Input from manufacturer on the 2nd draft assessment "GLASDEGIB IN COMBINATION WITH LOW-DOSE CYTARABINE, FOR THE TREATMENT OF NEWLY DIAGNOSED DE NOVO OR SECONDARY ACUTE MYELOID LEUKAEMIA (AML) IN ADULT PATIENTS WHO ARE NOT CANDIDATES FOR STANDARD INDUCTION CHEMOTHERAPY"

Project ID: PTJA12

eunethta

EUROPEAN NETWORK FOR HEALTH TECHNOLOGY ASSESSMENT

The objective of this reviewer form is to standardise the process of the factual accuracy check of the rapid relative effectiveness assessments.

The 2nd version of the Rapid Assessment of glasdegib in combination with low-dose cytarabine, for the treatment of newly diagnosed de novo or secondary acute myeloid leukaemia (aml) in adult patients who are not candidates for standard induction chemotherapy was open to review by the manufacturer [Pfizer] between **13/07/2020 and 17/07/2020**.

Comments received from:

Market Authorisation Holder
Pfizer

All received comments are formally responded in this combined document, to be published on the EUnetHTA website, name of organisation/institution (or individual names of the reviewers/affiliations) disclosed.

Comments from Market Authorisation Holder [Pfizer]

Page	Line	Comment	Character of comment [1]	Reply from author
Topic: Treatment Guidelines				
18 30	17-20 18-20	**Error:** LDAC is referred to as a second choice treatment option after HMAs. **Text from EUnetHTA draft assessment:** "However, if no clinical trial is available, decitabine or azacitidine are currently the first choice in newly diagnosed unfit (i.e. not eligible for standard induction and consolidation chemotherapy) AML patients. Further treatment options are low dose cytarabine (LDAC) or best supportive care (BSC) for patients who cannot tolerate any antileukaemic therapy, or who do not wish any therapy (2, 3)." **Comment from Pfizer:** ELN guidelines recommend azacitidine, decitabine and LDAC as equal treatment alternatives for patients who are not candidates for standard induction therapy. For patients who cannot tolerate any antileukaemic therapy, or who do not wish any therapy, only BSC is recommended. ESMO guidelines (Heuser et al., 2020) recommend as first-line treatment of AML patients not eligible for standard induction and consolidation chemotherapy the HMAs azacytidine and decitabine. Given the moderate effects of HMAs, LDAC remains an alternative to HMAs in the first-line treatment of AML patients who are ineligible for standard induction and consolidation chemotherapy, except in patients with adverse-risk cytogenetics, where LDAC has very poor	2	No change. The ESMO guidelines mention HMAs as first choice therapies. In addition, they state, that LDAC remains an alternative due to the moderate effects of HMAs. The ELN Guidelines state that - even though LDAC is mentioned alongside HMAs - OS improvements of LDAC is unsatisfactory. In general, the guidelines state that treatment of unfit and older patients is unsatisfactory therefore enrolment in a clinical trial is strongly recommended.

[1] Character of comment
- 'major'=1
- minor'= 2

'linguistic'=3

Page	Line	Comment	Character of comment [1]	Reply from author
		activity. BSC or LDAC are remaining options for MDS patients progressing to AML under HMA treatment, if no clinical trial is available. **Proposed amendment:** However, if no clinical trial is available, *the HMAs decitabine and azacitidine or low dose cytarabine (LDAC)* are currently the first choice in newly diagnosed unfit (i.e. not eligible for standard induction and consolidation chemotherapy) AML patients. *Best supportive care (BSC) is an alternative for patients who cannot tolerate any antileukemic therapy, or who do not wish any therapy.*		
Topic: Indirect comparison				
20 21 22 Section 4.11.2	45-47 36-39 6-8	**Error:** The indirect comparison was not considered to provide reliable results and therefore was not included in the assessment. **Comment from Pfizer:** The reliability of the ITC/STC was not formally assessed and therefore, is out of scope for the assessment. The indirect comparison was conducted using both classical ITC and simulated treatment comparison, which are validated evidence-based methods supported by literature as well as expert opinion. These various approaches are accepted to reliably control for within- and between-trial differences. Indirect comparisons showed that glasdegib+LDAC was statistically significantly associated with longer OS when compared to azacitidine and decitabine. After adjusting for potential differences in patient baseline characteristics, the OS benefit remained consistently in favour of glasdegib+LDAC, with all estimated HRs below 0.6 and confidence intervals not including the null value (HR=1) (Tremblay et	1	No change Generally, we agree that ITC and simulated treatment comparisons are accepted evidence based methods. However, results from these methods can only be regarded as valid and meaningful (with a reasonable degree of certainty), when central assumptions are plausibly met. We provide a number of reasons and explanations within the assessment why in this case central assumptions were judged to be likely violated.

Page	Line	Comment	Character of comment [1]	Reply from author
		al, 2019). While STC methods account for patient characteristics, both STC and classical ITC methods can be complementary owing to their distinctive strengths and limitations. In sum, STC can increase the reliability of comparative-effectiveness methods, inform healthcare decision making, and support clinical inferences for different and broader patient populations (Phillippo et al, 2016), such as those with AML who are not candidates for intensive chemotherapy. Indirect comparisons showed that glasdegib+LDAC was statistically significantly associated with longer OS when compared to azacitidine and decitabine. **Proposed amendment:** *Indirect comparisons showed that glasdegib+LDAC was statistically significantly associated with a better OS HR when compared to azacitidine and decitabine. After adjusting for potential differences in patient baseline characteristics, the OS benefit remained consistently in favour of glasdegib+LDAC.*		
Topic: Multiple testing and pre-planning				
21 72 Multiple pages	10-18 10-14 Multiple lines	**Text from EUnetHTA draft assessment:** "Even though further efficacy endpoints (CR, ORR, transfusion independency, quality of survival measured by Q-TWiST) suggest improvements of glasdegib+LDAC in comparison to LDAC alone, these results were considered exploratory as supportive evidence for the primary endpoint. Reasons were that the pre-planned endpoint CR was not controlled for multiple testing and ORR, transfusion independence and quality of survival were not pre-planned endpoints and the Q-TWiST results suffer from unexplained inconsistencies. Due	1	No change. The assessment is based on phase 2 results from a single phase 1b/2 study. For assessing findings from a single study, adequate predefined multiplicity control was judged to be important for

Page	Line	Comment	Character of comment [1]	Reply from author
		to these limitations and due to the fact that mature OS data are available, the most important patient-relevant endpoint, no risk of bias or certainty of evidence assessment according to GRADE had been performed for these exploratory efficacy outcomes." **Comment from Pfizer on multiple testing:** Since this study was originally designed as a Phase 2 study, a procedure to control the overall type I error for multiple testings was not pre-specified in the Statistical Analysis Plan (SAP). As such, to control the overall type I error at 0.10, the gatekeeping testing procedure was retrospectively proposed to adjust for multiple statistical testing as follows: • Test the primary endpoint OS in the intent-to-treat (ITT) population (N=132). If the 1-sided p-value from this analysis is <0.10, then declare statistical significance for OS in this population and proceed to the next testing; otherwise stop the testing. • Test the primary endpoint OS in AML patients of the ITT population (N=116). If the 1-sided p-value from this analysis is <0.10, then declare statistical significance for OS in this population and proceed to the next testing; otherwise stop the testing. • Test the secondary endpoint of complete response (CR) in the ITT population. If the 1-sided p-value from this analysis is <0.10, then declare the statistical significance for CR in this population and proceed to the next testing; otherwise stop the testing. • Test the secondary endpoint CR in AML patients in the ITT population. If the 1-sided p-value from this analysis is <0.10, then declare statistical significance for CR in this population.		confirmatory efficacy claims, in line with accepted regulatory standards for clinical trials (e.g. ICH E9 statistical principles for clinical trials) and within national HTA procedures. Post-hoc multiplicity adjustments are statistically not justifiable. Therefore, apart from the primary endpoint (OS) all other endpoints are regarded exploratory and as supportive evidence for the primary endpoint.

Page	Line	Comment	Character of comment [1]	Reply from author

This statistical testing procedure ensures that the overall alpha level is controlled at or below 0.10.

Table 1 summarizes these analyses. Using the gatekeeping testing procedures described above to control the overall type I error, the study demonstrated statistically significant improvements in OS and CR in both the ITT population and AML patients in the ITT populations.

Table 1. Summary of Analysis Results Overall Survival and Complete Remission in the ITT Population and AML Patients in the ITT Population

	ITT Population N=132		**AML Patients in the ITT Population N=116**	
Endpoint	**Glasdegib + LDAC N=88**	**LDAC N=44**	**Glasdegib + LDAC N=78**	**LDAC N=38**
Overall Survival				
No (%) patients with event	68 (77.3)	41 (93.2)	59 (75.6)	35 (92.1)
Median[a] (95% CI) (months)	8.8 (5.0, 11.7)	4.9 (2.9, 6.5)	8.3 (4.7, 12.2)	4.3 (1.9, 5.7)
HR	0.513		0.463	
1-sided p-value	0.0004		0.0002	
Complete Remission				
No (%) patients with CR	15 (17.0)	1 (2.3)	14 (17.9)	1 (2.6)
1-sided p-value	0.0071		0.0105	
2-sided p-value reported in submission	0.0142		0.0210	

		Source: SCE intext tables 32 and 35; Day 120 Table 14.2.2.3.2.2.2.1.E1AM; Day 120 Table 14.2.2.3.2.2.2.1.E1A. a. Time to event The intervention, in this particular case glasdegib+LDAC, should be comprehensively investigated with regard to the domains of mortality, morbidity, safety and health-related quality of life. If the simple question of accepting or rejecting a null hypothesis would be of primary interest, health technology assessments would not be necessary. Nevertheless, the results should always be interpreted in the context of the other analyses. As missing adjustments for multiple tests have not been generally criticized in EUnetHTA assessments before, we would recommend to harmonize the EUnetHTA assessments with regard to multiple testing. In summary, the totality of the data reflected in the improvement in CR, ORR, transfusion independence and quality of survival along with the statistically significant and clinically meaningful improvement in OS demonstrated the treatment benefit of glasdegib+LDAC. **Comment from Pfizer on pre-planning of ORR, transfusion independence and quality of survival:** In the project plan within the framework of the PICO criteria, the authors defined ORR, transfusion independency and health-related quality of survival as relevant for this assessment. Therefore, it cannot be assumed that the selection and presentation of the requested endpoints was selectively performed by Pfizer and their interpretation limited. **Comment from Pfizer on inconsistencies of Q-TWiST analysis:** Please see Section**Error! Reference source not found.** below. **Proposed amendment:** *Further efficacy endpoints (CR, ORR, transfusion independency) and quality of survival measured by Q-TWiST showed improvements of*		

Page	Line	Comment	Character of comment [1]	Reply from author
		glasdegib+LDAC in comparison to LDAC alone. Due to the fact that mature OS data are available, the most important patient-relevant endpoint, no risk of bias or certainty of evidence assessment according to GRADE had been performed for these additional efficacy outcomes."		
21	43-45	**Text from EUnetHTA draft assessment:** "Further efficacy outcomes suggest an advantage of glasdegib+LDAC over LDAC alone but were considered exploratory and thus as supportive evidence for the primary endpoint." **Comment from Pfizer on multiple testing and pre-planning of further efficacy outcomes:** Please see comment above. **Proposed amendment:** Further *supportive* efficacy outcomes *showed* an advantage of glasdegib+LDAC over LDAC alone *and support evidence of the primary endpoint*.	2	No change. We still think that the current formulation is appropriate.
Topic: Risk of bias assessment for OS:				
Multiple pages	Multiple lines	**Error in the Risk of Bias assessment for OS due to concomitant/subsequent therapies:** Authors judged some concerns regarding deviations from intended interventions due to the fact that allowed concomitant/subsequent therapies during the study were not pre-specified in the protocol. This statement is not explained further in section 4.8. **Comment from Pfizer:** According to the RoB2 guidance (Sterne et al., 2019), domain (2.3) assesses problems that arise when changes from assigned intervention that are inconsistent with the trial protocol arose because of the trial context. As 75/78 patients randomized to glasdegib+LDAC	1	Minor change. Overall, due to the open-label design of the B1371003 trial and potential differences in the administration of (known) concomitant interventions ("performance bias") across 81 study sites in Europe and North America the risk of bias judgement is regarded appropriate.

Page	Line	Comment	Character of comment [1]	Reply from author
		were treated with glasdegib+LDAC and 36/38 patients randomized to LDAC alone were treated with LDAC alone, most patients received their assigned intervention. The study protocol states the following restriction regarding concomitant therapies: • Concomitant administration of multiple moderate/strong CYP3A4/5 inhibitors, TdP drugs, and/or QT prolonging medications (without a risk of TdP) is not recommended and must be discussed with the Sponsor Medical Monitor. • Prior or concurrent treatment with a Hedgehog inhibitor or concurrent treatment with other investigational agents or approved oncology agent not specified in the protocol is not permitted. • Use of low dose dexamethasone is allowed. Once a patient had discontinued therapy, they were followed for survival. As they were no longer on protocol, they could be treated according to local practice. This reflects the state of the art in oncological studies. A table of concomitant therapies/medications given to patients in BRIGHT AML 1003 for both, the glasdegib+LDAC arm and the LDAC alone arm was provided with the submission file as an attachment to EUnetHTA. There is no evidence that non-protocol interventions are not balanced between treatment arms. In addition, a new figure was provided to EUnetHTA as an attachment (Figure 14.2.3.1.3.12.1 – based on long-term follow-up data), demonstrating that when patients are censored at the start of follow-up systemic therapies, the OS HR benefit and statistical significance is maintained.		However, the sentence was adapted to: *"Due to the open-label design of the study, administration of concomitant therapies could have been influenced by knowledge of the assigned intervention."*

Page	Line	Comment	Character of comment [1]	Reply from author
21	34-35	**Proposed amendment:** Delete sentence "concomitant/subsequent therapies were not pre-specified in the study protocol."		
53	27-28	**Proposed amendment:** Delete sentence "In addition, concomitant/subsequent therapies allowed during and after the study were not pre-specified in the protocol."		
53	Figure 4	**Proposed amendment:** Assign "*low risk of bias*" for deviations from intended interventions		
78	Table A3	**Proposed amendment:** Assign risk of bias judgement "*low*" for bias due to deviations from intended interventions Assign risk of bias judgement "*low*" for 'overall' bias for OS.		
81	Table A4	**Proposed amendment:** Assign "*not serious*" for risk of bias for OS.		
Topic: GRADE assessment for OS:				
21 23 82 Multiple pages	31 16 4 Multiple lines	**Error in the GRADE assessment for OS due to a potential overestimation of the treatment effect:** Based on the authors' rationale, the certainty of evidence for OS was downgraded 1 level because BRIGHT AML 1003 itself was not stopped early for benefit, no phase 3 trial was conducted due to the observed large treatment effect in this small, single phase 1b/2 trial. A substantial overestimate of the treatment effect can be considered likely"	1	No change. Because of the observed effect size of the phase 2 primary endpoint, no subsequent (phase 3) trial was conducted like indicated in your core submission dossier and the EPAR. Given approval was

Page	Line	Comment	Character of comment [1]	Reply from author
		Comment from Pfizer: A substantial overestimate of the treatment effect should be considered unlikely. The treatment effect considering mature OS data is HR [95 %-CI]: 0.46 [0.30, 0.72]; one-sided p = 0.0002, in favour of glasdegib+LDAC. A p-value of 0.0002 meant that there was 0.02% probability for the cohort to observe this impressive OS results if the true HR were 1.0. That is, the result of the OS analysis is extremely unlikely (2 in 10000) under the null hypothesis. This, along with the overall maturity of the data (with more than 80% of death events reported), support the rationale that this was not a small sample size for demonstrating the benefit:risk of glasdegib in this setting, and an overestimation is extremely unlikely. Additionally, the statistically significant and clinically meaningful improvement in OS demonstrated by glasdegib+LDAC was sufficiently robust for a full EMA approval. No confirmatory phase 3 trial was required for approval and therefore, EMA concluded that there was no substantial uncertainty in the evidence.		granted based on phase 2 results from a small phase 1b/2 trial, a large observed effect size in the primary endpoint can be considered mandatory. This leads to an increased probability of an overestimated effect size for the primary endpoint compared to trial programs with a subsequent phase 3 trial. Naturally, the certainty of evidence regarding the magnitude of the effect would have been considered higher, if a subsequent phase 3 trial had been conducted and had been able to confirm an effect of comparable magnitude. This reduced certainty is reflected by downgrading in GRADE. The small p-value of 0.0002 can be seen as evidence against the hypothesis that there is no effect at all, but is not informative for the question whether the true

Page	Line	Comment	Character of comment [1]	Reply from author
				effect might be of smaller magnitude than initially observed in this single small phase 1b/2 trial. We are not arguing that the hypothesis of no treatment effect is likely, merely that a smaller treatment effect is very much plausible.
Multiple pages	Multiple lines	**Comment on the GRADE assessment for OS due to the optimal information size:** The authors applied an additional 1 level downgrade of the certainty of evidence of OS when they assumed the optimal information size criterion was not met. According to the GRADE handbook (GRADE-Handbook, 2013), the optimal information size criterion is based on the "total number of patients generated by a conventional sample size calculation for a single adequately powered trial". No specific number is recommended for the significance level. The sample size estimation for BRIGHT AML 1003 was done correctly (see comment on sample size calculation on fact check page 18-21). It is not based on a significance level of 5%, which is often used in p3 clinical trials, but on a significance level of 10% which is not uncommon in p2 clinical trials. A 1 level downgrade might be appropriate when strictly following the steps proposed in the GRADE handbook. However, the author's assessment of the imprecision within the GRADE framework does not	1	No change. Phase 2 of the B1371003 trial was initially planned for a 1-sided alpha level of 10% (and corresponding 80% confidence intervals), as correctly described in the assessment. However, applying the accepted standard alpha level of 5% (2-sided) for efficacy claims in the regulatory context, the study was underpowered, the optimal information size criterion was not met, and hence certainty of evidence was downgraded in accordance to GRADE guidelines.

Page	Line	Comment	Character of comment [1]	Reply from author
		consider the significant treatment and the extremely small p-value for OS (one-sided p = 0.0002). Considering these aspects, the downgrading of 1 level due to the optimal information size does not lead to a downgrading of the certainty of evidence. Given the reasons above on an assumed overestimation of the treatment effect and on the optimal information size criterion, the downgrading of the certainty of evidence of OS by 2 levels total is not warranted. Even if the concerns regarding the optimal information size are taken into account by the downgrading of 1 level, the certainty of evidence for OS within the GRADE assessment is high.		
		Given the comments on the GRADE assessment for OS above we propose the following amendments:	1	See reply above.
21	4-5	**Proposed amendment:** ~~However,~~ the certainty of evidence according to GRADE was considered *high*.		
22	1-2	**Proposed amendment:** According to GRADE, the certainty of the evidence is considered *high* for OS, fatal AEs and SAEs and *moderate* for treatment discontinuations due to AE.		
23	Table 0.2	**Proposed amendment:** Do not downgrade the certainty of the evidence for OS due to a highly unlikely overestimation (0.02% probability). Assign *high* certainty of evidence for OS. Adjust footnotes and "certainty of the evidence" score accordingly.		

Page	Line	Comment	Character of comment [1]	Reply from author
71 74 81	26 9 Table A4	**Proposed amendment:** ~~However,~~ the certainty of evidence according to GRADE is considered *high*. **Proposed amendment:** According to GRADE, the certainty of the evidence is considered *high* for OS, fatal AEs and SAEs and *moderate* for treatment discontinuations due to AE. **Proposed amendment:** Do not downgrade the certainty of the evidence for OS due to a highly unlikely overestimation (0.02% probability). Assign *high* certainty of evidence for OS. Adjust footnotes and "certainty of the evidence" score accordingly.		
Topic: Risk of bias assessment for safety endpoints				
Multiple pages	Multiple lines	**Error in the Risk of Bias assessment for safety endpoints due to open-label:** Authors judged high risk of bias regarding measurement of the outcome for the safety endpoints SAE, fatal AE and treatment discontinuations due to AE due to the open-label design of BRIGHT AML 1003 Phase 2 non-intensive portion. **Comment from Pfizer:** The reporting of AEs by non-blinded investigators is not associated with a high risk of bias, as the safety assessment follows strict and pre-specified procedures. Especially for SAE and fatal AEs, the subjective component can be disregarded: The seriousness of AEs is determined by clearly defined and objective criteria that can hardly be	1	Minor amendment. Concerning fatal AEs: In GRADE, we changed the overall risk of bias from "very serious" to "serious" resulting in a low certainty of evidence in GRADE. For SAEs and treatment discontinuation, we still think that the open-label design in addition to investigator

Page	Line	Comment	Character of comment [1]	Reply from author
		controlled by the investigator (CTCAE Version 4.0). In accordance with the judgement for OS, fatal AEs should not be downgraded due to the open-label design of the study as death is a hard and objective endpoint. For this reason, EUnetHTA's assessment of Polatuzumab vedotin states regarding the risk of bias of SAE: "Even though this is an open-label study, the risk of bias was considered to be low, as the determination of death and of SAEs is considered sufficiently certain." (PTJA06 – Polatuzumab vedotin-Final Assessment report). We would recommend to harmonize the EUnetHTA assessments with regard to the risk of bias assessment. Given the reasons above, including the domain deviations from intended interventions, the risk of bias for "fatal AEs" and "SAEs" is low. Some concerns regarding a potential bias may remain for "treatment discontinuations due to AE" due to the fact that the investigator was not blinded.		assessed outcomes might have influenced the results and downgrading by two levels is thus still considered appropriate. Concerning the polatuzumab trial: Despite the statement that determination of death and SAEs is considered sufficiently certain, the RoB on bias on study level was rated as high, because the study was open-label and not free from potential sources of bias […]. This also led to a high RoB on endpoint level even though no further limitations were identified on outcome level for SAEs.
53	25-26	**Proposed amendment:** Due to the open-label design of the study, *some concerns remain regarding the risk of bias for treatment discontinuations due to AE.*		
53	Figure 4	**Proposed amendment:** Assign "*low risk of bias*" for "'measurement of the outcome" and for "overall" for "fatal AEs" and 'SAEs". Assign "*some concerns*" of bias for "measurement of the outcome" and for "overall" for "treatment discontinuations due to AE".		

Page	Line	Comment	Character of comment [1]	Reply from author
78	Table A3	**Proposed amendment:** Assign risk of bias judgement "*low*" for bias due to "deviations from intended interventions" for "fatal AEs", "SAEs" and "treatment discontinuations due to AE". Assign risk of bias judgement "*low*" for bias due to "measurement of the outcome" for "fatal AEs" and "SAEs". Assign risk of bias judgement "*some concerns*" for bias due to "measurement of the outcome" for "treatment discontinuations due to AE". Assign risk of bias judgement "*low*" for "overall" bias for "fatal AEs" and "SAEs". Assign risk of bias judgement "*some concerns*" for "overall" bias for "treatment discontinuations due to AEs".		
81	Table A4	**Proposed amendment:** Assign "*not serious*" for risk of bias for "fatal AEs" and "SAEs". Assign "*serious*" for risk of bias for "treatment discontinuations due to AEs".		
Topic: GRADE assessment for safety endpoints				
Multiple pages	Multiple lines	**Error in the GRADE assessment for SAEs due to number of events and width of CI:** The authors incorrectly downgraded the certainty of evidence by 2 levels due to allegedly very few events and CIs around both relative and absolute estimates of effect that include both appreciable benefit and appreciable harm for SAE.		No change. Even though the majority of patients enrolled in BRIGHT AML might have had these AEs (SAEs, fatal AEs, and

Page	Line	Comment	Character of comment [1]	Reply from author
		Comment from Pfizer: As 78.7% of patients in the glasdegib+LDAC arm and 77.8% of patients in the LDAC alone arm had at least one SAE, there were more than a few events reported. The fact that there is no significant difference between the treatments alone cannot lead to a downgrading of the certainty of evidence. Rather, it shows that there is no evidence that treatment with glasdegib+LDAC is less safe than LDAC alone regarding SAEs. Given the reasons above on the risk of bias and the alleged imprecision, the downgrading of the certainty of evidence of SAEs by 4 levels total is not warranted, thus the certainty of evidence for SAEs within the GRADE assessment is high.		treatment discontinuations due to AEs), the overall sample size was small and the median treatment duration was short. Therefore, we consider downgrading for "few events" as appropriate. For SAEs, certainty of evidence was downgraded twice since the CI (0.82 to 1.25) potentially includes both appreciable benefit and appreciable harms.
Multiple pages	Multiple lines	**Error in the GRADE assessment for fatal AEs and treatment discontinuations due to AE due to number of events and width of CI:** The authors incorrectly downgraded the certainty of evidence by 1 level due to allegedly "few events and wide CIs that overlaps no effect but potentially clinical relevant relative effects" for fatal AEs and treatment discontinuations due to AE. **Comment from Pfizer:** The AE rate and the confidence intervals do not warrant downgrading the certainty of evidence by 1 level. There were more than a "few events" reported. With regards to fatal AEs, 29.3% of patients in the glasdegib+LDAC arm and 44.4% of patients in the LDAC alone arm died. With regards to treatment discontinuations due to AE, 30.7% of patients in the glasdegib+LDAC arm and 47.2% of patients in the LDAC alone arm discontinued treatment due to AEs.		See reply above.

Page	Line	Comment	Character of comment [1]	Reply from author
		The absence of a statistically significant effect (e.g. wide CIs that overlaps no effect) alone cannot directly lead to a downgrading of the certainty of evidence (GRADE-Handbook, 2013). The relative benefit was RR [95%CI]: 0.66 [0.40-1.10] for Fatal AEs and 0.65 [0.40-1.05] for Treatment Discontinuation due to AE. In addition, there is a statistically significant and clinically relevant benefit for fatal AEs within the first 90 days of therapy (RR [95 % CI]: 0.44 [0.23; 0.87], p-value = 0.0184). Given the reasons above on the risk of bias and the alleged imprecision, the downgrading of the certainty of evidence for "fatal AEs" by 3 levels in total is not correct, thus the certainty of evidence for fatal AEs within the GRADE assessment is high. For "treatment discontinuations due to AE", a 1 level downgrade due to the concerns regarding the risk of bias may be appropriate. The downgrading by 2 levels due to number of events and width of CI is not warranted. Thus, the certainty of evidence for "treatment discontinuations due to AE" is at least moderate.		
21 22	24-25 1-2	**Given the comments on the GRADE assessment for safety endpoints above we propose the following amendments:** **Proposed amendment:** ~~However,~~ the certainty of evidence according to GRADE was considered *high* for fatal AEs and SAEs and *moderate* for treatment discontinuation due to AE. **Proposed amendment:** According to GRADE, the certainty of the evidence is considered *high* for OS, fatal AEs and SAEs and *moderate* for treatment		See reply above.

Page	Line	Comment	Character of comment [1]	Reply from author
		discontinuation due to AE.		
23	Table 0.2	**Proposed amendment:** Do not downgrade the certainty of the evidence for fatal AEs and SAEs by 2 levels due open-label design of the study. A downgrade of the certainty of the evidence by 1 level for treatment discontinuation due to AE due open-label may be appropriate. Do not downgrade the certainty of the evidence for the safety outcomes by 2 levels due to number of events and width of CI. Assign *high* certainty of evidence for fatal AEs and SAEs. Assign *moderate* certainty of evidence for treatment discontinuations due to AEs. Adjust footnotes and 'certainty of the evidence' score accordingly.		
71	40-41	**Proposed amendment:** ~~However,~~ the certainty of evidence according to GRADE was considered *high* for fatal AEs and SAEs and *moderate* for treatment discontinuation due to AE.		
74	9-10	**Proposed amendment:** According to GRADE, the certainty of the evidence is considered *high* for OS, fatal AEs and SAEs and *moderate* for treatment discontinuation due to AE.		
81	Table A4	**Proposed amendment:** Do not downgrade the certainty of the evidence for fatal AEs and SAEs by 2 levels due open-label design of the study.		

Page	Line	Comment	Character of comment [1]	Reply from author
		A downgrade of the certainty of the evidence by 1 level for treatment discontinuation due to AE due open-label may be appropriate. Do not downgrade the certainty of the evidence for the safety outcomes by 2 levels due to number of events and width of CI. Assign "*not serious*" for imprecision for "SAEs", "fatal AEs" and "treatment discontinuations due to AEs". Assign *high* certainty of evidence for fatal AEs and SAEs. Assign *moderate* certainty of evidence for treatment discontinuations due to AEs. Adjust footnotes and "certainty of the evidence" score accordingly.		
Topic: AML patients as a subpopulation of BRIGHT AML 1003 Phase 2 non-intensive portion				
20 21 Multiple Pages	29-30 28-29 Multiple lines	**Error:** The AML patients of the BRIGHT AML 1003 Phase 2 non-intensive portion (Cohort E1A) were considered as a subgroup. **Text from EUnetHTA draft assessment:** "However, the core submission dossier contains information only for the subgroup of AML patients (n = 116) (BRIGHT AML 1003)" **Comment from Pfizer:** Data are presented for all AML patients in the BRIGHT AML 1003 Phase 2 non-intensive portion. These patients reflect the marketing authorization and the population defined in the project plan within the framework of the PICO criteria by the authors. Analyses for patients with different indications (e.g. MDS) are not part of the assessment and thus not relevant. AML patients are not just a subgroup of the BRIGHT AML 1003 Phase 2 cohort of unfit patients but comprise 89%	2	No change. Although the core submission dossier submitted by the MAH contains information on AML patients only (BRIGHT AML 1003), these AML patients still remain a subgroup of the Phase 2 part of the B1371003 trial that included both AML and MDS patients considered unfit for intensive chemotherapy.

Page	Line	Comment	Character of comment [1]	Reply from author
		(78/88) of patients in the glasdegib+LDAC arm and 86% (38/44) of patients in the LDAC alone arm. The remaining 11% and 14% of patients in the respective arms are MDS patients that are not included in the label and do not meet the PICO criteria for this assessment. To reflect the PICO criteria defined by the authors of this assessment, analyses for AML patients only were performed especially for this submission file. **Proposed amendment:** The core submission dossier contains information *for AML patients* (n = 116) (BRIGHT AML 1003)"		
21	30-31	**Text from EUnetHTA draft assessment:** Data are only available for a subgroup of patients from a small, single phase 1b/2 trial. **Comment from Pfizer:** Erroneous sentence. All the patients included in the assessment were from the Phase 2 portion of the study. See also comment Number 12. **Proposed amendment:** Data are available *for AML patients from a phase 2 trial*.	2	Minor amendments. Phase 2 was added where appropriate.
21	41-43	**Text from EUnetHTA draft assessment:** "[…], showed a statistically and clinically significant improvement in median OS by 4.0 months for the combination of glasdegib+LDAC in comparison to LDAC alone in a subgroup of 116 newly diagnosed AML patients." **Comment from Pfizer:** Erroneous sentence.	2	No change. Although the core submission dossier submitted by the MAH contains information on AML patients only (BRIGHT AML 1003), these AML patients still remain a subgroup of the

Page	Line	Comment	Character of comment [1]	Reply from author
		The newly diagnosed AML patients (n=116) constituted 88% of the intent-to-treat (ITT) population. The ITT population included 16 MDS patients, which is not included in the approved indication by the EMA. With the specified testing procedure in the response to Comment 5, the improvement in OS for the AML patients was considered statistically significant. **Proposed amendment:** […], showed a statistically and clinically significant improvement in median OS by 4.0 months for the combination of glasdegib+LDAC in comparison to LDAC alone *in 116 newly diagnosed AML patients*.		Phase 2 part of the B1371003 trial that included both AML and MDS patients considered unfit for intensive chemotherapy.
Topic: Estimation of treatment effect				
21 Multiple pages	31-32 Multiple lines	**Text from EUnetHTA draft assessment:** (example: page 21, line 31-32) "Thus, a substantial overestimation of the treatment effect can be considered likely." **Comment from Pfizer on the estimation of the treatment effect:** The conclusion above is in error given the weight of the evidence below. The treatment effect considering mature OS data is HR [95 %-CI]: 0.46 [0.30, 0.72]; p = 0.0002, in favour of glasdegib+LDAC. A p-value of 0.0002 means that there was 0.02% probability for the cohort to observe this impressive OS results if the true HR were 1.0. That is, the result of the OS analysis is extremely unlikely (2 in 10000) under the null hypothesis. This, along with the overall maturity of the data (with more than 80% of death events reported), support the rationale that this was not a small sample size for demonstrating the benefit:risk of glasdegib in this setting, and an	1	No change. Because of the observed effect size of the phase 2 primary endpoint, no subsequent (phase 3) trial was conducted like indicated in your core submission dossier and the EPAR. Given approval was granted based on phase 2 results from a small phase 1b/2 trial, a large observed effect size in the primary endpoint can be considered mandatory. This leads to an increased probability of an overestimated effect size for

Page	Line	Comment	Character of comment [1]	Reply from author
		overestimation is extremely unlikely. **Proposed amendment:** Please delete: "Thus, a substantial overestimation of the treatment effect can be considered likely"		the primary endpoint compared to trial programs with a subsequent phase 3 trial. The small p-value of 0.0002 can be seen as evidence against the hypothesis that there is no effect at all, but is not informative for the question whether the true effect might be of smaller magnitude than initially observed in this single small phase 1b/2 trial.
Topic: Statistical power and sample-size calculation				
Multiple pages	Multiple lines	**Error:** Assessment states that the study is not appropriately powered for the primary endpoint. **Comment from Pfizer on the concept of statistical power:** The conclusion above is in error given the study design and the weight of the evidence below. Statistical power is the long-run proportion of studies that can be expected to demonstrate a statistically significant result given the sample size, the population effect size, and the chosen significance level. Thus, potential under-powering may pose an issue for the interpretation of studies failing to demonstrate a significant result (since the probability that an existing effect would have been detected is insufficient). However, in studies demonstrating a statistically significant result, the statistical power does not affect the validity of	1	No change. Phase 2 of the B1371003 trial was powered for an alpha level of 10% 1-sided, as correctly described in the statistics chapter of the assessment. However, applying the accepted standard alpha level of 5% (2-sided) for efficacy claims in the regulatory context, the study was underpowered. This is regarded relevant for

Page	Line	Comment	Character of comment [1]	Reply from author
		the findings. **Comment from Pfizer on sample size calculation:** In the BRIGHT AML 1003 Phase 2 randomized (2:1) cohort of unfit patients, it's specified in the protocol and the statistical analysis plan (SAP) that a total of 92 events out of 132 randomized patients would provide 80% power to detect a 60% improvement in OS, which translated to a hazard ratio of 0.625 for the glasdegib+LDAC arm versus the LDAC arm, at 1-sided alpha of 0.10. As of the data cutoff date of 03 January 2017, a total of 109 OS events were observed, and the primary analysis of OS demonstrated a statistically significant and clinically meaningful improvement in OS with HR of 0.513 (95% CI: 0.343, 0.766) and 1-sided p-value of 0.0004 for AML+MDS patients and HR of 0.46 [95 %-CI: 0.30, 0.72] and 1-sided p-value of 0.0002 for AML patients only). The results suggested that the assumption in the original study design might have been too conservative. The study could also have been designed to detect a difference in OS assuming the true HR were 0.513. This would require a total of 88 events out of 132 patients to provide 80% power at the usual 1-sided alpha of 0.025 for a Phase 3 study. Under this hypothetical design, the study could still have been declared positive at the final analysis with the observed 1-sided p-value of 0.0004.		the assessment. First, power directly corresponds to the precision of the effect estimate of the primary endpoint. Second, power is related to the false discovery rate, that is, the expected proportion of false positives among significant findings.
21	32-33	**Proposed amendment:** Please delete: "Furthermore, the study was not appropriately powered for the primary endpoint"		
48	40-42	**Proposed amendment:** Please delete: "Of note, although within the trial 94 OS events were observed, 160 OS events would have been necessary to obtain a		

Page	Line	Comment	Character of comment [1]	Reply from author
71	44-47	statistical power of at least 80 % to detect a true HR of 0.625 at a conventional alpha level of 5 % (two-sided) (49)" **Proposed amendment:** Please delete: "However, the study was underpowered for the primary endpoint because the final analysis was planned as soon as 92 OS events were observed and was conducted at 94 OS events but to obtain a statistical power of at least 80 % to detect a true HR of 0.625 at a conventional alpha level of 5 % (two-sided) 160 OS events would have been necessary."		
Topic: Subgroup analyses by cytogenetic risk				
21 and multiple pages	6-9 And multiple pages	**Text from EUnetHTA draft assessment:** "Exploratory results from a subgroup analysis according to IVRS based cytogenetic risk, descriptively showed a lower median OS improvement in AML patients with poor cytogenetic risk (median OS improvement of 1.3 months) compared to AML patients with good/intermediate cytogenetic risk (median OS improvement of 6.7 months)." **Comment from Pfizer:** HR reflects the entire Kaplan-Meier curves throughout the study follow-up since randomization while the median is a key point on the KM curve but doesn't fully reflect the area between the entire KM curves. Interpretation of subgroup analyses should be taken with caution given the nature of subgroup analyses with the limited sample size (49 versus 21 patients with good/intermediate cytogenetic risk and 29 versus 17 patients with poor cytogenetic risk). Both subgroups showed an advantage for glasdegib+LDAC and HR [95% CI] does not substantially differ with overlapped 95% CIs (0.417 [0.233, 0.744] for patients with good/intermediate cytogenetic risk and 0.528 [0.273, 1.022] for patients with poor cytogenetic risk).The width of the	1	No change. The only subgroup analysis that was prespecified in the B1371003 trial was for cytogenetic risk. Therefore, consistent with reporting of other OS results, results including Hazard Ratios and 95 % confidence intervals are presented in the text and/or respective table not only for the AML subpopulation but also for ITT population of the B1371003 trial containing AML and MDS patients.

Page	Line	Comment	Character of comment [1]	Reply from author
		confidence interval depends also on the sample size and is therefore different between the subgroups. The corresponding interaction test does not indicate different effects between subgroups, so the treatment effect should be interpreted for all patients. **Proposed amendment:** *Exploratory results from a subgroup analysis according to IVRS based cytogenetic risk, showed no substantial differences between AML patients with poor cytogenetic risk (HR [95% CI]: 0.528 [0.273, 1.022]) compared to AML patients with good/intermediate cytogenetic risk (HR [95% CI]: 0.417 [0.233, 0.744]).* Please delete sentence: "as well as for the 55 AML + MDS patients with poor cytogenetic risk within 19 the prespecified subgroup analysis (HR [95 % CI]: 0.63 [0.35, 1.15]; p = 0.0640; median OS reduction 20 of 0.2 months)" as MDS patients are not included in the label and do not meet the PICO criteria for this assessment.		
21 56 56	6-9 17-19 19-21			
Topic. Q-TWiST Analysis				
21	46-47	**Error:** Q-TWiST analysis was stated to be inconsistent, not valid and not reliable. **Text from EUnetHTA draft assessment:** "Patient reported outcomes were not proactively collected and the validity and reliability of the submitted explorative results from a quality of survival analysis is unclear." **Comment from Pfizer:** An evaluation of the validity and reliability of the Q-TWiST method was not conducted, therefore it is inappropriate to make a statement about the analysis being clear or unclear.	2	Minor amendments. After resolving most of the inconsistencies encountered in the submission file regarding the Q-TWiST analysis, we amended the Results section, based on your suggestion.

Page	Line	Comment	Character of comment [1]	Reply from author
		However, quality-adjusted mean survival favored the glasdegib+LDAC group by a range of 3.5 to 4.5 months, out of the total mean survival benefit of 4 months. This magnitude of benefit was consistent across the different methods used to derive weights, and is well demonstrated in Table 23 of the submission file. Sensitivity analysis varied the length of follow-up (6 to 24 month) and AE definitions (including all adverse events regardless of grade) and showed robust results; subgroup analyses were also performed and showed consistent effects. It therefore clearly establishes that the majority of survival benefit derived from therapy with glasdegib+LDAC is time with good quality. **Proposed amendment:** Patient reported outcomes were not proactively collected, *but the results from a quality of survival analysis suggest that glasdegib+LDAC is associated with improvement in quality-adjusted survival relative to LDAC alone*.		
Multiple pages	Multiple lines	Per our discussions, the following amendments were made to our Submission Dossier • The original Figure 19 was updated to correct for an error • 1 paragraph on page 80 was updated with new numbers to accurately reflect the derivation of the numbers and subsequent results in Figure 19 • Glasdegib+LDAC patients (n = 78) had significantly longer mean time in *TWiST (+ 3.4 [95 % CI]: [1.8, 5.2] months*) and TOX (*+ 0.4 [0.1, 0.8] months*), and longer but non-significant REL (*+ 0.2 [-1.0, 1.2] months*) when compared to LDAC patients (n = 38) (Figure 19). These mean times were multiplied by the respective utilities for *TOX*0.5, REL*0.5 and TWIST*1.0 (i.e., the base case)*.	3	Major changes. Due to the initial submission of an erroneous core submission dossier by the MAH, major adaptations had to be implemented by the authors to incorporate the corrected results of the Q-TWiST analyses. However, minor unexplained numerical inconsistencies still remain, i.e. conflicting results in text, table and figure.

Page	Line	Comment	Character of comment [1]	Reply from author
		This correction/amendment of the Q-TWiST data may impact the assessment report in the following places: • P21, line 15 • P58, lines 29-39 • P59, lines 1-2 P72, lines 50-51		
59	1-2	**Text from EUnetHTA draft assessment:** "Time spent in the TWiST, TOX, and REL states were computed to be 3.5, 0.8, and 0.3 months higher under glasdegib+LDAC arm compared to 1 LDAC alone, respectively." **Comment from Pfizer:** The numbers in the TWiST, TOX, and REL states are not correct. **Proposed amendment:** Time spent in the TWiST, TOX, and REL states were computed to be *3.4, 0.4, and 0.2* months higher under glasdegib+LDAC arm compared to LDAC alone, respectively.	3	See reply above. The numbers in the TWiST, TOX, and REL states were correctly derived from Table 23 of the initial core submission dossier for the unweighted case (utility value 1 for TWiST, TOX, and REL). However, following your corrections of errors within the initial core submission dossier, the Q-TWiST part of the assessment was adapted.
Mixed Topics				
31-32	Table 2/Features of intervention	**Text from EUnetHTA draft assessment:** "Table 2. Features of interventions comprises glasdegib, azacitidine and cytarabine. " **Comment from Pfizer on the interventions presented in this table:** Both azacitidine and decitabine should be included as comparators.	2	Amended.

Page	Line	Comment	Character of comment [1]	Reply from author
		Proposed amendment: Please include decitabine or delete azacitidine if only the interventions but not the comparators should be shown in this table.		
21 53 78	34 7 Table A3	**Text from EUnetHTA draft assessment:** "Multiple places referring to the random sequence generation missing" **Comment from Pfizer:** The random sequence generation information is found in CSR Table 16.1.7.2.3 **Proposed amendment:** Please review the CSR Table and make appropriate changes accordingly	2	No change. No information on the actual generation of the random sequence can be inferred from the now submitted table.
42	34-35	**Text from EUnetHTA draft assessment:** "Nevertheless, according to the CHMP assessment report, no patient in either group was considered to have completed treatment on trial (5)." **Comment from Pfizer :** There were multiple patients treated > 1 year (both fit and unfit), and we report these data as "treatment duration". We made the decision to not record patients treated > 1 year as "Completers" because the database only allows ONE reason per patient for discontinuing treatment (End of Treatment CRF). We were concerned that if these patients still continuing treatment at 1 year were listed as "Completers" in the database but continued taking study drug per protocol and then later discontinued due to AEs or death, we would be missing important safety data. In the BRIGHT AML 1003 Phase 2 randomized (2:1) cohort, there	2	Minor amendment. The sentence was deleted. The following information was descriptively added to section 4.7 Participant flow: *"At the cut-off date 15 patients (19 %) with glasdegib + LDAC and 1 patient (3 %) with LDAC alone were ongoing in the study. Of these, 4 patients with glasdegib + LDAC (5 %) remained on treatment (5)."*

Page	Line	Comment	Character of comment [1]	Reply from author
		were 14 patients in the Glasdegib + LDAC arm treated >1 year and among them, 2 were on study therapy for >4 years. In contrast, all patients in the LDAC alone arm were treated for less than 1 year. **Proposed amendment:** Delete the following sentence**:** "Nevertheless, according to the CHMP assessment report, no patient in either group was considered to have completed treatment on trial (5)." Replace with: "*In the BRIGHT AML 1003 Phase 2 randomized (2:1) cohort, there were 14 patients in the Glasdegib + LDAC arm treated >1 year and among them, 2 were on study therapy for >4 years. In contrast, all patients in the LDAC alone arm were treated for less than 1 year."*		
72	49-50	**Text from EUnetHTA draft assessment:** "methodological descriptions for e.g. calculation of relative risk for disease response were lacking" **Comment from Pfizer:** As stated in table 24 of the core submission file, the relative risk for CR was calculated using the Cochran-Mantel-Haenszel method stratified by prognosis stratum which was IVRS based cytogenetic risk. The relative risks for CRi, MLFS and ORR (CR+CRi+MLFS), which were presented in table 25 were also calculated using the Cochran-Mantel-Haenszel method, and we would like to add that the same stratification (IVRS based cytogenetic risk) was used. **Proposed amendment:** Please delete: "methodological descriptions for e.g. calculation of relative risk for disease response were lacking."	2	No change. Thanks for clarifying. However, this information is still not included, neither in the initial nor in the amended core submission dossier submitted.
72	51-53	**Text from EUnetHTA draft assessment:**	1	Minor change.

Page	Line	Comment	Character of comment [1]	Reply from author
73	1-2	"These inconsistencies could not be adequately addressed, since - even though requested by the authors - the MAH did not submit individual patient data. Hence, an independent evaluation of the reproducibility and robustness of reported results by the authoring team was not possible. Therefore, inconsistencies encountered in the submission dossier could not be resolved by the authoring team." **Comment from Pfizer:** This sentence is erroneous. Individual patient data were only requested to validate the ITC/STC analyses; however, due to data privacy and processing procedures, we were not able to provide the individual patient data in the timeline required by EUnetHTA. We did offer to run additional analyses as needed that were specific to the ITC/STC analyses. **Proposed amendment:** Delete these sentences and if needed, move the lack of individual patient data to the ITC/STC section.		Although IPD were initially asked for in the context of ITC and STC, of course they could have been used (and were intended) to check reproducibility of OS results, resolve any inconsistencies, and potentially check the robustness of results with additional sensitivity analyses.
54	21	**Error:** Wrong p-value for OS **Text from EUnetHTA draft assessment:** "As can be seen in Table 16 and Figure 5, glasdegib + LDAC showed superior OS (HR [95 % CI]: 0.46 [0.30, 0.72]; p = 0.0004)" **Comment from Pfizer:** Wrong p-Value. This should be 0.0002 (see Table 16 of the draft assessment) **Proposed amendment:** As can be seen in Table 16 and Figure 5, glasdegib + LDAC showed superior OS (HR [95 % CI]: 0.46 [0.30, 0.72]; *p = 0.0002*)	2	Minor amendments. We now consistently report 2-sided p-values (for potentially directional/1-sided hypotheses tests) throughout the assessment to avoid any misunderstandings. In particular all reported 1-sided p-values were converted to 2-sided p-values where necessary.

Page	Line	Comment	Character of comment [1]	Reply from author
54	24/Table 16	**Error:** Wrong p-value for unstratified analysis of OS (IVRS based) **Text from EUnetHTA draft assessment:** "Unstratified analysis: 0.45 [0.29, 0.69] p=0.0001" **Comment from Pfizer:** Wrong p-Value. This should be 0.0002 (see Submission file Attachment "BRIGHT AML 1003 Additional requested tables, page 36) **Proposed amendment:** Unstratified analysis: 0.45 [0.29, 0.69] *p=0.0002*	2	Minor amendments. See reply above. In addition, despite the fact that we now consistently report 2-sided p-values, your corrections are not consistent, since in the comment above you report the 1-sided p-value and in this comment the 2-sided p-value.
61	36-40	**Error:** Wrong cutoff for SAEs/AEs leading to discontinuation **Text from EUnetHTA draft assessment:** "Considering the entire study period, the most frequently reported SAEs that occurred in ≥5 % of patients were pneumonia (21 % vs. 19 %), sepsis (4 % vs. 14 %), febrile neutropenia (28 % vs. 17 %), anaemia (7 % vs. 0), pancytopenia (0 % vs. 6 %) and disease progression (9 % vs. 11 %). During the same period, the most common AEs leading to treatment discontinuation in ≥5 % of patients comprised pneumonia (5 % vs. 3 %), sepsis (1 % vs. 6 %), and febrile neutropenia (3 % vs. 6 %)." **Comment from Pfizer:** The cutoff of ≥5 % is not correct. SAEs/AEs leading to discontinuation occurring in ≥2 % of patients are reported. **Proposed amendment:** Considering the entire study period, the most frequently reported SAEs that occurred in *≥2%* of patients 36 were pneumonia (21 % vs.	2	Changed. This has been corrected.

Page	Line	Comment	Character of comment [1]	Reply from author
		19 %), sepsis (4 % vs. 14 %), febrile neutropenia (28 % vs. 17 %), anaemia 37 (7 % vs. 0), pancytopenia (0 % vs. 6 %) and disease progression (9 % vs. 11 %). During the same 38 period, the most common AEs leading to treatment discontinuation in ≥2% of patients comprised 39 pneumonia (5 % vs. 3 %), sepsis (1 % vs. 6 %), and febrile neutropenia (3 % vs. 6 %).		
Multiple pages	Multiple lines	**Error:** In multiple places, the authors refer to the submitted clinical trial as a phase 1b/2 study. **Comment from Pfizer:** While the overall protocol included both phase 1b and phase 2 cohorts, the submitted BRIGHT AML 1003 randomized clinical trial is a phase 2 study. Therefore, it should be referred to as a phase 2 study throughout the assessment (i.e., the phase 1b label should be deleted) **Proposed amendment:** Delete the phase 1b label throughout document.	3	No change. While B1371003 was a randomized Phase 1b/2 study, BRIGHT AML 1003 reports results only for the AML subgroup.

1 Amendments of the original submission file

The following amendments were made to our Submission Dossier, due to EUnetHTA flagging an inconsistency in the Submission Dossier in the draft Assessment Report that was shared for Fact Check:

- The original Figure 19 was updated to correct for an error
- The last paragraph on page 80 was updated with new numbers to accurately reflect the derivation of the numbers and subsequent results in Figure 19
    - Glasdegib + LDAC patients (n = 78) had significantly longer mean time in *TWiST (+ 3.4 [95 % CI]: [1.8, 5.2] months*) and TOX (*+ 0.4 [0.1, 0.8] months*), and longer but non-significant REL (*+ 0.2 [-1.0, 1.2] months*) when compared to LDAC patients (n = 38) (Figure 19). These mean times were multiplied by the respective utilities for *TOX*0.5, REL*0.5 and TWIST*1.0 (i.e., the base case)*.

This correction/amendment of the Q-TWiST data may impact the assessment report in the following places:

- P21, line 15
- P58, lines 29-39
- P59, lines 1-2
- P72, lines 50-51

2 . References

Common Terminology Criteria for Adverse Events (CTCAE) Version 5.0 Published: November 27, 2017 U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES (https://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm#ctc_50)

Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. (https://gdt.gradepro.org/app/handbook/handbook.html)

Heuser M, Ofran Y, Boissel N, Mauri SB, Craddock C, Janssen J, et al. Acute myeloid leukaemia in adult patients: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of oncology : official journal of the European Society for Medical Oncology. 2020.

Phillippo DM, Ades A, Dias S, Palmer S, Abrams KR, Welton NJ. Methods for Population- Adjusted Indirect Comparisons in Submissions to NICE. Sheffield: Decision Support Unit; 2016. Accessed at : http://nicedsu.org.uk/wp-content/uploads/2018/08/Populationadjustment-TSD-FINAL-ref-rerun.pdf

PTJA06 – Polatuzumab vedotin in combination with bendamustine and rituximab for the treatment of relapsed/refractory diffuse large B-cell lymphoma (DLBCL) who are not candidates for haematopoietic stem cell transplant. – Final Assessment report. (https://eunethta.eu/wp-content/uploads/2020/02/PTJA06-Final-Assessment-Report-V1.0.pdf)

Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, Cates CJ, Cheng H-Y, Corbett MS, Eldridge SM, Hernán MA, Hopewell S, Hróbjartsson A, Junqueira DR, Jüni P, Kirkham JJ, Lasserson T, Li T, McAleenan A, Reeves BC, Shepperd S, Shrier I, Stewart LA, Tilling K, White IR, Whiting PF, Higgins JPT. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019; **366**: l4898.

Tremblay G, Westley T, Cappelleri JC, et al. Overall survival of glasdegib in combination with low-dose cytarabine, azacitidine, and decitabine among adult patients with previously untreated AML: comparative effectiveness using simulated treatment comparisons. ClinicoEconomics and Outcomes Research. September 2019.