic blasts in the peripheral blood or at least 5% blasts in the bone marrow aspirate or biopsy specimen not attributable to any other causes, or reappearance or new appearance of extramedullary leukaemia
[12] The results of this RCT were reported in two different papers (Xuan et al 2020, Xu et al 2022)

Outcome	Evidence statement
	therapy (56.6% (95%CI 46.1 to 65.8)) (HR: 0.37 (95%CI 0.22 to 0.63), p<0.0001). **(HIGH)** • Xuan et al also reported 11/100 (11.0%) relapses with sorafenib and 32/102 (31.4%) relapses with no maintenance therapy at a median (IQR) follow-up of 21.3 months (15.0 to 37.0). The groups were not statistically compared. Median RFS was not reached for sorafenib or no maintenance therapy. **(MODERATE)** • Xuan et al also reported *statistically significantly lower cumulative incidence of relapse* at two years with sorafenib (11.9% (95%CI 6.2 to 19.6)) vs no maintenance therapy (31.6% (95%CI 22.6 to 41.1)) (HR: 0.29 (95%CI 0.15 to 0.58), p<0.0001). **(HIGH)** **At 1 year:** • Xuan et al (2020) reported *statistically significantly lower cumulative incidence of relapse* at one year with sorafenib (7.0% (95%CI 3.1 to 13.1)) vs no maintenance therapy (24.5% (95%CI 16.6 to 33.2)) (HR: 0.25 (95%CI 0.11 to 0.57), p=0.001). **(HIGH)** **One RCT provided high certainty evidence of statistically significantly higher RFS at 24 months for sorafenib compared to placebo and moderate certainty evidence of statistically significantly fewer relapse events for sorafenib compared to placebo at a median follow-up of 42 months. A second RCT provided high certainty evidence of statistically significantly higher RFS with sorafenib compared to no maintenance therapy at two years and three years. The same RCT also provided high certainty evidence of statistically significantly lower cumulative incidence of relapse with sorafenib compared to no maintenance therapy at one, two and three years.**
Overall survival **Certainty of evidence:** High to moderate	Overall survival is important to patients as individuals with relapsed AML have a high mortality rate due to disease. Improvement in survival is an important marker of effective treatment. In total, two RCTs provided evidence relating to overall survival in patients with FTL3-ITD AML who have undergone allo-HSCT. Overall survival, calculated as time from randomisation to death from any cause, was reported in one RCT comparing sorafenib to placebo. Overall survival, calculated as time from transplantation to death from any cause, was reported in a second RCT comparing sorafenib to no maintenance therapy. *Sorafenib vs placebo* **At median 55 months:** • One RCT (Burchert et al 2020) reported *no statistically significant difference* in deaths between sorafenib (11/43, 25.6%) and placebo (16/40, 40.0%) at median follow-up of 55.1 months (IQR for follow-up not reported) (HR for death at median follow-up: 0.52 (95%CI 0.24 to 1.11), p=0.086). Median overall survival was not reached for sorafenib or placebo. **(MODERATE)** **At 21-24 months:** • One RCT (Burchert et al 2020) reported *statistically significantly higher* overall survival at 24 months with sorafenib (90.5% (95%CI 77 to 96)) vs placebo (66.2% (95%CI 49 to 79)) (HR for death at 24 months: 0.24 (95%CI 0.08 to 0.74), p=0.007). **(HIGH)** *Sorafenib vs no maintenance therapy* **At 3 years:** • One RCT (Xu et al 2022) reported *statistically significantly higher* overall survival at three years with sorafenib (79.5% (95%CI 69.6 to 85.8)) vs no maintenance therapy (61.4% (95%CI 51.1 to 70.1)) (HR: 0.48 (95%CI 0.28 to 0.82), p=0.005). **(MODERATE)** • Xu et al also reported 21/100 (21.0%) deaths with sorafenib and 39/102 (38.2%) deaths with no maintenance therapy at a median (IQR) follow-up of

Outcome	Evidence statement
	36.8 months (2.5 to 67.1). The groups were not statistically compared. **(MODERATE)** **At 21-24 months:** • One RCT (Xuan et al 2020) reported *statistically significantly higher* overall survival at two years with sorafenib (82.1% (95%CI 72.6 to 88.5)) vs no maintenance therapy (68.0% (95%CI 57.8 to 76.2)) (HR: 0.48 (95%CI 0.27 to 0.86), p=0.012). **(MODERATE)** • Xuan et al also reported 17/100 (17.0%) deaths with sorafenib and 32/102 (31.4%) deaths with no maintenance therapy at a median (IQR) follow-up of 21.3 months (15.0 to 37.0). The groups were not statistically compared. Median overall survival was not reached for sorafenib or no maintenance therapy. **(MODERATE)** **One RCT provided high certainty evidence of statistically significantly higher overall survival at 24 months for sorafenib compared to placebo. In this RCT there was moderate certainty evidence that the difference in number of deaths between the groups was not statistically significant at a median follow-up of 55 months. A second RCT provided moderate certainty evidence of statistically significantly higher overall survival for sorafenib compared to no maintenance therapy at two and three years.**
Quality of life **Certainty of evidence:** Not applicable	Quality of life is important to patients as it provides an indication of an individual's general health and self-perceived well-being and their ability to participate in activities of daily living. Measurement of quality of life can help inform patient-centred decision making and inform health policy. **No evidence was identified for this outcome.**
Important outcomes	
Hospitalisation **Certainty of evidence:** Not applicable	This outcome is important to patients as it may represent either disease progression or treatment toxicity. It may have a bearing on the patient's quality of life and inform their treatment decision making. **No evidence was identified for this outcome.**
Treatment adherence **Certainty of evidence:** Moderate	Adherence to treatment is important to patients as it provides an indication of how the treatment is tolerated. If a treatment has adherence challenges, it can increase the risk of treatment failure and add to relapse risk. In total, two RCTs provided evidence for treatment adherence related outcomes in patients with FTL3-ITD AML who have undergone allo-HSCT. One RCT compared sorafenib to placebo. A second RCT compared sorafenib to no maintenance therapy, but only reported adherence to treatment for sorafenib patients. Treatment adherence related outcomes reported were duration of therapy, dose reductions and dose interruptions. *Sorafenib vs placebo* **At median 42 months:** • One RCT (Burchert et al 2020) reported the median (range) duration of therapy as 34.6 weeks (1.3 to 106.9) for sorafenib patients and 54.4 weeks (1.7 to 128.3) for placebo patients at a median (IQR) follow-up of 41.8 months (24.1 to 42.5). The groups were not statistically compared. **(MODERATE)** • One RCT (Burchert et al 2020) reported dose reductions for 21 of 43 (49%) sorafenib patients and 16 of 40 (40%) placebo patients at a median (IQR) follow-up of 41.8 months (24.1 to 42.5). The groups were not statistically compared. **(MODERATE)**

Outcome	Evidence statement
	Sorafenib (no comparator) **At median 21 months:** • One RCT (Xuan et al 2020) reported the median (IQR) duration of therapy as 134 days (116 to 150) for sorafenib patients at a median (IQR) follow-up of 21.3 months (15.0 to 37.0). **(MODERATE)** • One RCT (Xuan et al 2020) reported dose reductions due to adverse events for 42 of 100 (42%) sorafenib patients at a median (IQR) follow-up of 21.3 months (15.0 to 37.0). **(MODERATE)** • One RCT (Xuan et al 2020) reported dose interruptions due to adverse events for 12 of 100 (12%) sorafenib patients at a median (IQR) follow-up of 21.3 months (15.0 to 37.0). **(MODERATE)** **One RCT provided moderate certainty evidence that sorafenib patients had a shorter duration of therapy and a higher proportion of dose reductions than placebo patients at a median follow-up of 42 months. The groups were not statistically compared. A second RCT reported moderate certainty evidence that patients received sorafenib for a median of 134 days with dose reductions and dose interruptions due to adverse events in 42% and 12% respectively.**
Graft-versus-host-disease (GVHD) **Certainty of evidence:** Moderate	This is important to patients since acute or chronic GVHD is a potentially serious complication of allogeneic stem cell transplantation and reduced-intensity allogeneic stem cell transplantation which requires further management. In total, two RCTs provided evidence relating to GVHD in patients with FTL3-ITD AML who have undergone allo-HSCT. Acute and chronic GVHD[13] were reported by one RCT comparing sorafenib to placebo. Acute and chronic GVHD[14] were also reported by a second RCT comparing sorafenib to no maintenance therapy. *Sorafenib vs placebo* **At median 42 months:** • One RCT (Burchert et al 2020) reported acute GVHD (Grade ≥2) for 10 of 42 (24%) sorafenib patients and seven of 39 (18%) placebo patients at a median (IQR) follow-up of 41.8 months (24.1 to 42.5). The groups were not statistically compared. **(MODERATE)** • One RCT (Burchert et al 2020) reported severe chronic GVHD for 8 of 42 (19%) sorafenib patients and 4 of 39 (10%) placebo patients at a median (IQR) follow-up of 41.8 months (24.1 to 42.5). The groups were not statistically compared. **(MODERATE)** • One RCT (Burchert et al 2020) reported mild/moderate chronic GVHD for 18 of 42 (43%) sorafenib patients and 14 of 39 (36%) placebo patients at a median (IQR) follow-up of 41.8 months (24.1 to 42.5). The groups were not statistically compared. **(MODERATE)** *Sorafenib vs no maintenance therapy* **At up to 210 days:** • One RCT (Xuan et al 2020) reported acute GVHD (Grade ≥2) for 23 of 100 (23%) sorafenib patients and 21 of 102 (21%) no maintenance therapy patients at up to 210 days post-transplantation. The groups were not statistically compared. **(MODERATE)** • One RCT (Xuan et al 2020) reported acute GVHD (Grade 1) for eight of 100 (8%) sorafenib patients and six of 102 (6%) no maintenance therapy patients

[13] Acute GVHD categorised according to the Mount Sinai Acute GVHD International Consortium (Harris et al 2016). Chronic GVHD categorised according to the National Institutes of Health consensus criteria (Filipovich et al 2005)
[14] Acute and chronic GVHD were graded according to published guidelines (Prezpiorka et al 1995, Jagasia et al 2015)

Outcome	Evidence statement
	at up to 210 days post-transplantation. The groups were not statistically compared. **(MODERATE)** • One RCT (Xuan et al 2020) reported moderate/severe chronic GVHD for 18 of 99 (18%) sorafenib patients and 17 of 99 (17%) no maintenance therapy patients at up to 210 days post-transplantation. The groups were not statistically compared. **(MODERATE)** • One RCT (Xuan et al 2020) reported mild chronic GVHD for five of 99 (5%) sorafenib patients and five of 99 (5%) no maintenance therapy patients at up to 210 days post-transplantation. The groups were not statistically compared. **(MODERATE)** **One RCT provided moderate certainty evidence that a higher proportion of sorafenib patients had acute and chronic GVHD than placebo patients at a median follow-up of 42 months. A second RCT provided moderate certainty evidence that sorafenib and no maintenance therapy patients had similar proportions of acute and chronic GVHD at up to 210 days post-transplantation. The groups were not statistically compared in either RCT.**
Activities of daily living (ADLs) **Certainty of evidence:** Not applicable	ADLs are important outcomes to patients as they facilitate enablement and independence, allowing individuals to function in education, work, home and recreational settings. They encompass patients' individual needs and facilitate inclusion and participation. The complications of recurrence can lead to progressively worsening physical symptoms and altered ability to complete ADLs without assistance. **No evidence was identified for this outcome.**
Safety	
Safety **Certainty of evidence:** Moderate	Safety of sorafenib is important to patients as it allows comparison of interventional approaches. In total, two RCTs provided evidence relating to safety outcomes in patients with FTL3-ITD AML who have undergone allo-HSCT. Adverse events[15] and drug discontinuation due to toxicity were reported by one RCT comparing sorafenib to placebo. Adverse events[16] and drug discontinuation due to adverse events were reported by a second RCT comparing sorafenib to no maintenance therapy. *Sorafenib vs placebo* **At median 42 months:** • One RCT (Burchert et al 2020) reported study drug discontinuation due to toxicity for nine of 42 (21%) sorafenib patients and two of 39 (5%) placebo patients at a median (IQR) follow-up of 41.8 months (24.1 to 42.5). The groups were not statistically compared. **(MODERATE)** • Burchert et al also reported adverse events of Grade ≥3 for each group at a median (IQR) follow-up of 41.8 months (24.1 to 42.5). The proportion of patients in each group with any adverse event of Grade ≥3 was not reported and the groups were not statistically compared. The most common (>10% of patients) adverse events with sorafenib were infections (26%), GI toxicity (14%), electrolyte alterations (14%), skin toxicity (12%), cardiotoxicity and renal insufficiency (10%) and other (not further defined) (79%). The most common (>10% of patients) adverse events with placebo were infections (23%), GI toxicity (15%) and other (not further defined) (56%). **(MODERATE)**

[15] Grading criteria not specified

[16] Non-haematological adverse events defined using the National Cancer Institute Common Terminology Criteria for Adverse Events (version 4.0) where Grade 1 = mild; Grade 2 = moderate; Grade 3 = severe or medically significant but not immediately life threatening; Grade 4 = life-threatening consequences; Grade 5 = death related to adverse event. Grade 3 haematological adverse events were defined as either an absolute neutrophil count <1.0 x 10^9 cells/L but ≥ 0.5 x 10^9 cells/L or a platelet count <30 x 10^9/L but ≥20 x 10^9/L. Grade 4 haematological adverse events were defined as either an absolute neutrophil count <0.5 x 10^9 cells/L or a platelet count <20 x 10^9/L

Outcome	Evidence statement
	• Burchert et al also reported drug-related adverse events of Grade ≥3 for each group at a median (IQR) follow-up of 41.8 months (24.1 to 42.5). The proportion of patients in each group with any drug-related adverse event of Grade ≥3 was not reported and the groups were not statistically compared. The most common (>5% of patients) drug-related adverse events with sorafenib were electrolyte alterations (7%), skin toxicity (5%), GI toxicity (5%) and other (not further defined) (19%). The most common (>5% of patients) drug-related adverse events with placebo were GI toxicity (8%), infections (5%), liver toxicity (5%) and other (not further defined) 10%. **(MODERATE)** *Sorafenib vs no maintenance therapy* **At up to 210 days:** • One RCT (Xuan et al 2020) reported study drug discontinuation due to adverse events for five of 100 (5%) sorafenib patients at up to 210 days post-transplantation. **(MODERATE)** • Xuan et al also reported deaths due to adverse events for four of 100 (4%) sorafenib patients and five of 102 (5%) no maintenance therapy patients at up to 210 days post-transplantation. The groups were not statistically compared. **(MODERATE)** • Xuan et al also reported that 50/100 (50%) sorafenib patients and 47/102 (46%) no maintenance therapy patients experienced at least one adverse event of Grade 3 or 4, at up to 210 days post-transplantation. The groups were not statistically compared. **(MODERATE)** • In Xuan et al the most common (>10% of patients) Grade 3-4 adverse events with sorafenib were infections (25%), haematologic toxicity (15%) and gastrointestinal (11%). The most common (>10% of patients) Grade 3-4 adverse event with no maintenance therapy was infections (24%). Adverse events were assessed up to 210 days post-transplantation. **(MODERATE)** • In Xuan et al the most common (>10% of patients) Grade 1-2 adverse events with sorafenib were gastrointestinal (25%), renal or genitourinary (23%), skin related (20%), hepatobiliary or pancreatic (16%) and cardiac (14%). The most common (>10% of patients) adverse events with no maintenance therapy were renal or genitourinary (25%), gastrointestinal (20%), hepatobiliary or pancreatic (17%) and cardiac (12%). Adverse events were assessed up to 210 days post-transplantation. **(MODERATE)** • Xuan et al also stated that the most common Grade 3-4 treatment-related adverse events with sorafenib were skin-related (7%) or haematological (5%) and that no patients died from treatment-related adverse events. Adverse events were assessed up to 210 days post-transplantation. The proportion of patients in each group with any drug-related adverse events was not reported. **(MODERATE)** **One RCT provided moderate certainty evidence of higher drug discontinuation due to toxicity with sorafenib than placebo at a median follow-up of 42 months. The groups were not statistically compared. The same RCT specified adverse events and drug-related adverse events of Grade ≥3 experienced by each group at a median follow-up of 42 months. However, the proportion of patients in each group with any adverse event of Grade ≥3 was not reported and the groups were not statistically compared. A second RCT provided moderate certainty evidence of similar proportions of deaths due to adverse events and patients experiencing at least one Grade 3 or 4 adverse event with sorafenib or no maintenance therapy, assessed up to 210 days post-transplantation. The groups were not statistically compared. The same RCT reported discontinuation of sorafenib due to adverse events in 5% of patients. This RCT also specified adverse events and treatment-related adverse events experienced by each group.**

Outcome	Evidence statement
Abbreviations ADLs: activities of daily living; Allo-HSCT: allogeneic haematopoietic stem cell transplantation; AML: acute myeloid leukaemia; CI: confidence intervals; FLT3-ITD: fms-like tyrosine kinase 3-internal tandem duplication; GI: gastrointestinal; GVHD: graft-versus-host-disease; HR: hazard ratio; HSCT: haematopoietic stem cell transplantation; IQR: interquartile range; RCT: randomised controlled trial; RFS: relapse free survival	

In patients who have FLT3-ITD AML and have undergone allo-HSCT, what is the cost effectiveness of sorafenib compared with standard of care?

Outcome	Evidence statement
Cost effectiveness	**No evidence was identified for cost effectiveness.**

From the evidence selected, are there any subgroups of patients that may benefit from sorafenib more than the wider population of interest?

Outcome	Evidence statement
Subgroups	Subgroup results comparing sorafenib and placebo for relapse free survival were reported by one RCT. A second RCT reported cumulative incidence of relapse separately by age group. Neither RCT reported outcomes according to whether patients had received one, or more than one, allo-HSCT. The subgroup analyses were exploratory and post-hoc. **Relapse free survival (RFS)** *Sorafenib vs placebo* • One RCT (Burchert et al 2020) reported that RFS was *statistically significantly higher* with sorafenib than placebo for the following subgroups: • Patients with undetectable minimal residual disease (MRD) before allo-HSCT (0/9 relapsed or died with sorafenib vs 5/12 with placebo, p=0.028) • Patients with detectable MRD after allo-HSCT (p=0.015) (n not reported) Median follow-up was 41.8 months (24.1 to 42.5). *Sorafenib vs no maintenance therapy* • One RCT (Xuan et al 2020) reported a *statistically significantly lower* cumulative incidence of relapse at two years for sorafenib (8.0% (95%CI 2.5 to 17.7)) vs no maintenance therapy (38.7% (95%CI 24.4 to 52.7)) for patients aged ≥35 years (n=99) (HR 0.17 (95%CI 0.06 to 0.50), p not reported). For patients aged <35 years (n=103) the difference between sorafenib (16.1% (95%CI 6.8 to 28.9)) and no maintenance therapy (25.1% (95%CI 14.2 o 37.7)) was *not statistically significant* (HR 0.45 (95%CI 0.18 to 1.11), p not reported). **One RCT reported that RFS was statistically significantly higher for sorafenib vs placebo for patients with undetectable MRD before allo-HSCT and patients with detectable MRD after allo-HSCT. A second RCT reported that cumulative incidence of relapse at two years was statistically significantly lower for sorafenib vs no maintenance therapy for patients aged ≥35 years, but not for patients aged <35 years.**
Abbreviations Allo-HSCT: allogeneic haematopoietic stem cell transplantation; CI: confidence intervals; HR: hazard ratio; HSCT: haematopoietic stem cell transplantation; MRD: minimal residual disease; RCT: randomised controlled trial; RFS: relapse free survival	

From the evidence selected, what were the dose regimens of sorafenib used?

Outcome	Evidence statement
Dose regimens of sorafenib	In the RCT by Burchert et al (2020), the starting dose of sorafenib was 2 x 200mg orally per day for two weeks (dose level 1). This was followed by 3 x 200mg orally per day for four weeks (dose level 2), then 4 x 200mg orally per day (dose level 3). Dose reductions were permitted. Treatment started between 60 and 100 days after allo-HSCT and continued for 24 months or until relapse or intolerable toxicity. In the RCT by Xuan et al (2020), the sorafenib dose was 2 x 400mg orally per day. Dose reductions or interruptions were allowed if adverse events of Grade ≥3 occurred. Dose reductions were to 200mg once or twice a day with return to 400mg twice a day after the resolution of adverse events. Treatment started between 30 and 60 days after allo-HSCT and continued up to 180 days post-transplantation.
Abbreviations Allo-HSCT: allogeneic haematopoietic stem cell transplantation; mg: milligram; RCT: randomised controlled trial	

6. Discussion

This evidence review considered the clinical effectiveness and safety of sorafenib as maintenance therapy compared with standard of care for the treatment of patients who have FLT3-ITD AML and have undergone allo-HSCT. The critical outcomes of interest were relapse free survival, overall survival and quality of life. Important outcomes were hospitalisation, treatment adherence, GVHD, ADLs and safety. Evidence on cost effectiveness was also sought.

Evidence was available from two RCTs. One RCT (Burchert et al 2020) compared sorafenib to placebo and was conducted at 15 centres in Germany and Austria. The second RCT (Xuan et al 2020) compared sorafenib to no maintenance therapy and was conducted at seven centres in China. It is not clear to what extent the results of these studies might be generalisable to the NHS in England. Both RCTs provided data for the critical outcomes of relapse free survival and overall survival and the important outcomes of treatment adherence, GVHD and safety. No evidence was identified for the critical outcome of quality of life, or the important outcomes of hospitalisation and activities of daily living. No evidence was identified on cost effectiveness.

Both RCTs included adults with FLT3-ITD AML who had undergone allo-HSCT. In both RCTs, the groups were similar at baseline for prognostic factors. In one RCT (Burchert et al 2020) there were more males and more patients with an Eastern Cooperative Oncology Group performance status (ECOG-PS) of zero (fully active) in the placebo group. Other patients in this RCT had an ECOG-PS score of one (restrictions in physically strenuous activity), where this was known. In the second RCT (Xuan et al 2020) patients had an ECOG-PS score of between zero and two (ambulatory and capable of self-care but not work activities). No evidence about patients' capacity in relation to quality of life or ADLs during treatment or during the study follow-up periods (up to three to four years) was reported by either RCT.

The dosing regimen differed between the RCTs. In the RCT by Burchert et al, patients received an escalating dose equating to 400mg per day initially and rising to a maximum of 800mg per day. In the RCT by Xuan et al, patients started at a dose of 800mg per day with dose reductions or interruptions allowed if adverse events of Grade ≥3 occurred. Treatment started between 60 and 100 days after allo-HSCT in Burchert et al and continued for 24 months or until intolerable toxicity. In Xuan et al, treatment started between 30 and 60 days after allo-HSCT and continued for up to 180 days post-transplantation. The median length of time that patients received sorafenib was approximately 35 weeks in Burchert et al and approximately 19 weeks in Xuan et al. In addition, no patients in Burchert et al had received sorafenib pre-transplant, whereas this was approximately 25% in Xuan et al. These differences may have contributed to the difference in the proportion of patients who discontinued sorafenib due to adverse events observed in the two RCTs.

The definitions of relapse-free survival and overall survival differed in the two RCTs, with the RCT by Burchert et al calculating survival from time of randomisation and the RCT by Xuan et al calculating survival from time of transplantation. The time between allo-HSCT and randomisation was not stated in either RCT. However, the maximum time between transplantation and starting sorafenib was 100 days and patients in both RCTs were in remission when randomised. The difference in the definitions is unlikely to have had an impact for the follow-up timepoints reported.

The RCT by Burchert et al included 43 sorafenib patients and 40 placebo patients. The power calculation, based on the primary endpoint of relapse-free survival, estimated that 200 patients were needed. However, study recruitment was terminated early due to slow patient recruitment. The study may not have been sufficiently powered. The RCT by Xuan

et al was sufficiently powered to show a difference between treatment groups with a two-sided significance level of 5% and 90% power, based on the primary endpoint of cumulative incidence of relapse at one year.

Both RCTs followed-up all patients and conducted an intention-to-treat analysis. The duration of follow-up was sufficient for the outcomes reported.

In the RCT by Burchert et al, the patients and investigators were blinded to study group, but it was not clear if the outcome assessors were. In the RCT by Xuan et al, the patients and investigators were not blinded to study group but the staff who did the data analysis and assessment of outcomes were. As the outcomes reported in both RCTs were objective and/or had standardised definitions or assessment criteria the risk from any lack of blinding impacting the results is low.

No information about what any minimal clinically important thresholds or differences might be was reported for any of the outcomes considered. However, Burchert et al commented in their discussion that the relapse rate observed after two years appeared to be a clinically meaningful improvement.

The RCTs were well conducted with few risk of bias issues that were likely to impact on the outcomes reported. However, statistical comparison between the groups was not reported for treatment adherence, GVHD or safety outcomes and some outcomes were downgraded for imprecision due to wide confidence intervals around a hazard ratio. In the RCT by Xuan et al, the comparator used (no maintenance therapy) limited the availability of comparative data for treatment adherence and some safety outcomes.

Subgroup results comparing sorafenib and placebo for relapse free survival were reported by Burchert et al. The RCT by Xuan et al reported cumulative incidence of relapse separately for the (adult) patients who were aged more or less than 35 years. The subgroup analyses were exploratory and post-hoc. Neither RCT reported outcomes according to whether patients had received one, or more than one allo-HSCT. In the RCT by Burchert et al, some patients received a second allo-HSCT after relapse. However, it is not clear if any patients received sorafenib after a second allo-HSCT. In the RCT by Xuan et al, no patients were reported to have received more than one allo-HSCT.

7. Conclusion

This evidence review includes one RCT comparing sorafenib to placebo and a second RCT comparing sorafenib to no maintenance therapy. The populations of both studies were adults with FLT3-ITD AML after allo-HSCT. There was no evidence on cost effectiveness.

There were RCT data comparing sorafenib to placebo or no maintenance therapy for the critical outcomes of relapse free survival and overall survival and the important outcomes of treatment adherence, GVHD and safety. Both RCTs reported a statistically significant advantage for sorafenib for relapse free survival and overall survival up to at least two years.

The two RCTs were both well conducted with few risk of bias issues that were likely to impact the outcomes reported. However, statistical comparison between groups was not reported for treatment adherence, GVHD or safety outcomes which limits the interpretation of these results. In addition, some outcomes were downgraded for imprecision due to wide confidence intervals around a hazard ratio. In the RCT by Xuan et al, the comparator used (no maintenance therapy) limited the availability of comparative data for treatment adherence and some safety outcomes.

No evidence was identified for the critical outcome of quality of life, or the important outcomes of hospitalisation and activities of daily living. The extent to which the improved relapse free survival and overall survival might improve or maintain patients' quality of life or allow patients to participate in and perform activities of daily living is therefore unclear.

Both RCTs reported relapse free survival for subgroups. Patients who may benefit more from sorafenib more than the wider population of interest included those with undetectable MRD before allo-HSCT, those with detectable MRD after allo-HSCT and patients aged 35 years or older.

The studies identified for this review therefore provide high to moderate evidence of better relapse free survival and overall survival with sorafenib compared to placebo or no maintenance therapy in adults with FLT3-ITD AML after allo-HSCT. The impact of sorafenib on outcomes that might relate to patients' quality of life is unclear.

Appendix A PICO Document

The review questions for this evidence review are:

1. In patients who have FLT3-ITD AML and have undergone allogeneic HSCT, what is the clinical effectiveness of sorafenib compared with standard of care?
2. In patients who have FLT3-ITD AML and have undergone allogeneic HSCT, what is the safety of sorafenib compared with standard of care?
3. In patients who have FLT3-ITD AML and have undergone allogeneic HSCT, what is the cost effectiveness of sorafenib compared with standard of care?
4. From the evidence selected, are there any subgroups of patients that may benefit from sorafenib more than the wider population of interest?
5. From the evidence selected, what were the dose regimens of sorafenib used?

P-Population and Indication	Individuals with FLT3-ITD AML who have undergone allogeneic haematopoietic stem cell transplantation (allo-HSCT). Subgroups of particular interest • One or more than one allo-HSCT treatments • Age [Patients may or may not have received FLT3 Tyrosine Kinase Inhibitors prior to HSCT]
I-Intervention	Sorafenib as maintenance therapy [A tyrosine kinase inhibitor given at a starting dose of 200mg twice daily as an oral agent to a maximum of 400mg BD usually commenced at 60-100 days post allo-HSCT] [Sorafenib is to be administered for a maximum of 24 months, or until toxicity occurs] [Maintenance therapy in this context is defined as the ongoing treatment of FLT3-ITD AML after the patient has received allo-HSCT]
C-Comparator	No treatment Placebo [There is currently no maintenance therapy available for these patients to prevent disease relapse post allo-HSCT]
O-Outcomes	**Clinical Effectiveness** Minimally clinically important difference (MCIDs) are not known unless stated. *Critical to decision-making:* • **Relapse free survival** *This outcome is important to patients as it represents the time for which their disease is not progressing. Stable disease might represent longer survival and that patients experience less symptoms from the disease itself.* [The time interval from transplant until relapse of AML or death from any cause, whichever occurs first. This may be measured in days, weeks, months, or years. Relapse is defined as any

	blast appearance in the peripheral blood, in the bone marrow (>5%) or extramedullary blasts (chloroma).] • **Overall survival** *Overall survival is important to patients as individuals with relapsed AML have a high mortality rate due to disease. Improvement in survival is an important marker of effective treatment.* • **Quality of life** *Quality of life is important to patients as it provides an indication of an individual's general health and self-perceived well-being and their ability to participate in activities of daily living. Measurement of quality of life can help inform patient-centred decision making and inform health policy.* [Examples include, but not limited to: o Acute Myeloid Leukaemia – Quality of Life (AML-QOL) o EuroQol Eq-5D-3L o Haematological Malignancy Patient Reported Outcome (HM-PRO)] *Important to decision-making:* • **Hospitalisation** *This outcome is important to patients as it may represent either disease progression or treatment toxicity. It may have a bearing on the patient's quality of life and inform their treatment decision making.* • **Treatment adherence** *Adherence to treatment is important to patients as it provides an indication of how the treatment is tolerated. If a treatment has adherence challenges, it can increase the risk of treatment failure and add to relapse risk.* [Examples of relevant outcome measures include, but are not limited to: o Missed doses (observed by research staff review of medication/returned medication) o Self-reported adherence measures (e.g., questionnaire methods)] • **Graft-versus-host-disease (GVHD)** *This is important to patients since acute or chronic GVHD is a potentially serious complication of allogeneic stem cell transplantation and reduced-intensity allogeneic stem cell transplantation which requires further management.* • **Activities of daily living (ADLs)** *ADLs are important outcomes to patients as they facilitate enablement and independence, allowing individuals to function in education, work, home, and recreational settings. They encompass patients' individual needs and facilitate inclusion and participation. The complications of recurrent LGSOC can lead to progressively worsening physical symptoms and altered ability to complete ADLs without assistance.* [ADLs can be measured using assessments such as:

	○ Timed task completion (e.g., timed repeatable test such as dressing, meal preparation or patient specific ADL goal) ○ ADLs assessment using a tool (e.g., Barthel Index (BI) or Independence in Activities of Daily Living (ADL) ○ Subjective/self-reported assessment (e.g., by the individual, carer, or MDT. This could include self-reported questionnaires such as participation in work and other activities).] **Safety** Safety of sorafenib is important to patients as it allows comparison of interventional approaches. [Examples include, but not limited to: • Toxicity • Frequency of adverse events • Frequency of grade 3 or 4 adverse events • Adverse events leading to discontinuation • Treatment related adverse events] **Cost effectiveness**
Inclusion criteria	
Study design	Systematic reviews, randomised controlled trials, controlled clinical trials, cohort studies. If no higher-level quality evidence is found, case series can be considered.
Language	English only
Patients	Human studies only
Age	All ages
Date limits	2012-2022
Exclusion criteria	
Publication type	Conference abstracts, non-systematic reviews, narrative reviews, commentaries, letters, editorials and guidelines
Study design	Case reports, resource utilisation studies

Appendix B Search strategy

Medline, Embase, the Cochrane Library, PubMed and the TRIP database were searched limiting the search to papers published in the English language in the last 10 years. Conference abstracts, commentaries, letters, editorials, case reports and trial registrations were excluded.

Search dates: 1 January 2012 to 14 November 2022

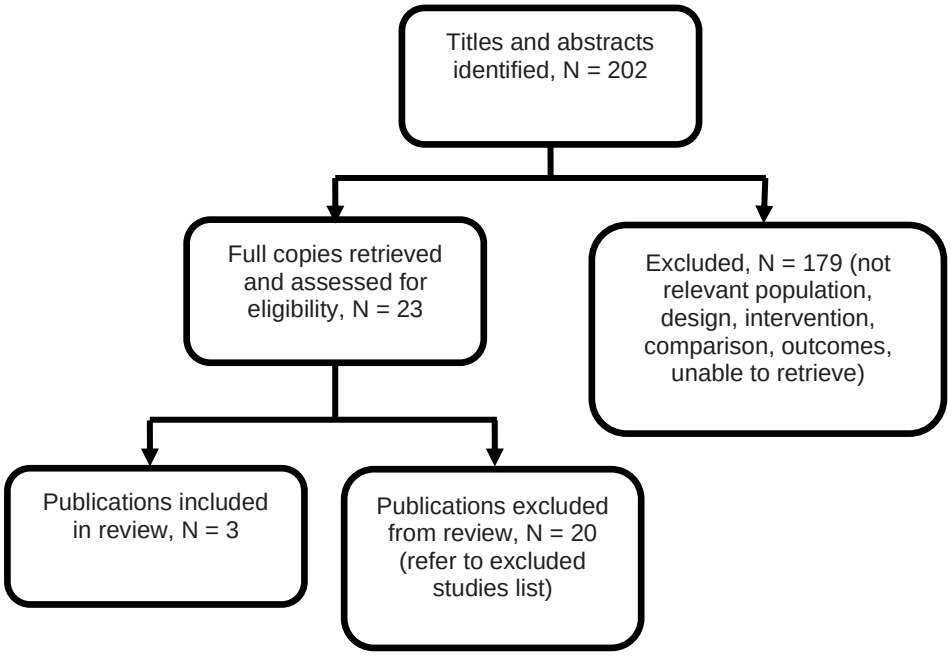
Medline search strategy:

1	leukemia, myeloid/ or exp leukemia, myeloid, acute/
2	(leukaemia? or leukemia? or aml).ti,ab,kf.
3	1 or 2
4	stem cell transplantation/ or exp hematopoietic stem cell transplantation/
5	(allogenic or allogeneic or allograft* or allo-graft*).ti,ab,kf.
6	(h?ematopoietic adj5 (transplant* or stem cell)).ti,ab,kf.
7	(hsct or allo-hsct or allhsct).ti,ab,kf.
8	((stem or cell or bone marrow) adj2 transplant*).ti,ab,kf.
9	4 or 5 or 6 or 7 or 8
10	Sorafenib/
11	(sorafenib or nexavar).ti,ab,kf.
12	10 or 11
13	3 and 9 and 12
14	(comment or editorial or letter or preprint).pt.
15	13 not 14
16	limit 15 to (english language and yr="2012 -Current")

Appendix C Evidence selection

The literature search identified 202 potential references. These were screened using their titles and abstracts and 23 references potentially relating to the use of sorafenib for FLT3-ITD AML after allo-HSCT were obtained in full text and assessed for relevance. Of these, three references are included in this evidence review. The 20 references excluded are listed in Appendix D.

Figure 1- Study selection flow diagram



References submitted with Preliminary Policy Proposal

Reference	Paper selection decision and rationale if excluded
Burchert A, Bug G, Fritz LV, Finke J, Stelljes M, Röllig C et al. Sorafenib maintenance after allogenic hematopoietic stem cell transplantation for acute myeloid leukemia with FLT3-internal tandem duplication mutation (SORMAIN). Journal of Clinical Oncology 2020, 38(26):2993-3003.	Included in the review
Gagelmann N, Wolschke C, Klyuchnikov E, Christopeit M, Ayuk F, Kröger N. TKI maintenance after stem-cell transplantation for FLT3-ITD positive acute myeloid leukaemia: a systematic review and meta-analysis. Frontiers in Immunology 2021, 12(630429):1-10.	Meta analysis combines different comparators and study designs. Individual studies considered separately for eligibility for inclusion in this review.
Xuan L, Wang Y, Huang F, Fan Z, Xu Y, Sun J et al. Sorafenib maintenance in patients with FLT3-ITD acute myeloid leukaemia undergoing allogenic haematopoietic stem cell transplantation: an open-label, multicentre, randomised phase 3 trial. The Lancet 2020, 21(1):1201-1212.	Included in the review

Appendix D Excluded studies table

Study reference	Reason for exclusion
Antar A, Kharfan-Dabaja MA, Mahfouz R, Bazarbachi A. Sorafenib Maintenance Appears Safe and Improves Clinical Outcomes in FLT3-ITD Acute Myeloid Leukemia After Allogeneic Hematopoietic Cell Transplantation. Clin Lymphoma Myeloma Leuk. 2015;15(5):298-302.	Non-comparative study. RCT evidence available for the outcomes reported in this study.
Aydin S, Passera R, Scaldaferri M, Dellacasa CM, Poggiu M, Cattel F, et al. Sorafenib maintenance after hematopoietic stem cell transplantation improves outcome of FLT3-ITD-mutated acute myeloid leukemia. Int J Hematol. 2022;09:09.	Non-randomised study. RCT evidence available for the outcomes reported in this study.
Battipaglia G, Massoud R, Ahmed SO, Legrand O, El Cheikh J, Youniss R, et al. Efficacy and Feasibility of Sorafenib as a Maintenance Agent After Allogeneic Hematopoietic Stem Cell Transplantation for Fms-like Tyrosine Kinase 3 Mutated Acute Myeloid Leukemia: An Update. Clin Lymphoma Myeloma Leuk. 2019;19(8):506-8.	Non-comparative study. RCT evidence available for the outcomes reported in this study.
Battipaglia G, Ruggeri A, Massoud R, El Cheikh J, Jestin M, Antar A, et al. Efficacy and feasibility of sorafenib as a maintenance agent after allogeneic hematopoietic stem cell transplantation for Fms-like tyrosine kinase 3-mutated acute myeloid leukemia. Cancer. 2017;123(15):2867-74.	Non-comparative study. RCT evidence available for the outcomes reported in this study.
Bazarbachi A, Labopin M, Battipaglia G, Djabali A, Forcade E, Arcese W, et al. Allogeneic Stem Cell Transplantation for FLT3-Mutated Acute Myeloid Leukemia: In vivo T-Cell Depletion and Posttransplant Sorafenib Maintenance Improve Survival. A Retrospective Acute Leukemia Working Party-European Society for Blood and Marrow Transplant Study. Clin Hematol Int. 2019;1(1):58-74.	Non-comparative study. RCT evidence available for the outcomes reported in this study.
Bewersdorf JP, Allen C, Mirza AS, Grimshaw AA, Giri S, Podoltsev NA, et al. Hypomethylating Agents and FLT3 Inhibitors As Maintenance Treatment for Acute Myeloid Leukemia and Myelodysplastic Syndrome After Allogeneic Hematopoietic Stem Cell Transplantation-A Systematic Review and Meta-Analysis. Transplant Cell Ther. 2021;27(12):997.e1-.e11.	Meta analysis combines different FLT3 inhibitors, including interventions not in scope. Individual studies considered separately for eligibility for inclusion in this review.
Brunner AM, Li S, Fathi AT, Wadleigh M, Ho VT, Collier K, et al. Haematopoietic cell transplantation with and without sorafenib maintenance for patients with FLT3-ITD acute myeloid leukaemia in first complete remission. Br J Haematol. 2016;175(3):496-504.	Non-randomised study. RCT evidence available for the outcomes reported in this study.
Chen YB, Li S, Lane AA, Connolly C, Del Rio C, Valles B, et al. Phase I trial of maintenance sorafenib after allogeneic hematopoietic stem cell transplantation for fms-like tyrosine kinase 3 internal tandem duplication acute myeloid leukemia. Biol Blood Marrow Transplant. 2014;20(12):2042-8.	Non-comparative study. RCT evidence available for the outcomes reported in this study.
De Freitas T, Marktel S, Piemontese S, Carrabba MG, Tresoldi C, Messina C, et al. High rate of hematological responses to sorafenib in FLT3-ITD acute myeloid leukemia relapsed after allogeneic hematopoietic stem cell transplantation. Eur J Haematol. 2016;96(6):629-36.	Population receiving sorafenib as salvage therapy. Not the population of interest.
Gagelmann N, Wolschke C, Klyuchnikov E, Christopeit M, Ayuk F, Kröger N. TKI Maintenance After Stem-Cell Transplantation for FLT3-ITD Positive Acute Myeloid Leukemia: A Systematic Review and Meta-Analysis. Front Immunol. 2021;12:630429.	Meta analysis combines different comparators and study designs. Individual studies considered separately for eligibility for inclusion in this review.

Study reference	Reason for exclusion
Metzelder SK, Schroeder T, Finck A, Scholl S, Fey M, Gotze K, et al. High activity of sorafenib in FLT3-ITD-positive acute myeloid leukemia synergizes with allo-immune effects to induce sustained responses. Leukemia. 2012;26(11):2353-9.	Population receiving sorafenib as salvage therapy. Not the population of interest.
Metzelder SK, Schroeder T, Lubbert M, Ditschkowski M, Gotze K, Scholl S, et al. Long-term survival of sorafenib-treated FLT3-ITD-positive acute myeloid leukaemia patients relapsing after allogeneic stem cell transplantation. Eur J Cancer. 2017;86:233-9.	Population receiving sorafenib as salvage therapy. Not the population of interest.
Pasvolsky O, Shimony S, Yeshurun M, Shargian L, Wolach O, Raanani P, et al. Maintenance therapy after allogeneic hematopoietic transplant for acute myeloid leukemia: a systematic review and meta-analysis. Acta Oncol. 2021;60(10):1335-41.	Meta analysis combines different populations and FLT3 inhibitors, including populations and interventions not in scope. Individual studies considered separately for eligibility for inclusion in this review.
Pollard JA, Alonzo TA, Gerbing R, Brown P, Fox E, Choi J, et al. Sorafenib in Combination With Standard Chemotherapy for Children With High Allelic Ratio FLT3/ITD+ Acute Myeloid Leukemia: A Report From the Children's Oncology Group Protocol AAML1031. Journal of Clinical Oncology. 2022;40(18):2023-35.	Not a study about the effectiveness of sorafenib after allo-HSCT.
Pratz KW, Rudek MA, Smith BD, Karp J, Gojo I, Dezern A, et al. A Prospective Study of Peritransplant Sorafenib for Patients with FLT3-ITD Acute Myeloid Leukemia Undergoing Allogeneic Transplantation. Biol Blood Marrow Transplant. 2020;26(2):300-6.	Non-comparative study. RCT evidence available for the outcomes reported in this study.
Rautenberg C, Nachtkamp K, Dienst A, Schmidt PV, Heyn C, Kondakci M, et al. Sorafenib and azacitidine as salvage therapy for relapse of FLT3-ITD mutated AML after allo-SCT. Eur J Haematol. 2017;98(4):348-54.	Population receiving sorafenib as salvage therapy. Not the population of interest.
Tarlock K, Chang B, Cooper T, Gross T, Gupta S, Neudorf S, et al. Sorafenib treatment following hematopoietic stem cell transplant in pediatric FLT3/ITD acute myeloid leukemia. Pediatr Blood Cancer. 2015;62(6):1048-54.	Non-comparative study. RCT evidence available for the outcomes reported in this study.
Tschan-Plessl A, Halter JP, Heim D, Medinger M, Passweg JR, Gerull S. Synergistic effect of sorafenib and cGvHD in patients with high-risk FLT3-ITD+AML allows long-term disease control after allogeneic transplantation. Ann Hematol. 2015;94(11):1899-905.	Non-comparative study. RCT evidence available for the outcomes reported in this study.
Xuan L, Wang Y, Chen J, Jiang E, Gao L, Wu B, et al. Sorafenib Therapy Is Associated with Improved Outcomes for FMS-like Tyrosine Kinase 3 Internal Tandem Duplication Acute Myeloid Leukemia Relapsing after Allogeneic Hematopoietic Stem Cell Transplantation. Biol Blood Marrow Transplant. 2019;25(8):1674-81.	Population receiving sorafenib as salvage therapy. Not the population of interest.
Xuan L, Wang Y, Huang F, Jiang E, Deng L, Wu B, et al. Effect of sorafenib on the outcomes of patients with FLT3-ITD acute myeloid leukemia undergoing allogeneic hematopoietic stem cell transplantation. Cancer. 2018;124(9):1954-63.	Non-randomised study. RCT evidence available for the outcomes reported in this study.

Appendix E Evidence Table

For abbreviations see list after table

Study details	Population	Intervention	Study outcomes	Appraisal and Funding
Burchert A, Bug G, Fritz LV, Finke J, Stelljes M, Röllig C et al. Sorafenib maintenance after allogenic hematopoietic stem cell transplantation for acute myeloid leukemia with FLT3-internal tandem duplication mutation (SORMAIN). Journal of Clinical Oncology 2020, 38(26):2993-3003. **Study location** 15 centres in Germany and Austria **Study type** RCT **Study aim** To test the hypothesis that sorafenib can inhibit FLT3-ITD AML recurrence after allo-HSCT	Adults with FLT3-ITD AML in complete haematologic remission after allo-HSCT **Inclusion criteria** Adults with FLT3-ITD AML in complete haematologic remission at enrolment after HSCT from a 9/10 or 10/10 HLA-matched unrelated or sibling donor HSCT could be performed as part of the consolidation therapy upfront (i.e. in the first remission) or in the context of relapsed or refractory AML Conditioning therapy for HSCT could be given with or without prior achievement of a complete remission using	**Intervention** Sorafenib 2 x 200mg orally per day for 2 weeks (dose level 1), then 3 x 200mg orally per day for 4 weeks (dose level 2), then 4 x 200mg orally per day (dose level 3) Treatment started between 60 and 100 days after allo-HSCT and continued for 24 months or until relapse or intolerable toxicity **Comparison** Placebo for up to 24 months **Concurrent treatments** Patients could be treated with TKIs (including sorafenib),	Median (IQR) follow-up: 41.8 months (24.1 to 42.5) **Critical outcomes** **Relapse free survival (RFS)[17]** *Relapse events* • Sorafenib: 10/43 (23.3%) (8 relapses and 2 deaths) • Placebo: 19/40 (47.5%) (17 relapses and 2 deaths) HR for relapse or death at median (IQR) follow-up of 41.8 months (24.1 to 42.5) for sorafenib vs placebo: 0.39 (95%CI 0.18 to 0.85), p=0.013 Median RFS was not reached for sorafenib and was 30.9 months for placebo (CI not reported) *Estimated probability of 24 month RFS:* • Sorafenib: 85.0% (95% CI 70 to 93) • Placebo: 53.3% (95%CI 36 to 68)	This study was appraised using the JBI checklist for RCTs: 1. Yes 2. Yes 3. Yes 4. Yes 5. Yes 6. Unclear 7. Yes 8. Yes 9. Yes 10. Yes 11. Yes 12. Yes 13. Yes **Other comments** This was a double-blind multi-centre phase II RCT comparing sorafenib to placebo. Groups were similar at baseline for potential prognostic factors. There were more males and more patients with an ECOG-PS of 0 in the placebo group.

[17] Calculated as time from randomisation to the first occurrence of either AML relapse or death from any cause. Relapse was defined as loss of complete haematologic remission, according to the revised recommendations of the International Working Group (Cheson et al 2003)

Study details	Population	Intervention	Study outcomes	Appraisal and Funding
Study dates October 2010 to May 2016	either a dose-reduced or a myeloablative protocol Patients could be treated with FLT3-targeting agents (except sorafenib) before study enrolment **Exclusion criteria** The exclusion criteria included previous sorafenib therapy, secondary HSCT, severe concomitant conditions and active GVHD at randomisation despite adequate treatment **Total sample size** n=83 Sorafenib: n=43 Placebo: n=40 **Baseline characteristics** *Sorafenib* Age median (range): 54.2 years (23.6-74.6) Male: 41.9% ECOG-PS score: • 0: 30.2% • 1: 67.4% • Missing: 2.3%	chemotherapy or a second allo-HSCT for the treatment of relapse after study entry	HR for relapse or death at 24 months for sorafenib vs placebo: 0.26 (95%CI 0.10 to 0.65), p=0.002 *Subgroup analysis* RFS was statistically significantly higher with sorafenib than placebo for the following subgroups: • Patients with undetectable minimal residual disease (MRD) before allo-HSCT (0/9 relapsed or died with sorafenib vs 5/12 with placebo, p=0.028) • Patients with detectable MRD after allo-HSCT (p=0.015) (n not reported) **Overall survival[18]** *Deaths* • Sorafenib: 11/43 (25.6%) • Placebo: 16/40 (40.0%) HR for death at median follow-up of 55.1 months for sorafenib vs placebo: 0.52 (95%CI 0.24 to 1.11), p=0.086 Follow-up IQR not reported Median overall survival was not reached for sorafenib or placebo *Estimated probability of 24 month overall survival:* • Sorafenib: 90.5% (95% CI 77 to 96) • Placebo: 66.2% (95%CI 49 to 79)	Investigators and patients were blinded to treatment group. No statement was made about whether outcome assessors were blind to treatment group. However, the outcomes reported were objective and/or had standardised definitions or assessment criteria. The risk of any potential lack of blinding for assessors impacting the results is low. Three sorafenib patients and one placebo patient withdrew consent (reason not stated). Two of these, one in each group, did not receive any study treatment. The authors conducted an intention-to-treat analysis. Patients were included in the analysis of safety outcomes, including GVHD, if they received at least one dose of study medication. The length of follow-up was sufficient for the outcomes reported. 13 patients in the placebo group and 5 patients in the sorafenib group received sorafenib as treatment after a relapse during

[18] Calculated as time from randomisation to death from any cause

Study details	Population	Intervention	Study outcomes	Appraisal and Funding
	Placebo Age median (range): 53.6 years (18.6-75.6) Male: 57.5% ECOG-PS score: • 0: 45.0% • 1: 55.0% The authors stated that the groups were well balanced in relation to potential prognostic factors such as cytogenetic and genetic risk category and time of transplantation with regards to whether they were in first complete remission at transplantation. The proportion of males and the proportion of patients with an ECOG-PS of 0 were higher in the placebo group		HR for death at 24 months for sorafenib vs placebo: 0.24 (95%CI 0.08 to 0.74), p=0.007 **Important outcomes** **Treatment adherence** *Median (range) duration of therapy:* • Sorafenib: 34.6 weeks (1.3 to 106.9) • Placebo: 54.4 weeks (1.7 to 128.3) *Dose reductions:* • Sorafenib: 21/43 (49%) • Placebo: 16/40 (40%) No statistical comparison between groups **Graft-versus-host-disease (GVHD)[19]** *Acute GVHD (Grade ≥2):* • Sorafenib: 10/42 (24%) • Placebo: 7/39 (18%) *Chronic GVHD (severe):* • Sorafenib: 8/42 (19%) • Placebo: 4/39 (10%) *Chronic GVHD (mild/moderate):* • Sorafenib: 18/42 (43%) • Placebo: 14/39 (36%) No statistical comparison between groups	the study. Other treatments after a relapse included six patients who received a second allo-HSCT (5 in the placebo group and 1 in the sorafenib group) and 17 patients who received chemotherapy (11 in the placebo group and 6 in the sorafenib group). The authors stated that there was no significant difference in the administration of relapse therapies between the groups. The power calculation determined that 200 patients were needed. However, study recruitment was terminated prematurely in 2016 due to inadequate slow patient recruitment. The study may not have been sufficiently powered. Statistical comparison between the groups was not reported for treatment adherence, GVHD or safety outcomes. No minimally clinically important differences were reported. However, the authors stated that "a relapse rate of only 15% after 2 years in the sorafenib arm

[19] Acute GVHD categorised according to the Mount Sinai Acute GVHD International Consortium (Harris et al 2016). Chronic GVHD categorised according to the National Institutes of Health consensus criteria (Filipovich et al 2005)

Study details	Population	Intervention	Study outcomes	Appraisal and Funding
			Safety No statistical comparison between groups for safety outcomes *Study drug discontinuation due to toxicity:* • Sorafenib: 9/43 (21%) • Placebo: 2/40 (5%) *Adverse events Grade ≥3[20]:* Proportion of patients in each group with any adverse event Grade ≥3 not reported • Sorafenib (n=42) • Infections: 11 (26%) • GI toxicity[21]: 6 (14%) • Electrolyte alterations: 6 (14%) • Skin toxicity: 5 (12%) • Cardiotoxicity and renal insufficiency: 4 (10%) • Thrombocytopaenia: 2 (5%) • Liver toxicity[22]: 2 (5%) • Neutropaenia: 1 (2%) • Other[23]: 33 (79%) • Placebo: (n=39) • Infections: 9 (23%) • GI toxicity: 6 (15%) • Liver toxicity: 2 (5%) • Skin toxicity: 1 (3%) • Neutropenia: 1 (3%) • Thrombocytopenia: 1 (3%)	appears to be a clinically meaningful improvement". Some patients received a second allo-HSCT after relapse. It is not clear if any patients received sorafenib after a second allo-HSCT. Exploratory subgroup analysis was conducted to explore subgroups showing the strongest benefit with sorafenib. The study was conducted in 2 European countries with recruitment over a 6 year period. The generalisability of the results to the NHS in England is unclear. **Source of funding** Bayer HealthCare provided study drugs and partial financial support. Other funding/ grants declared were from the Deutsche Forschungsgemeinschaft and the German Carreras Leukemia Foundation.

[20] Grading criteria not specified
[21] Vomiting, nausea, diarrhoea
[22] ALT, AST increased
[23] Not further defined

Study details	Population	Intervention	Study outcomes	Appraisal and Funding
			• Cardiotoxicity and renal insufficiency: 1 (3%) • Electrolyte alterations: 1 (3%) • Other: 22 (56%) *Drug-related adverse events Grade ≥3:* Proportion of patients in each group with any drug-related adverse event Grade ≥3 not reported • Sorafenib (n=42) • Electrolyte alterations: 3 (7%) • Skin toxicity: 2 (5%) • GI toxicity: 2 (5%) • Infections: 1 (2%) • Cardiotoxicity and renal insufficiency: 1 (2%) • Neutropaenia: 1 (2%) • Other: 8 (19%) • Placebo: (n=39) • GI toxicity: 3 (8%) • Infections: 2 (5%) • Liver toxicity: 2 (5%) • Skin toxicity: 1 (3%) • Neutropenia: 1 (3%) • Other: 4 (10%)	
Xu X, Fan Z, Wang Y, Huang F, Xu Y, Sun J. et al. Effect of sorafenib maintenance on Epstein-Barr virus and cytomegalovirus infections in patients	Adults with FLT3-ITD AML in composite complete remission[24] before and after allo-HSCT	This paper reports longer-term outcomes from an RCT. The intervention group received sorafenib. The comparator	Median (IQR) follow-up: 36.8 months (2.5 to 67.1) **Critical outcomes** **Relapse free survival (RFS)** See Xuan et al 2020 for outcome definition	This study was appraised using the JBI checklist for RCTs. See Xuan et al 2020 for ratings and comments relating to the design and conduct of this RCT. **Other comments**

[24] Composite complete remission was complete remission, complete remission with incomplete platelet recovery and complete remission with incomplete haematological recovery

Study details	Population	Intervention	Study outcomes	Appraisal and Funding
with FLT3-ITD AML undergoing allogeneic haematopoietic stem cell transplantation: a secondary analysis of a randomized clinical trial. BMC Medicine 2022, 20: 282 **Study location** 7 centres in China **Study type** RCT follow-up study **Study aim** Longer-term follow-up of an RCT assessing the efficacy and tolerability of sorafenib as maintenance therapy post-transplantation for the prevention of relapse in patients with FLT3-ITD AML undergoing allo-HSCT **Study dates** June 2015 to July 2018	This paper reports outcomes at 3 years after transplantation from an RCT. See Xuan et al 2020 for the trial inclusion/ exclusion criteria and baseline characteristics **Total sample size** n=202 Sorafenib: n=100 No maintenance therapy: n=102	group received no maintenance therapy See Xuan et al 2020 for further details	*3 year RFS:* • Sorafenib: 75.9% (95% CI 66.2 to 83.1) • No maintenance therapy: 52.5% (95%CI 42.2 to 61.7) HR 0.41 (95%CI 0.25 to 0.67), p<0.001 *3-year cumulative incidence of relapse:* • Sorafenib: 13.0% (95% CI 7.3 to 20.4) • No maintenance therapy: 34.8% (95%CI 25.5 to 44.2) HR 0.31 (95%CI 0.16 to 0.58), p<0.001 **Overall survival** See Xuan et al 2020 for outcome definition *3 year overall survival:* • Sorafenib: 79.0% (95% CI 69.6 to 85.8) • No maintenance therapy: 61.4% (95%CI 51.1 to 70.1) HR 0.48 (95%CI 0.28 to 0.82), p=0.005 *Deaths during follow-up:* • Sorafenib: 21/100 (21.0%) • No maintenance therapy: 39/102 (38.2%) No statistical comparison between groups	This paper reports outcomes at 3 years after transplantation from an RCT. The primary focus of this analysis was on the effect of sorafenib maintenance on Epstein-Barr virus and cytomegalovirus infection risk. Results relating to this outcome were not extracted. The authors conducted an intention-to-treat analysis for the 3 year outcomes reported. **Source of funding** The authors stated that this study was supported by the National Natural Science Foundation of China, the National Key Research and Development Program of China, the Research and Development Program in Key areas of Guangdong Province and the Clinical Research Program of Nanfang Hospital Southern Medical University. The authors declared that they had no competing interests.

Study details	Population	Intervention	Study outcomes	Appraisal and Funding
Xuan L, Wang Y, Huang F, Fan Z, Xu Y, Sun J et al. Sorafenib maintenance in patients with FLT3-ITD acute myeloid leukaemia undergoing allogenic haematopoietic stem cell transplantation: an open-label, multicentre, randomised phase 3 trial. The Lancet 2020, 21(1):1201-1212 **Study location** 7 centres in China **Study type** RCT **Study aim** To assess the efficacy and tolerability of sorafenib as maintenance therapy post-transplantation for prevention of relapse in patients with FLT3-ITD AML undergoing allo-HSCT	Adults with FLT3-ITD AML in composite complete remission[25] before and after allo-HSCT **Inclusion criteria** Adults (aged 18-60) with FLT3-ITD AML undergoing first allo-HSCT who had an ECOG-PS score of 0-2. Patients had to have composite complete remission before and after allo-HSCT and haematopoietic recovery within 60 days post-transplantation An HLA-matched sibling donor was preferred for the allo-HSCT, followed by an HLA-matched unrelated donor. Patients could also have received a transplant from an HLA-haploidentical donor	**Intervention** Sorafenib 2 x 400mg orally per day Dose reductions or interruptions were allowed if adverse events of Grade ≥3 occurred. Dose reductions were to 200mg once or twice a day with return to 400mg twice a day after the resolution of adverse events Treatment started between 30 and 60 days after allo-HSCT and continued up to 180 days post-transplantation Sorafenib maintenance was withdrawn if an alternative FLT3 inhibitor was started post-transplantation, if any intolerable	Median (IQR) follow-up: 21.3 months (15.0 to 37.0) **Critical outcomes** **Relapse free survival (RFS)[26]** *2 year RFS:* • Sorafenib: 78.9% (95% CI 69.0 to 85.9) • No maintenance therapy: 56.6% (95%CI 46.1 to 65.8) HR 0.37 (95%CI 0.22 to 0.63), p<0.0001 *Number of patients who relapsed:* • Sorafenib: 11/100 (11.0%) • No maintenance therapy: 32/102 (31.4%) No statistical comparison between groups Median RFS was not reached for either the sorafenib or no maintenance therapy groups *1-year cumulative incidence of relapse:* • Sorafenib: 7.0% (95% CI 3.1 to 13.1) • No maintenance therapy: 24.5% (95%CI 16.6 to 33.2) HR 0.25 (95%CI 0.11 to 0.57), p=0.001 *2-year cumulative incidence of relapse:* • Sorafenib: 11.9% (95% CI 6.2 to 19.6)	This study was appraised using the JBI checklist for RCTs: 1. Yes 2. Yes 3. Yes 4. No 5. No 6. Yes 7. Yes 8. Yes 9. Yes 10. Yes 11. Yes 12. Yes 13. Yes **Other comments** This was an open label phase III RCT comparing sorafenib to no maintenance therapy. Groups were similar at baseline for prognostic and demographic factors. Investigators and participants were not blind to treatment group. However, the staff who did the data analysis and

[25] Composite complete remission was complete remission, complete remission with incomplete platelet recovery or complete remission with incomplete haematological recovery

[26] Calculated as time from transplantation until relapse or death from any cause. Relapse was defined as either reappearance of leukaemic blasts in the peripheral blood or at least 5% blasts in the bone marrow aspirate or biopsy specimen not attributable to any other causes, or reappearance or new appearance of extramedullary leukaemia

Study details	Population	Intervention	Study outcomes	Appraisal and Funding
Study dates June 2015 to July 2018	All patients received myeloablative conditioning with a modified busulfan-cyclophosphamide regimen Patients could receive sorafenib before transplantation **Exclusion criteria** Patients were excluded if they had acute promyelocytic leukaemia, intolerance to sorafenib pretransplantation, life expectancy shorter than 30 days post-transplantation, active acute GVHD or uncontrolled infections within 60 days post-transplantation, liver dysfunction or renal dysfunction or severe concomitant conditions not suitable for the trial such as cardiovascular disease or psychiatric disorders **Total sample size** n=202	adverse events related to study treatment arose, if the patient withdrew informed consent or if any clinical adverse event or laboratory test result indicated that study treatment was not in the patient's best interest **Comparison** No maintenance therapy with sorafenib or another FLT3 inhibitor **Concurrent treatments** GVHD and infection prophylaxis were permitted Patients could be treated with TKIs (including sorafenib), chemotherapy or donor lymphocyte infusion after relapse	• No maintenance therapy: 31.6% (95%CI 22.6 to 41.1) HR 0.29 (95%CI 0.15 to 0.58), p<0.0001 *Subgroup analysis* 2-year cumulative incidence of relapse was reported separately for age groups: *For patients aged <35 years (n=103):* • Sorafenib: 16.1% (95%CI 6.8 to 28.9) • No maintenance therapy: 25.1% (95%CI 14.2 to 37.7) HR 0.45 (95%CI 0.18 to 1.11), p not reported *For patients aged* ≥35 years (n=99)*:* • Sorafenib: 8.0% (95%CI 2.5 to 17.7) • No maintenance therapy: 38.7% (95%CI 24.4 to 52.7) HR 0.17 (95%CI 0.06 to 0.50), p not reported **Overall survival[27]** *2 year overall survival:* • Sorafenib: 82.1% (95% CI 72.6 to 88.5) • No maintenance therapy: 68.0% (95%CI 57.8 to 76.2) HR 0.48 (95%CI 0.27 to 0.86), p=0.012 *Deaths during follow-up:* • Sorafenib: 17/100 (17.0%) • No maintenance therapy: 32/102 (31.4%) No statistical comparison between groups	assessment of outcomes were blinded. The outcomes reported were objective and/or had standardised definitions or assessment criteria. The risk of any lack of blinding impacting the results is low. One patient from the no maintenance therapy group withdrew consent after randomisation (reason not stated). The authors conducted an intention-to-treat analysis for both the efficacy and safety outcomes. The analysis of chronic GVHD only included the 198 patients who were alive 100 days after allo-HSCT. The length of follow-up was sufficient for the outcomes reported. 27 patients in the no maintenance therapy group and 9 patients in the sorafenib group received salvage therapy for relapse during the study. All patients received sorafenib as salvage therapy, with or without chemotherapy and donor lymphocyte infusion.

[27] Calculated as time from transplantation to death from any cause

Study details	Population	Intervention	Study outcomes	Appraisal and Funding
	Sorafenib: n=100 No maintenance therapy: n=102 **Baseline characteristics** *Sorafenib* Age median (IQR): 35 years (26-42) Male: 50% Received sorafenib pre-transplantation: 24% *No maintenance therapy* Age median (IQR): 35 years (26-43) Male: 51% Received sorafenib pre-transplantation: 25% The authors stated that prognostic factors were well balanced between groups. The groups were also well balanced for age, gender and sorafenib use pre-transplantation		Median overall survival was not reached for either the sorafenib or no maintenance therapy groups **Important outcomes** **Treatment adherence** *Median (IQR) duration of therapy:* • Sorafenib: 134 days (116 to 150) *Sorafenib dose amendments due to adverse events:* • Dose reductions: 42/100 (42%) • Dose interruptions: 12/100 (12%) **Graft-versus-host-disease (GVHD)[28]** Assessed up to 210 days post-transplantation. No statistical comparison between groups for GVHD outcomes *Acute GVHD (Grade ≥2):* • Sorafenib: 23/100 (23%) • No maintenance therapy: 21/102 (21%) *Acute GVHD (Grade I)* • Sorafenib: 8/100 (8%) • No maintenance therapy: 6/102 (6%) *Chronic GVHD (moderate/severe):* • Sorafenib: 18/99 (18%) • No maintenance therapy: 17/99 (17%)	The number of patients included in the analysis was sufficient to show a difference between treatment groups with a two-sided significance level of 5% and 90% power. Statistical comparison between the groups was not reported for treatment adherence, GVHD or safety outcomes. No statement on minimally clinically important differences was made. The authors reported that post-hoc exploratory analysis of two year cumulative incidence of relapse favoured sorafenib over maintenance theory in almost all subgroups. These included subgroups considering gender, white blood cell count at diagnosis, NPM1 mutation, sorafenib pre-transplantation, disease status at transplantation, complete remission (CR) status at transplantation, MRD at transplantation, MRD at the time of enrolment post-transplantation and chronic GVHD. Patient age group did show a difference in results for subgroups (results

[28] Acute and chronic GVHD were graded according to published guidelines (Prezpiorka et al 1995, Jagasia et al 2015)

Study details	Population	Intervention	Study outcomes	Appraisal and Funding
			Chronic GVHD (mild): • Sorafenib: 5/99 (5%) • No maintenance therapy: 5/99 (5%) **Safety** Adverse events[29] were recorded up to 210 days post-transplantation. No statistical comparison between groups for safety outcomes *Study drug discontinuation due to adverse events:* • Sorafenib: 5/100 (5%) Reasons for discontinuation were death (n=4) and a treatment-related adverse event (n=1) *Patients with ≥1 adverse events of Grade 3-4:* • Sorafenib: 50/100 (50%) • No maintenance therapy: 47/102 (46%) Proportion of patients in each group with any drug-related adverse events not reported *Deaths due to adverse events:* • Sorafenib: 4/100 (4%) • No maintenance therapy: 5/102 (5%) No patients died from treatment-related adverse events. Causes of deaths in the	extracted). The only other subgroups that showed a difference in results for subgroups were cytogenetic risk, transplant modality and acute GVHD. These data were not extracted as they were not specified as subgroups of interest in the PICO. The authors conducted post-hoc multivariable analysis of risk factors for survival outcomes. This analysis confirmed that sorafenib maintenance post-transplantation was the only protective factor for relapse free survival and overall survival. Other variables included in the multivariable analysis were patient's age (<35 years vs ≥35 years), CR status at transplantation (≥CR2 vs CR1), MRD at transplantation (positive vs negative), sorafenib pre-transplantation (use vs no use), MRD at time of enrolment post-transplantation (positive vs negative) and chronic GVHD (yes or no).

[29] Non-haematological adverse events defined using the National Cancer Institute Common Terminology Criteria for Adverse Events (version 4.0) where Grade 1 = mild; Grade 2 = moderate; Grade 3 = severe or medically significant but not immediately life threatening; Grade 4 = life-threatening consequences; Grade 5 = death related to adverse event. Grade 3 haematological adverse events were defined as either an absolute neutrophil count <1.0 x 10^9 cells/L but ≥ 0.5 x 10^9 cells/L or a platelet count <30 x 10^9/L but ≥20 x 10^9/L. Grade 4 haematological adverse events were defined as either an absolute neutrophil count <0.5 x 10^9 cells/L or a platelet count <20 x 10^9/L

Study details	Population	Intervention	Study outcomes	Appraisal and Funding
			sorafenib group were infections (n=2), acute GVHD (n=1) and cardiotoxicity (n=1). Causes of deaths in the no maintenance therapy group were infections (n=3), acute GVHD (n=1) and thrombotic microangiopathy (n=1) *Adverse events Grade 3-4:* • Sorafenib (n=100) • Infections: 25 (25%) • Haematologic toxicity: 15 (15%) • Gastrointestinal: 11 (11%) • Skin related: 7 (7%) • Hepatobiliary or pancreatic: 5 (5%) • Renal or genitourinary: 4 (4%) • Secondary malignant disease: 2 (2%) • Vascular: 1 (1%) • No maintenance therapy: (n=102) • Infections: 24 (24%) • Gastrointestinal: 8 (8%) • Haematologic toxicity: 7 (7%) • Hepatobiliary or pancreatic: 6 (6%) • Renal or genitourinary: 5 (5%) • Secondary malignant disease: 2 (2%) • Skin related: 1 (1%) • Cardiac: 1 (1%) • Vascular: 1 (1%) The authors stated that the most common Grade 3-4 treatment-related adverse events with sorafenib were skin-related (7%) or haematological (5%)	Subgroup analyses based on whether or not patients had sorafenib pre-transplantation and/or post-transplantation were only presented graphically in a data supplement. The study was conducted in 7 centres in China with recruitment over a 3 year period. The generalisability of the results to the NHS in England is unclear. **Source of funding** The authors stated that there was no funding source for this study. The authors declared that they had no competing interests.

Study details	Population	Intervention	Study outcomes	Appraisal and Funding
			Adverse events Grade 1-2: • Sorafenib (n=100) • Gastrointestinal: 25 (25%) • Renal or genitourinary: 23 (23%) • Skin related: 20 (20%) • Hepatobiliary or pancreatic: 16 (16%) • Cardiac: 14 (14%) • Infections: 8 (8%) • Vascular: 6 (6%) • No maintenance therapy: (n=102) • Renal or genitourinary: 25 (25%) • Gastrointestinal: 20 (20%) • Hepatobiliary or pancreatic: 17 (17%) • Cardiac: 12 (12%) • Infections: 9 (9%) • Skin related: 9 (9%) • Vascular: 5 (5%)	

Abbreviations

Allo-HSCT: allogeneic haematopoietic stem cell transplantation; ALT: alanine aminotransferase; AML: acute myeloid leukaemia; AST: aspartate aminotransferase; CI: confidence intervals; CR: complete remission; ECOG-PS: Eastern Cooperative Oncology Group performance status; FLT3-ITD: fms-like tyrosine kinase 3-internal tandem duplication; GI: gastrointestinal; GVHD: graft-versus-host-disease; HLA: human leukocyte antigen; HR: hazard ratio; HSCT: haematopoietic stem cell transplantation; IQR: interquartile range; L: litre; mg: milligram; MRD: minimal residual disease; RCT: randomised controlled trial; RFS: relapse free survival; TKI: tyrosine kinase inhibitor

Appendix F Quality appraisal checklists

JBI Critical Appraisal Checklist for RCTs

1. Was true randomisation used for assignment of participants to treatment groups?
2. Was allocation to treatment groups concealed?
3. Were treatment groups similar at the baseline?
4. Were participants blinded to treatment assignment?
5. Were those delivering treatment blind to treatment assignment?
6. Were outcomes assessors blind to treatment assignment?
7. Were treatment groups treated identically other than the intervention of interest?
8. Was follow-up complete and if not, were differences between groups in terms of their follow-up adequately described and analysed?
9. Were participants analysed in the groups to which they were randomised?
10. Were outcomes measured in the same way for treatment groups?
11. Were outcomes measured in a reliable way?
12. Was appropriate statistical analysis used?
13. Was the trial design appropriate, and any deviations from the standard RCT design (individual randomisations, parallel groups) accounted for in the conduct and analysis of the trial

Appendix G GRADE profiles

In patients who have FLT3-ITD AML and have undergone allo-HSCT, what is the clinical effectiveness and safety of sorafenib compared with standard of care?

For abbreviations and footnotes see end of tables.

Table 2. Sorafenib compared to placebo

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sorafenib	Placebo	Result		
Relapse free survival (RFS) (1 RCT)									
Relapse events (number, %). Median (IQR) follow-up 41.8 months (24.1 to 42.5)									
1 RCT Burchert et al 2020	No serious limitations	No serious indirectness	Not applicable	Serious imprecision[1]	10/43 (23.3%)	19/40 (47.5%)	HR for relapse or death at median follow-up: 0.39 (95%CI 0.18 to 0.85), p=0.013 Median RFS not reached for sorafenib; 30.9 months for placebo (CI not reported)	Critical	Moderate
RFS at 24 months (%)									
1 RCT Burchert et al 2020	No serious limitations	No serious indirectness	Not applicable	No serious imprecision	43	40	• Sorafenib: 85.0% (95% CI 70 to 93) • Placebo: 53.3% (95%CI 36 to 68) HR for relapse or death at 24 months: 0.26 (95%CI 0.10 to 0.65), p=0.002	Critical	High
Overall survival (1 RCT)									
Deaths (number, %). Median follow-up 55.1 months (IQR not reported)									
1 RCT	No serious limitations	No serious indirectness	Not applicable	Serious imprecision[1]	11/43 (25.6%)	16/40 (40.0%)	HR for death at median follow-up: 0.52 (95%CI 0.24 to 1.11), p=0.086	Critical	Moderate

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	**Risk of bias**	**Indirectness**	**Inconsistency**	**Imprecision**	**Sorafenib**	**Placebo**	**Result**		
Burchert et al 2020							Median overall survival not reached for sorafenib or placebo		
Overall survival at 24 months (%)									
1 RCT Burchert et al 2020	No serious limitations	No serious indirectness	Not applicable	No serious imprecision	43	40	• Sorafenib: 90.5% (95% CI 77 to 96) • Placebo: 66.2% (95%CI 49 to 79) HR for death at 24 months: 0.24 (95%CI 0.08 to 0.74), p=0.007	Critical	High
Treatment adherence (1 RCT)									
Median (range) duration of therapy. Median (IQR) follow-up 41.8 months (24.1 to 42.5)									
1 RCT Burchert et al 2020	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	43	40	• Sorafenib: 34.6 weeks (1.3 to 106.9) • Placebo: 54.4 weeks (1.7 to 128.3) No statistical comparison between groups	Important	Moderate
Dose reductions (number, %). Median (IQR) follow-up 41.8 months (24.1 to 42.5)									
1 RCT Burchert et al 2020	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	21/43 (49%)	16/40 (40%)	No statistical comparison between groups	Important	Moderate
Graft-versus-host-disease (GVHD) (1 RCT)									
Acute GVHD (Grade ≥2) (number, %). Median (IQR) follow-up 41.8 months (24.1 to 42.5)									
1 RCT Burchert et al 2020	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	10/42 (24%)	7/39 (18%)	No statistical comparison between groups	Important	Moderate
Chronic GVHD (severe) (number, %). Median (IQR) follow-up 41.8 months (24.1 to 42.5)									
1 RCT	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	8/42 (19%)	4/39 (10%)	No statistical comparison between groups	Important	Moderate

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sorafenib	Placebo	Result		
Burchert et al 2020									
Chronic GVHD (mild/moderate) (number, %). Median (IQR) follow-up 41.8 months (24.1 to 42.5)									
1 RCT Burchert et al 2020	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	18/42 (43%)	14/39 (36%)	No statistical comparison between groups	Important	Moderate
Safety (1 RCT)									
Study drug discontinuation due to toxicity (number, %). Median (IQR) follow-up 41.8 months (24.1 to 42.5)									
1 RCT Burchert et al 2020	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	9/42 (21%)	2/39 (5%)	No statistical comparison between groups	Important	Moderate
Adverse events Grade ≥3 (%). Median (IQR) follow-up 41.8 months (24.1 to 42.5)									
1 RCT Burchert et al 2020	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	42	39	Proportion of patients in each group with any adverse event Grade ≥3 not reported Most common (>10% of patients) adverse events with sorafenib: infections (26%), GI toxicity (14%), electrolyte alterations (14%), skin toxicity (12%), cardiotoxicity and renal insufficiency (10%), other (not further defined) (79%) Most common (>10% of patients) adverse events with placebo: infections (23%), GI toxicity (15%), other (not further defined) (56%)	Important	Moderate
Drug-related adverse events Grade ≥3 (%). Median (IQR) follow-up 41.8 months (24.1 to 42.5)									
1 RCT	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	42	39	Proportion of patients in each group with any drug-related	Important	Moderate

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sorafenib	Placebo	Result		
Burchert et al 2020							adverse event Grade ≥3 not reported Most common (>5% of patients) drug-related adverse events with sorafenib: electrolyte alterations (7%), skin toxicity (5%), GI toxicity (5%), other (not further defined) (19%) Most common (>5% of patients) drug-related adverse events with placebo: GI toxicity (8%), infections (5%), liver toxicity (5%), other (not further defined) 10%		

Abbreviations
CI: confidence intervals; GI: gastrointestinal; GVHD: graft-versus-host-disease; HR: hazard ratio; IQR: interquartile range; RCT: randomised controlled trial; RFS: relapse free survival

1. Imprecision: Serious imprecision due to wide 95% confidence intervals that cross the default minimal clinically important difference lower threshold
2. Risk of bias: Serious limitations due to lack of statistical analysis

Table 3. Sorafenib compared to no maintenance therapy

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sorafenib	No maintenance therapy	Result		
Relapse free survival (RFS) (1 RCT)									
RFS at 3 years (%)									
1 RCT Xu et al 2022	No serious limitations	No serious indirectness	Not applicable	No serious imprecision	100	102	• Sorafenib: 75.9% (95% CI 66.2 to 83.1) • No maintenance therapy: 52.5% (95%CI 42.2 to 61.7) HR: 0.41 (95%CI 0.25 to 0.67), p<0.001	Critical	High
RFS at 2 years (%)									
1 RCT Xuan et al 2020	No serious limitations	No serious indirectness	Not applicable	No serious imprecision	100	102	• Sorafenib: 78.9% (95% CI 69.0 to 85.9) • No maintenance therapy: 56.6% (95%CI 46.1 to 65.8) HR: 0.37 (95%CI 0.22 to 0.63), p<0.0001	Critical	High
Relapses (number, %). Median (IQR) follow-up 21.3 months (15.0 to 37.0)									
1 RCT Xuan et al 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	11/100 (11.0%)	32/102 (31.4%)	No statistical comparison between groups Median RFS not reached for sorafenib or no maintenance therapy	Critical	Moderate
Cumulative incidence of relapse at 3 years (%)									
1 RCT Xu et al 2022	No serious limitations	No serious indirectness	Not applicable	No serious imprecision	100	102	• Sorafenib: 13.0% (95% CI 7.3 to 20.4) • No maintenance therapy: 34.8% (95%CI 25.5 to 44.2) HR: 0.31 (95%CI 0.16 to 0.58), p<0.001	Critical	High

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sorafenib	No maintenance therapy	Result		
Cumulative incidence of relapse at 2 years (%)									
1 RCT Xuan et al 2020	No serious limitations	No serious indirectness	Not applicable	No serious imprecision	100	102	• Sorafenib: 11.9% (95% CI 6.2 to 19.6) • No maintenance therapy: 31.6% (95%CI 22.6 to 41.1) HR: 0.29 (95%CI 0.15 to 0.58), p<0.0001	Critical	High
Cumulative incidence of relapse at 1 year (%)									
1 RCT Xuan et al 2020	No serious limitations	No serious indirectness	Not applicable	No serious imprecision	100	102	• Sorafenib: 7.0% (95% CI 3.1 to 13.1) • No maintenance therapy: 24.5% (95%CI 16.6 to 33.2) HR: 0.25 (95%CI 0.11 to 0.57), p=0.001	Critical	High
Overall survival (1 RCT)									
Overall survival at 3 years (%)									
1 RCT Xu et al 2022	No serious limitations	No serious indirectness	Not applicable	Serious imprecision[2]	100	102	• Sorafenib: 79.0% (95% CI 69.6 to 85.8) • No maintenance therapy: 61.4% (95%CI 51.1 to 70.1) HR: 0.48 (95%CI 0.28 to 0.82), p=0.005	Critical	Moderate
Overall survival at 2 years (%)									
1 RCT Xuan et al 2020	No serious limitations	No serious indirectness	Not applicable	Serious imprecision[2]	100	102	• Sorafenib: 82.1% (95% CI 72.6 to 88.5) • No maintenance therapy: 68.0% (95%CI 57.8 to 76.2) HR: 0.48 (95%CI 0.27 to 0.86), p=0.012	Critical	Moderate

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sorafenib	No maintenance therapy	Result		
Deaths (number, %). Median (IQR) follow-up 36.8 months (2.5 to 67.1)									
1 RCT Xu et al 2022	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	21/100 (21.0%)	39/102 (38.2%)	No statistical comparison between groups	Critical	Moderate
Deaths (number, %). Median (IQR) follow-up 21.3 months (15.0 to 37.0)									
1 RCT Xuan et al 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	17/100 (17.0%)	32/102 (31.4%)	No statistical comparison between groups Median overall survival not reached for sorafenib or no maintenance therapy	Critical	Moderate
Treatment adherence (1 RCT)									
Median (IQR) duration of therapy. Median (IQR) follow-up 21.3 months (15.0 to 37.0)									
1 RCT Xuan et al 2020	No serious limitations	Serious indirectness[3]	Not applicable	Not calculable	100	N/a	• Sorafenib: 134 days (116 to 150)	Important	Moderate
Dose reductions due to adverse events (%). Median (IQR) follow-up 21.3 months (15.0 to 37.0)									
1 RCT Xuan et al 2020	No serious limitations	Serious indirectness[3]	Not applicable	Not calculable	42/100 (42%)	N/a	Sorafenib dose reductions: 42%	Important	Moderate
Dose interruptions due to adverse events (%). Median (IQR) follow-up 21.3 months (15.0 to 37.0)									
1 RCT Xuan et al 2020	No serious limitations	Serious indirectness[3]	Not applicable	Not calculable	12/100 (12%)	N/a	Sorafenib dose interruptions: 12%	Important	Moderate
Graft-versus-host-disease (GVHD) (1 RCT)									
Acute GVHD (Grade ≥2) (number, %). Assessed up to 210 days post-transplantation									
1 RCT	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	23/100 (23%)	21/102 (21%)	No statistical comparison between groups	Important	Moderate

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sorafenib	No maintenance therapy	Result		
Xuan et al 2020									
Acute GVHD (Grade 1) (number, %). Assessed up to 210 days post-transplantation									
1 RCT Xuan et al 2020	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	8/100 (8%)	6/102 (6%)	No statistical comparison between groups	Important	Moderate
Chronic GVHD (moderate/severe) (number, %). Assessed up to 210 days post-transplantation									
1 RCT Xuan et al 2020	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	18/99 (18%)	17/99 (17%)	No statistical comparison between groups	Important	Moderate
Chronic GVHD (mild) (number, %). Assessed up to 210 days post-transplantation									
1 RCT Xuan et al 2020	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	5/99 (5%)	5/99 (5%)	No statistical comparison between groups	Important	Moderate
Safety (1 RCT)									
Study drug discontinuation due to adverse events (number, %). Assessed up to 210 days post-transplantation									
1 RCT Xuan et al 2020	No serious limitations	Serious indirectness[3]	Not applicable	Not calculable	5/100 (5%)	N/a	Reasons for discontinuation: death (n=4); treatment-related adverse event (n=1)	Important	Moderate
Deaths due to adverse events (number,%). Assessed up to 210 days post-transplantation									
1 RCT Xuan et al 2020	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	4/100 (4%)	5/102 (5%)	No statistical comparison between groups	Important	Moderate
Patents with ≥1 adverse events Grade 3-4 (number,%). Assessed up to 210 days post-transplantation									
1 RCT	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	50/100 (50%)	47/102 (46%)	No statistical comparison between groups	Important	Moderate

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	**Risk of bias**	**Indirectness**	**Inconsistency**	**Imprecision**	**Sorafenib**	**No maintenance therapy**	**Result**		
Xuan et al 2020									
Adverse events Grade 3-4 (%). Assessed up to 210 days post-transplantation									
1 RCT Xuan et al 2020	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	100	102	Most common (>10% of patients) adverse events with sorafenib: infections (25%), haematologic toxicity (15%), gastrointestinal (11%) Most common (>10% of patients) adverse event with no maintenance therapy: infections (24%)	Important	Moderate
Adverse events Grade 1-2 (%). Assessed up to 210 days post-transplantation									
1 RCT Xuan et al 2020	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	100	102	Most common (>10% of patients) adverse events with sorafenib: gastrointestinal (25%), renal or genitourinary (23%), skin related (20%), hepatobiliary or pancreatic (16%), cardiac (14%) Most common (>10% of patients) adverse events with no maintenance therapy: renal or genitourinary (25%), gastrointestinal (20%), hepatobiliary or pancreatic (17%), cardiac (12%)	Important	Moderate
Treatment-related adverse events (%). Assessed up to 210 days post-transplantation									
1 RCT Xuan et al 2020	No serious limitations	Serious indirectness[3]	Not applicable	Not calculable	100	N/a	The most common Grade 3-4 treatment-related adverse events with sorafenib were skin-related (7%) or haematological (5%). No patients died from treatment-related adverse events	Important	Moderate

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sorafenib	No maintenance therapy	Result		
							Proportion of patients in each group with any drug-related adverse events not reported		

Abbreviations
CI: confidence intervals; GVHD: graft-versus-host-disease; HR: hazard ratio; IQR: interquartile range; RCT: randomised controlled trial; RFS: relapse free survival

1. Risk of bias: Serious limitations due to lack of statistical analysis
2. Imprecision: Serious imprecision due to wide 95% confidence intervals that cross the default minimal clinically important difference lower threshold
3. Indirectness: Serious indirectness due to no comparison across treatment arms

Glossary

Adverse event	Any undesirable event experienced by a person while they are having a drug or any other treatment or intervention, regardless of whether or not the event is suspected to be related to or caused by the drug, treatment or intervention.
Bias	Systematic (as opposed to random) deviation of the results of a study from the 'true' results, which is caused by the way the study is designed or conducted.
Blinding	A way to prevent researchers, doctors and patients in a clinical trial from knowing which study group each patient is in so they cannot influence the results. The best way to do this is by sorting patients into study groups randomly. The purpose of 'blinding' or 'masking' is to protect against bias.
Clinical importance	A benefit from treatment that relates to an important outcome such as length of life and is large enough to be important to patients and health professionals.
Confidence interval (CI)	A way of expressing how certain we are about the findings from a study, using statistics. It gives a range of results that is likely to include the 'true' value for the population. A wide confidence interval indicates a lack of certainty about the true effect of the test or treatment - often because a small group of patients has been studied. A narrow confidence interval indicates a more precise estimate (for example, if a large number of patients have been studied).
Control group	A group of people in a study who do not have the intervention or test being studied. Instead, they may have the standard intervention. The results for the control group are compared with those for a group having the intervention being tested. The aim is to check for any differences. Ideally, the people in the control group should be as similar as possible to those in the intervention group, to make it as easy as possible to detect any effects due to the intervention.
GRADE (Grading of recommendations assessment, development and evaluation)	A systematic and explicit approach to grading the quality of evidence and the strength of recommendations developed by the GRADE working group.
Hazard ratio (HR)	The hazard or chance of an event occurring in the treatment arm of a study as a ratio of the chance of an event occurring in the control arm over time.
Intention-to-treat analysis (ITT)	An assessment of the people taking part in a trial, based on the group they were initially (and randomly) allocated to. This is regardless of whether or not they dropped out, fully adhered to the treatment or switched to an alternative treatment. ITT analyses are often used to assess clinical effectiveness because they mirror actual practice, when not everyone adheres to the treatment, and the treatment people have may be changed according to how their condition responds to it. Studies of drug treatments often use a modified ITT analysis, which includes only the people who have taken at least one dose of a study drug.
Minimal clinically important difference	The smallest change in a treatment outcome that people with the condition would identify as important (either beneficial or harmful), and that would lead a person or their clinician to consider a change in treatment.
Objective measure	A measurement that follows a standardised procedure which is less open to subjective interpretation by potentially biased observers and people in the study.
PICO (population, intervention, comparison and outcome) framework	A structured approach for developing review questions that divides each question into 4 components: the population (the population being studied); the interventions (what is being done); the comparators (other main treatment options); and the outcomes (measures of how effective the interventions have been).
P-value (p)	The p value is a statistical measure that indicates whether or not an effect is statistically significant. For example, if a study comparing 2 treatments found that 1 seems to be more effective than the other, the p value is the probability of obtaining these results by chance. By convention, if the p value is below 0.05 (that is, there is less than a 5% probability that the results occurred by chance), it is considered that there probably is a real difference between treatments. If the p value is 0.001 or less (less than a 0.1% probability that the results occurred by chance), the result is seen as highly significant. If the p value shows that there is

	likely to be a difference between treatments, the confidence interval describes how big the difference in effect might be.
Randomised controlled trial (RCT)	A study in which a number of similar people are randomly assigned to 2 (or more) groups to test a specific drug, treatment or other intervention. One group (the experimental group) has the intervention being tested, the other (the comparison or control group) has an alternative intervention, a dummy intervention (placebo) or no intervention at all. The groups are followed up to see how effective the experimental intervention was. Outcomes are measured at specific times and any difference in response between the groups is assessed statistically. This method is also used to reduce bias.
Statistical significance	A statistically significant result is one that is assessed as being due to a true effect rather than random chance.

References

Included studies

- Burchert A, Bug G, Fritz LV, Finke J, Stelljes M, Röllig C et al. Sorafenib maintenance after allogenic hematopoietic stem cell transplantation for acute myeloid leukemia with FLT3-internal tandem duplication mutation (SORMAIN). Journal of Clinical Oncology 2020, 38(26):2993-3003.
- Xu X, Fan Z, Wang Y, Huang F, Xu Y, Sun J. et al. Effect of sorafenib maintenance on Epstein-Barr virus and cytomegalovirus infections in patients with FLT3-ITD AML undergoing allogeneic haematopoietic stem cell transplantation: a secondary analysis of a randomized clinical trial. BMC Medicine 2022, 20: 282.
- Xuan L, Wang Y, Huang F, Fan Z, Xu Y, Sun J et al. Sorafenib maintenance in patients with FLT3-ITD acute myeloid leukaemia undergoing allogenic haematopoietic stem cell transplantation: an open-label, multicentre, randomised phase 3 trial. The Lancet 2020, 21(1):1201-1212.

Other references

- Cheson BD, Bennett JM, Kopecky KJ, Büchner T, Wilman CL, Estey EH et al. Revised recommendations of the International Working Group for diagnosis, standardization of response criteria, treatment outcomes and reporting standards for therapeutic trials in acute myeloid leukemia. Journal of Clinical Oncology 2003, 21:4642-4649.
- Filipovich AH, Weisdorf D, Pavletic S, Socie G, Wingard JR, Lee SJ et al. National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host-disease: I. Diagnosis and staging working group report. Biol. Blood Marrow Transplant 2005, 11: 945-956.
- Harris AC, Young R, Devine S, Hogan WJ et al. International, multicenter standardization of acute graft-versus-host-disease clinical data collection. A report from the Mount Sinai Acute GVHD International Consortium. Biol. Blood Marrow Transplant 2016, 22: 4-10.
- Jagasia MH, Greinix HT, Arora M et al. National Institutes of Health consensus development project on criteria for clinical trials for chronic graft-versus-host disease: I – the 2004 Diagnosis and Staging Working Group report. Biol. Blood Marrow Transplant 2015, 21: 389-400.
- Prezepiork D, Weisdorf D, Martin P, Klingemann HG, Beatty P, Hows J, Thomas ED. 1994 consensus conference on acute GVHD grading. Bone Marrow Transplant 1995, 15: 825-28.

NHS England
Wellington House
133-155 Waterloo Rd
London
SE1 8UG