HTA CG | MEMBER STATE COORDINATION GROUP ON HEALTH TECHNOLOGY ASSESSMENT

Joint Clinical Assessment Report

Tovorafenib

Version 1.0

Endorsed by the Coordination Group pursuant to Article 12(2) of Regulation (EU) 2021/2282 on 30 April 2026

Date of the procedural review by the European Commission pursuant to Article 28, point (d), of Regulation (EU) 2021/2282: 19 May 2026

European Commission

This document has been produced and endorsed by the Member State Coordination Group pursuant to Regulation (EU) 2021/2282 on health technology assessment and financed by the European Commission under the framework contract FWC HADEA/2024/OP/0031. The European Commission performed a procedural review pursuant to Article 28, point (d), of Regulation (EU) 2021/2282. The joint clinical assessment is the responsibility of the Member State Coordination Group, and not of the European Commission or the European Health and Digital Executive Agency (HADEA).

Brussels: Directorate-General for Health and Food Safety, European Commission, 2026

How to cite this report: Member State Coordination Group on Health Technology Assessment, *Joint Clinical Assessment Report of Tovorafenib, Version 1.0*, European Union, Brussels, 2026

General instructions

The report shall follow international standards of evidence-based medicine and take into account, if available, the methodological guidance, adopted by the HTACG pursuant to Article 3(7), point (d), of the HTAR.

List of abbreviations

Abbreviation	Definition
1L	first line
2+L	second or later line
AE	adverse event
ATC	anatomical therapeutic chemical
ATMP	advanced therapy medicinal product
BOR	best overall response
BRAF	v-Raf murine sarcoma viral oncogene homologue B
BSA	body surface area
CR	complete response
CRAF/RAF1	proto-oncogene serine/threonine-protein kinase
CSR	clinical study report
DOR	duration of response
EEA	European Economic Area
EMA	European Medicines Agency
ESS	effective sample size
EU	European Union
FAS	full analysis set
FLAIR	fluid-attenuated inversion recovery
HGG	high-grade glioma
HTA	health technology assessment
HTACG	Member State Coordination Group on Health Technology Assessment
HTAR	Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU (OJ L 458, 22.12.2021, p. 1, ELI: http://data.europa.eu/eli/reg/2021/2282/oj)
HTD	health technology developer
INV	investigator
IPD	individual participant data
IQWiG	Institute for Quality and Efficiency in Health Care

Abbreviation	Definition
IR	Commission Implementing Regulation (EU) 2024/1381 of 23 May 2024 laying down, pursuant to Regulation (EU) 2021/2282 on health technology assessment, procedural rules for the interaction during, exchange of information on, and participation in, the preparation and update of JCA of medicinal products for human use at Union level, as well as templates for those joint clinical assessments
IRC	independent review committee
ITC	indirect treatment comparison
JCA	joint clinical assessment
JSC	joint scientific consultation
KM	Kaplan-Meier
LCH	Langerhans cell histiocytosis
LGG	low-grade glioma
MAIC	matching adjusted indirect comparison
MR	minor response
MRI	magnetic resonance imaging
NCPE	National Centre for Pharmacoeconomics
ORR	objective response rate
OS	overall survival
PD	progressive disease
PFS	progression-free survival
PH	proportional hazards
PICO	A set of parameters for the JCA comprising: patient population – intervention(s) – comparator(s) – health outcomes
PR	partial response
PRIME	priority medicines scheme by the European Medicines Agency
PVs/EMs	prognostic and effect-modifying variables
RAF	rapidly accelerated fibrosarcoma
RANO	Response Assessment in Neuro-Oncology
RAPNO	Response Assessment in Paediatric Neuro-Oncology
RCT	randomised controlled trial
RMST	restricted mean survival time
RP2D	recommended phase 2 dose
R/R	relapsed/refractory
RR	relative risk

Abbreviation	Definition
RoB	risk of bias
SAP	statistical analysis plan
SD	stable disease
SEGA	subependymal giant cell astrocytoma
SLR	systematic literature review
SmPC	summary of product characteristics
STC	simulated treatment comparison
TTR	time to response
WHO	World Health Organization

Table of Contents

List of Tables

List of Figures

1 General information on the JCA

This section shall provide:

— information on the assessor and co-assessor,

— an overview of the procedural steps and their dates,

— information on the involvement of patients, clinical experts and other relevant experts, as well as on the input received from patient organisations, healthcare professional organisations and clinical and learned societies. The input from experts and stakeholders shall be provided in Appendix A,

— information on previous JSC under the HTAR.

The assessors and co-assessors for this JCA are listed in the following table.

Table 1: Information on the assessor and co-assessor

	Member State	**Agency/Institution**	**Representative**
Assessor	Ireland	NCPE	Emer Fogarty
Co-Assessor	Germany	IQWiG	Beate Wieseler

The main procedural steps are summarised in the next table.

Table 2: Overview of the procedural steps and their dates

Procedural Steps	**Date**
Submission of information for the development of the assessment scope	11 March 2025
Consolidated assessment scope proposal shared with individual experts	21 May 2025
Assessment scope shared with HTD	06 June 2025
Dossier submission Version 0.1	15 September 2025
Dossier submission Version 0.2	17 October 2025
Commission’s confirmation of the JCA dossier	10 November 2025
Dossier submission Version 0.3	05 December 2025
First draft JCA report shared with JCA SG	22 December 2025
Revised draft JCA report shared with individual experts	17 February 2026
Revised draft JCA report shared with HTD	17 February 2026
Finalisation of revised draft JCA report by JCA SG	07 April 2026
Endorsement of JCA report by HTACG	30 April 2026

Expert involvement

The following table provides information about the experts involved in the assessment. Patients/carers and clinical experts were given the opportunity to provide input on the assessment scope and on the revised draft JCA and summary reports, in compliance with Article 8(6), second paragraph, and Article 11(4) of the HTAR. This input was considered by the JCA subgroup prior to finalisation of the assessment scope and the JCA and summary reports. The clinical expert also provided input during the preparation of the draft JCA report, in response to questions from the assessors. The input from patients/carers and clinical experts is available in Appendix A.

Table 3: Information on the involvement of patients, clinical experts and other relevant experts

	Name of the expert (except for patients)/organisation or affiliation (if any)	**Type and timing of the involvement**
Patient/Carer	one carer	Written input on the consolidated assessment scope proposal, and on the revised draft JCA and summary reports
Clinical experts	Christine Dahl ▪ NOPHO [Nordic Society of Paediatric Haematology and Oncology], ▪ SIOP [International Society of Paediatric Oncology], ▪ PanCare	Written input on the consolidated assessment scope proposal, during the preparation of the draft JCA and summary reports, and on the revised draft JCA and summary reports
Other experts	–	–

Joint Scientific Consultation

The next table documents whether there has been a Joint Scientific Consultation in the indication under assessment for the medicinal product being assessed.

Table 4: Information on JSC under the HTAR

JSC under the HTAR	**Detail**
No	–

According to the HTAR, the health technology developer (HTD) should explain any deviations from recommended propositions for evidence generation in the dossier. The HTD identified the following deviations and explanations.

Table 5: Deviations from JSC propositions for evidence generation

Deviation from evidence generation proposition in JSC	Explanation
Not applicable	–

Requests for additional information

The following table provides information on the assessors' requests for additional information from the HTD. The requests and the responses to those requests by the HTD, together with their assessment, are available in Appendix D.

Table 6: Requests for additional information

	Date of request	Date of response	New dossier version submitted [yes/no]
Request 1	14 Nov 2025	21 Nov 2025	no
Request 2	28 Nov 2025	05 Dec 2025	yes

Factual accuracy check by the HTD

According to the HTAR, a factual accuracy check was performed by the HTD. The outcome of this step including the assessors' feedback is provided in a separate document "*Annex_Comments_JCA-MP-2024-06_tovorafenib"*.

The JCA report is based on information submitted by the HTD in the latest version of the dossier and as a response to the additional information requests from the assessors. The HTD is responsible for ensuring the accuracy of the submitted dossier. Information provided by the HTD in the dossier, that was identified by the HTD as incorrect only during the factual accuracy check, was therefore not acted on in the report. The nature of the incorrect information is such that there is no effect on the estimates of relative effectiveness and safety, and there is no meaningful impact on the associated uncertainties identified in the JCA. The information can be found in the document "*Annex_Comments_JCA-MP-2024-06_tovorafenib"*.

2 Background

2.1 Overview of the medical condition

This section shall provide:

— a summary of the medical condition, including the symptoms and the burden and natural progression of the medical condition, its prevalence or incidence in the EEA States in which the HTAR is applicable, as available,

— a brief description of the target patient population and its characteristics reflected in the assessment scope as set out pursuant to Article 8(6) of the HTAR,

— a brief description of the care pathway for the medical condition and whether it varies substantially between the EEA States in which the HTAR is applicable, as well as, if relevant, for different stages and/or subtypes or sub-populations of the medical condition.

Paediatric low-grade gliomas (LGG) form a heterogeneous group of grade 1 and 2 tumours (classified according to World Health Organization (WHO) criteria). In general, these tumours have a favourable prognosis with a 10-year overall survival between 85% and 96% [1]. Based on data from the US, paediatric LGGs account for an estimated 30% of all childhood brain tumours [2].

The development of paediatric LGGs is mainly determined by the presence of genetic alterations, such as the v-Raf murine sarcoma viral oncogene homologue B (BRAF) V600E mutation and KIAA1549-BRAF fusion, as well as other molecular alterations [2]. There are few published data on the incidence and prevalence of BRAF-altered paediatric LGG in Europe. In the population aged 0 to 19 years, the incidence of BRAF-altered paediatric LGG was estimated by the HTD to be 0.33 per 100,000 individuals in the EU member states [3].

Although overall survival (OS) is generally favourable, patients with paediatric LGG suffer from a significant clinical disease burden that correlates with the location of the tumour in the brain [1,4]. Tumour-related morbidities include seizures, hemiparesis, focal neurological findings, behavioural changes, visual disturbances, cognitive dysfunction and endocrine dysfunction [5]. Radiographic tumour response assessment plays a crucial role in managing patients with central nervous system tumours. Since 2010, different sets of standardised criteria were developed: Response Assessment in Neuro-Oncology (RANO) high-grade glioma (HGG), RANO-LGG and Response Assessment in Paediatric Neuro-Oncology (RAPNO) (see Section 4.2.1).

The target patient group according to the full claimed indication for this JCA consists of patients aged 6 months and older with paediatric LGG harbouring a BRAF fusion or rearrangement or BRAF V600 mutation, who have received one or more prior systemic therapies.

Although surgery and radiation therapy are broadly used to treat LGG, many patients with paediatric LGG are not candidates for surgery or radiation therapy due to treatment limitations or long-term health risks and effects and thus require systemic treatment [1]. Carboplatin in combination with vincristine is the most commonly used first-line systemic treatment for patients with paediatric LGG. In addition, vinblastine is also commonly used for first-line treatment in Europe [1]. A current European standard for treatment of paediatric LGG patients who have already received one or more prior systemic therapies has not been established. Different chemotherapy regimens are used for the treatment of these patients. In addition, targeted therapies are available for distinct patient groups.

For BRAF-altered paediatric LGG, a combination of the two targeted therapies, dabrafenib and trametinib, is available [6,7]. This combination is approved for the treatment of paediatric LGG patients aged 1 year and older who require systemic therapy and have a specific type of BRAF alteration, i.e., a BRAF V600E mutation. Everolimus is another targeted treatment option that is recommended for the distinct patient group with subependymal giant cell astrocytoma (SEGA) [1].

2.2 Characterisation of the medicinal product

2.2.1 Characteristics of the medicinal product

This section shall describe characteristics of the medicinal product under assessment ('the medicinal product') and report the following information:

— proprietary name,
— active substance(s),
— pharmaceutical formulation(s),
— therapeutic indication,
— marketing authorisation holder,
— mechanism of action,
— ATC code where already assigned.

The following table characterises the medicinal product under assessment.

Table 7: Characteristics of the medicinal product under assessment

Proprietary name	OJEMDA
Active substance(s)	Tovorafenib
Pharmaceutical formulation(s)	Powder for oral suspension or film-coated tablets
Claimed indication as submitted to EMA by the HTD	Tovorafenib is indicated as monotherapy for the treatment of patients 6 months of age and older with paediatric LGG harbouring a BRAF fusion or rearrangement, or BRAF V600 mutation, who have received one or more prior systemic therapies

Final therapeutic indication	Tovorafenib is indicated as monotherapy for the treatment of patients 6 months of age and older with paediatric LGG harbouring a BRAF fusion or rearrangement, or BRAF V600 mutation, who have progressed after one or more prior systemic therapies
Marketing authorisation holder	Ipsen Pharma
Mechanism of action	Tovorafenib is a Type II RAF kinase inhibitor of mutant BRAF V600E, wild-type BRAF, and wild-type CRAF kinases.
ATC code where already assigned	L01EC04
BRAF: v-Raf murine sarcoma viral oncogene homolog B; CRAF/RAF1: proto-oncogene serine/threonine-protein kinase; LGG: low-grade glioma; MAPK: mitogen-activated protein kinase; RAF: rapidly accelerated fibrosarcoma	

The JCA took into account the changes between the final therapeutic indication and the claimed indication, which formed the basis for the assessment scope (see Section 3). The assessment and interpretation of the available results are unaffected by the changes, and the JCA subgroup agreed that there was no need to define a new assessment scope.

2.2.2 Requirements/instructions for use

This section shall include a description of the methods of administration, dosing of the medicinal product and duration of treatment.

Details on the administration and dosing of the medicinal product under assessment are described in the following table.

Table 8: Administration and dosing of the medicinal product under assessment

Method of administration	Tovorafenib is for oral use. The powder for oral suspension and the film coated tablets may be used interchangeably. Tovorafenib can be taken with or without food and should be taken at a regularly scheduled time once weekly. It should be administered to paediatric patients under adult supervision. The tablets should be swallowed whole with water and must not be chewed, cut, or crushed. For patients who are not able to swallow or with BSA less than 0.9 m^2 the oral suspension should be provided. Tovorafenib powder for oral suspension must be reconstituted prior to being dispensed. Prior to first time use of the oral suspension, caregivers (and if appropriate, patients) should be instructed on the proper preparation, dose, and administration of Tovorafenib.

Prerequisites for administration	Treatment with tovorafenib should be initiated and supervised by a qualified physician experienced in the use of anti-cancer medicinal products. Before taking tovorafenib, patients must have confirmation of BRAF fusion or rearrangement, or BRAF V600 mutation assessed by a CE-marked IVD medical device with the corresponding intended purpose. If the CE-marked IVD is not available, confirmation of BRAF fusion or rearrangement, or BRAF V600 mutation should be assessed by an alternative validated test.
Doses and dosing frequency	The recommended dose of tovorafenib based on BSA is 380 mg/m^2 orally once weekly. The maximum recommended dose is 600 mg orally once weekly. A recommended dose for patients with BSA less than 0.3 m^2 has not been established.
Duration of treatment (including end of treatment criteria if necessary)	Continue once weekly dosing until disease progression, loss of clinical benefit, or unacceptable toxicity.
BRAF: v-Raf murine sarcoma viral oncogene homolog B; BSA: body surface area; IVD: in vitro diagnostic	

2.2.3 Regulatory status of the medicinal product

This section shall describe the regulatory information on the medicinal product and provide details of the procedural pathway of the medicinal product in the EU, such as orphan medicinal product designation, conditional marketing authorisation with any specific obligations of the conditional marketing authorisation, ATMP or PRIME. It shall also provide details of ongoing or planned early access/compassionate use programs in the EEA.
When appropriate links to SmPC shall be inserted for details of other licensed therapeutic indications and to the dossier for further regulatory information.

Regulatory information on the medicinal product under assessment is provided in the following table.

Table 9: Regulatory information on the medicinal product under assessment

Orphan medicinal product (yes/no)	yes
Conditional marketing authorisation (yes/no)	yes
Specific obligations of the conditional Marketing Authorisation	see EPAR
Exceptional circumstances (yes/no)	no
ATMP (yes/no)	no
PRIME (yes/no)	no
First indication (yes/no)	yes
Details of ongoing early access programmes in the EU (as provided by the HTD)	NA
EMA: European Medicines Agency; EPAR: European public assessment report; NA: not applicable	

3 Assessment scope

This section shall reproduce the assessment scope as set out pursuant to Article 8(6) of the HTAR.

Assessment scope

As set out pursuant to Article 8(6) of Regulation (EU) 2021/2282 on health technology assessment

Table 10 provides an overview of the assessment scope and the submitted data per PICO[1] for the Joint Clinical Assessment (JCA) of tovorafenib. Table 11 and Table 12 show the outcomes and subgroup analyses required according to the assessment scope.

Table 10: Assessment scope and overview of submitted data per PICO

Population Intervention PICO no.	**Comparator**	**Results submitted [yes/no]**	**HTD's reason for omission**	**Data included in the JCA report [yes/no]**
Population 1 (full claimed indication)				
Patients 6 months of age and older with paediatric LGG harbouring a BRAF fusion or rearrangement, or BRAF V600 mutation, who have received one or more prior systemic therapies				
Intervention: Tovorafenib				
PICO 1	Individualised treatment, comprising multiple treatment options[a]: ▪ Vinblastine[b] ▪ Combination of carboplatin and vincristine[b,c] ▪ TPCV combination therapy[b] ▪ Combination of dabrafenib and trametinib[d] ▪ Everolimus[e] ▪ Bevacizumab in combination with chemotherapy[b]	no	No comparator data available	no
PICO 2	Individualised treatment, comprising multiple treatment options[a]: ▪ Vinblastine[f] ▪ Combination of carboplatin and vincristine[f,c] ▪ TPCV combination therapy[f] ▪ Combination of dabrafenib and trametinib[d]	no	No comparator data available	no
PICO 3	Combination of carboplatin and vincristine[c]	no	No comparator data available	no
PICO 4	Vinblastine	no	No comparator data available	no

[1] Patient Population – Intervention(s) – Comparator(s) – Health Outcomes

Population Intervention PICO no.	Comparator	Results submitted [yes/no]	HTD's reason for omission	Data included in the JCA report [yes/no]
Population 2 (BRAF V600E mutation in patients > 1 year)				
Patients > 1 year of age with paediatric LGG harbouring a BRAF V600E mutation, who have received one or more prior systemic therapies				
Intervention: Tovorafenib				
PICO 5	Combination of dabrafenib and trametinib	yes	–	yes
PICO 6	Individualised treatment, comprising multiple treatment options[a]: ▪ Vinblastine ▪ Combination of carboplatin and vincristine[c]	no	No comparator data available	no
Population 3 (BRAF fusion, rearrangement, or V600 [non-E] mutation)				
Patients 6 months of age and older with paediatric LGG harbouring a BRAF fusion or rearrangement, or BRAF V600 (non-E) mutation, who have received one or more prior systemic therapies				
Intervention: Tovorafenib				
PICO 7	Trametinib	yes	–	no[g]
PICO 8	Individualised treatment, comprising multiple treatment options[a] ▪ Vinblastine ▪ Combination of carboplatin and vincristine[c]	no	No comparator data available	no

a: The choice of treatment depends on a patient's individual characteristics.
b: Chemotherapy-based treatment regimens are a treatment option for patients covered by the full claimed indication with the following characteristics: a BRAF alteration other than a BRAF V600E mutation or with a BRAF V600E mutation and insufficient response to dabrafenib in combination with trametinib; aged 6-11 months and from 18 years; no SEGA associated with TSC.
c: A treatment regimen starting with a combination of carboplatin & vincristine may be finished with alternating courses of cisplatin & vincristine with cyclophosphamide & vincristine, due to the risk of carboplatin hypersensitivity.
d: Dabrafenib in combination with trametinib should be used for paediatric patients aged 12 months and older with a BRAF V600E mutation, except for patients who did not respond sufficiently to a first-line therapy with dabrafenib in combination with trametinib.
e: Everolimus is a treatment option for patients with SEGA associated with TSC.
f: Chemotherapy-based treatment regimens are a treatment option for patients covered by the full claimed indication with the following characteristics: a BRAF alteration other than a BRAF V600E mutation or with a BRAF V600E mutation and insufficient response to dabrafenib in combination with trametinib; aged 6-11 months and from 18 years.
g: Results submitted by the HTD to inform PICO 7 are not included in the assessment, due to insufficient information available for the assessment of the study on the comparator (see Section 4.1.1 for a detailed explanation).

BRAF: v-Raf murine sarcoma viral oncogene homologue B; LGG: low-grade glioma; SEGA: subependymal giant cell astrocytoma; TPCV: Thioguanine, procarbazine, lomustine, and vincristine; TSC: tuberous sclerosis complex

Table 11: Assessment scope table of outcomes

Outcome	Notes
Relative effect estimates are required for the following outcomes:	
Safety outcomes according to the Guidance on outcomes for JCA	-
Overall survival (OS)	-
Progression-free survival (PFS)	PFS should be defined as a time-to-event outcome. Rates of PFS at 6- and 12-months (binary outcomes) are also requested. This outcome should be included with 'progression' defined according to both RAPNO and RANO criteria, wherever these data are available for both intervention and comparator in a given comparison.
Objective response, a composite outcome including complete response and partial response.	This outcome should be included with 'response' defined according to both RAPNO and RANO criteria, wherever these data are available for both intervention and comparator in a given comparison.
Health-related quality of life (HrQoL), generic	Measured preferably as PedsQL 4.0 Generic Core Scales and/or SF-36 (generic questionnaires).
Health-related quality of life (HrQoL), disease-specific	Measured preferably as PedsQL 3.0 Cancer Module and/or EORTC QLQ-C30 (disease-specific questionnaires).
Symptoms of the disease, including all of the following: symptoms concerning neurological function; impaired vision, balance, and hearing; motor function; epileptic seizures; severe haemorrhagic events; general and cognitive fatigue.	Measured preferably by a validated instrument.
Symptomatic disease control: a composite outcome, preferably comprising stable symptoms and improved symptoms.	-
Composite outcome comprising of complete response, partial response, or stable disease lasting a minimum of (i) 6 and (ii) 12 months.	This outcome should be included with 'response' and 'stable disease' defined according to both RAPNO and RANO criteria, provided these data are available for both intervention and comparator in a given comparison.
General and cognitive fatigue	Measured by an instrument that can be used to generate QALY-weights. Only required as a separate outcome if fatigue is *not* included within the instrument(s) used to measure symptoms of the disease and/or quality of life.

Outcome	Notes
Descriptive summary data are required for the following outcomes for both intervention and comparator(s) (relative effect estimates not required):	
Duration of response	Descriptive data should include median time-to-event estimates, Kaplan-Meier curves, and landmark 6- and 12-month rates, together with 95% confidence intervals/bands in all cases, reported for both the intervention and comparator. Data should be provided with 'response' defined by both RAPNO and RANO criteria, provided these data are available for both intervention and comparator in a given comparison.
Time-to-response	Descriptive data should include median time-to-event estimates and Kaplan-Meier curves, together with 95% confidence intervals/bands in all cases, reported for both the intervention and comparator. Data should be provided with 'response' defined by both RAPNO and RANO criteria, provided these data are available for both intervention and comparator in a given comparison.
Best overall response, a categorical outcome preferably consisting of complete response, partial response, minor response, stable disease and progressed disease.	Descriptive data should include the number and percentage of subjects achieving each response category, for both intervention and comparator. Data should be provided with 'response' defined by both RAPNO and RANO criteria, provided these data are available for both intervention and comparator in a given comparison.

Table 12: Assessment scope table of subgroup analyses

Note to the HTD:	The results of the following specific subgroup analyses should be provided for all comparisons for which the relevant data are available, if not already included among the subgroup analysis reported in the submission dossier (please see the "HTACG Guidance on filling in the joint clinical assessment (JCA) dossier template – Medicinal products" for further details of the requirements).
Variable	**Notes**
Prior lines of therapy	Preferably categorised as (i) 1 prior line and (ii) 2 or more prior lines.
Age	-

Three different populations were defined in the assessment scope (see Table 10). Population 1 corresponds to the full claimed indication of patients 6 months of age and older with paediatric LGG harbouring a BRAF fusion, rearrangement or V600 mutation who have received one or more prior systemic therapies. Population 2 refers to the subpopulation of patients >1 year of age with a BRAF V600E mutation and Population 3 to the subpopulation of patients with a BRAF fusion, rearrangement or V600 (non-E) mutation. Eight different research questions were specified for these populations in accordance with the PICO framework. PICOs 1 to 4 were defined for Population 1 (full claimed indication), PICOs 5 and 6 for Population 2

(BRAF V600E mutation in patients >1 year), and PICOs 7 and 8 for Population 3 (BRAF fusion, rearrangement or V600 [non-E] mutation).

The HTD provides evidence informing PICOs 5 and 7 and argues that PICOs 1 to 4, 6 and 8 could not be addressed due to the unavailability of data for the specified comparators in the target population. Thus, no comparative effectiveness or safety analyses were provided in the dossier for PICOs 1 to 4, 6 and 8.

PICOs with comparator data

For PICO 5 (subpopulation of patients > 1 year of age with BRAF V600E mutation), the HTD provides an unanchored matching-adjusted indirect comparison (MAIC) of tovorafenib with the combination of dabrafenib and trametinib.

For PICO 7 (subpopulation of patients with BRAF fusion, rearrangement or a V600 (non-E) mutation), the HTD provides an unanchored MAIC of tovorafenib with trametinib. These data are not included in the JCA report due to insufficient information being available for the assessment of the study on the comparator (see Section 4.1.1 for details).

PICOs without comparator data

For PICOs 1, 2, 6 and 8 with an individualised treatment comparator, the HTD describes that no studies were identified in the relapsed/refractory (R/R), paediatric LGG BRAF-altered population that included all the treatments comprising the individualised treatment. The HTD concludes that the lack of such comparator data in the target population precludes the possibility of performing a comparison with tovorafenib for these PICOs.

The HTD did not aim to include studies with an individualised treatment comparator which did not consider all of the treatment options. The HTD's approach to include studies with an individualised treatment comparator, only if the comparator included all treatment options, is not adequate. If comparator data that include only a selection of the requested treatment options had been identified and submitted, the suitability of these data to address the relevant PICOs would then have been a matter of assessment (see [8] for further information). The availability of data on only a relevant selection of treatment options in the comparator study may result in the description of effects only for a subpopulation of the original population of the PICO. However, the assessors did not identify studies on a relevant selection of treatment options for PICOs 1, 2, 6 and 8 (see Section 4.1 for details).

For PICOs 3 and 4, the HTD describes that no studies were identified in the R/R, paediatric LGG BRAF-altered population that investigated treatment with the combination of carboplatin and vincristine or with vinblastine.

For the PICOs with a lack of published data on the comparator (PICOs 1 to 4, 6 and 8), the HTD described an attempt to construct external control arms using patient data acquired outside of clinical studies to allow comparative effectiveness analyses. The HTD performed a feasibility study to identify suitable data sources to construct the external control arms. According to the HTD, one US database initially demonstrated feasibility for use but was deemed unsuitable due to an insufficient number of patients with BRAF alterations and adequate follow-up in the R/R setting. The HTD concluded that finding a fit-for-purpose and reliable comparator data source was not possible. Details on the feasibility study are presented by the HTD in Appendix D.5.2 of the dossier.

For each of the populations defined in the assessment scope, the HTD provides single-arm descriptive outcomes on treatment with tovorafenib from the pivotal study. These data are not discussed in the JCA report because they do not provide information on the relative effects of the medicinal product under assessment against the comparators defined in the assessment scope.

For a list of studies included in the current assessment for results on relative effectiveness or safety, see Section 4.1.

4 Results

The results presented in this section shall follow international standards of evidence-based medicine and take into account, if available, the methodological guidance adopted by the HTACG pursuant to Article 3(7), point (d), of the HTAR. Any deviations shall be described and justified.

The results section provides the findings of the systematic information retrieval, characterises the included studies and presents the results on relative effectiveness and relative safety of the medicinal product under assessment versus the comparators defined in the PICO question(s). Factors which may affect the certainty of the relative effects are identified, taking into account the strengths and limitations of the available evidence.

4.1 Information retrieval

This section shall include:

— a description of the information retrieval performed by the HTD,

— an assessment of the appropriateness of the sources and the search strategies of the HTD.

It shall provide the date of the list of the studies, performed or sponsored by the HTD or by third parties, referred to in Annex I, point (b), of the HTAR, as well as the date of the last searches for the medicinal product and the comparator(s) in bibliographic databases and in study registries and study results registries (clinical trial databases).

Detailed information shall be provided in Appendix B.

The assessment of the appropriateness of the sources and the search strategies demonstrates some deficiencies further detailed in Appendix B.

The studies included in the assessment were compiled using the following information:

Sources provided by the HTD in the dossier:

- list of HTD-sponsored studies on tovorafenib (no status date reported by the HTD),
- bibliographic search for tovorafenib and comparators (last search on 14 July 2025),
- search in study registries/study result databases for tovorafenib and comparators (last search on 14 October 2025),
- search for HTA reports for tovorafenib and comparators (last search on 25 July 2025),
- submission files to the European Medicines Agency (EMA) for tovorafenib (no date reported by the HTD),
- search in patient registries for paediatric LGG (last search on 25 July 2025).

In their own searches, the assessors did not identify additional studies (see Appendix B).

Consistent with the findings of the HTD, neither randomised, controlled trials (RCTs) with a direct comparison of the intervention versus PICO comparators nor studies enabling an anchored indirect comparison of the intervention with PICO comparators via a common comparator were identified by the assessors for any of the PICOs. Thus, studies with either the intervention or comparators for unanchored indirect comparisons were selected. Through this process, the HTD and the assessors identified the following 3 studies:

- the pivotal study on monotherapy with tovorafenib (FIREFLY-1)
- the study by Bouffet et al. 2023 on the combination of dabrafenib and trametinib (hereinafter referred to as dabrafenib + trametinib)
- the TRAM-01 study on monotherapy with trametinib

The 3 studies are referred to as FIREFLY-1, Bouffet 2023 and TRAM-01 in the following text. In the dossier, the HTD provides data from these studies for the different populations according to the assessment scope as described below.

No additional relevant study was identified through the review of completeness for any of the PICOs. However, a number of additional issues relating to information retrieval and study selection were identified. These are described in detail in Appendix B.

The relevance of the studies identified in the information retrieval process for the JCA is described in the following section.

4.1.1 Resulting list of included studies overall and by PICO

This section shall provide in tabular format:
— an overview of all included studies and the associated references for these studies overall and by PICO,
— the list of studies included by the HTD which were excluded within the assessment, with a justification for their exclusion.

The following table lists the studies included in the assessment and the available documentation and identifies which studies are relevant for each of the PICO questions of the assessment.

Table 13: Included studies – list of relevant studies by PICO question

Study reference/ID **Study type** **Study interventions**	**Study for marketing authorisation**	**Sponsored[a] or third-party study of the medicinal product under assessment**	**Available documentation in the submission dossier**
Population 1 (full claimed indication): PICOs 1, 2, 3, 4			
No evidence informing PICOs included			
Population 2 (BRAF V600E mutation in patients > 1 year): PICO 5			
Studies providing indirect evidence: tovorafenib vs. dabrafenib + trametinib			
DAY101-001/PNOC026 (FIREFLY-1[b]) *Non-randomised, non-comparative, open-label, multi-centre, multi-arm study* Assessed intervention: tovorafenib	yes[c]	sponsored[d]	▪ CSR[e]: [9] ▪ Registry entry: ▫ Clinicaltrials.gov: NCT04775485 [10] ▫ EudraCT: 2020-003657-30 [11] ▪ Publication or other reference: ▫ Kilburn et al., 2024 [12]
CTMT212X2101 (Bouffet 2023[b]) *Non-randomised, non-comparative, open-label, multi-centre, multi-arm study* Assessed intervention: dabrafenib + trametinib	no	not sponsored	▪ CSR: – ▪ Registry entry: ▫ Clinicaltrials.gov: NCT02124772 [13] ▫ EudraCT: 2013-003596-35 [14] ▪ Publication or other reference: ▫ Bouffet 2023[f] [15]
Population 2 (BRAF V600E mutation in patients > 1 year): PICOs 6			
No evidence informing PICO included			
Population 3 (BRAF fusion, rearrangement, or V600 [non-E] mutation): PICO 7, 8			
No evidence informing PICOs included[g]			
a: Study sponsored by the HTD or in which the HTD participated financially in some other way. b: In the following tables, the study is referred to in this abbreviated form. c: FIREFLY-1 is the pivotal study for the EMA marketing authorisation that is related to this JCA. d: FIREFLY-1 was sponsored by Day One Biopharmaceuticals, Inc. e: The 2-year data cut for FIREFLY-1 (used in this submission) is available from the CSR; a post-hoc analysis based on this data cut was presented at ASCO 2025 [16]. f: Bouffet 2023 is a multi-arm study that included several study parts and cohorts of patients. The combination of dabrafenib and trametinib and trametinib monotherapy were each assessed in two of these parts. Some efficacy analyses presented in the manuscript were performed on regrouped patient cohorts based on the interventions received. g: Results informing PICO 7 submitted by the HTD are not included in the JCA report due to insufficient information available for the assessment of the study on the comparator (see subsequent text for details)			
ASCO: American Society of Clinical Oncology; CSR: clinical study report; EMA: European Medicines Agency; EudraCT: European Union Drug Regulating Authorities Clinical Trials Database; HTD: health technology developer; ID: identification; JCA: Joint Clinical Assessment; NCT: National Clinical Trial			

The following table lists studies that were included by the HTD in the submission dossier but are not included in the assessment.

Table 14: List of excluded studies – studies included by the HTD but not used in the assessment

Study reference/ID	Reason for exclusion
TRAM-01 [17-20]	Insufficient information available to assess study methods and results (see subsequent text for details)

For the subpopulation of patients with a BRAF fusion, rearrangement or V600 (non-E) mutation (Population 3), the HTD provides an unanchored MAIC of tovorafenib with trametinib (PICO 7). For tovorafenib, this unanchored MAIC is based on a subpopulation from FIREFLY-1; for trametinib, it is based on results of TRAM-01. For this study, a design publication [17] and abstracts reporting results on 2 different data cut-offs are available (Perreault 2021: 12 February 2021 [19], Perreault 2022: 31 January 2022 [18]). Thus, the reported results for this study are based on abstracts only. The comparison based on FIREFLY-1 and TRAM-01 is not included in the JCA report due to insufficient information available for the assessment of TRAM-01.

For studies to be included in the JCA, sufficient documentation is required to allow for the assessment of the study methods and results. Therefore, studies for which only limited summary data are available (e.g., from a conference abstract, presentation or poster) should be excluded [21]. Contrary to this, the HTD describes in Section 4.1 of the dossier that it was decided to include results of studies presented only in conference abstracts in the HTD's assessment due to the rarity of the disease and the paucity of evidence. Regardless of the rarity of the disease, the information on the reported results of TRAM-01 is not sufficient to include the study in the JCA report. In particular, it is unclear whether the results presented in the abstracts were generated using the methods reported in the design publication.

In addition, the information on patient characteristics is insufficient to show that the reported results refer to the subpopulation of patients within the scope of PICO 7 (i.e., patients with a BRAF fusion, rearrangement or V600 [non-E] mutation). For the data cut-off of 31 January 2022 included in the HTD's assessment, age is the only patient characteristic reported and therefore the only variable for which adjustment was possible. For this reason, including the results in an unanchored MAIC of tovorafenib with trametinib for PICO 7 is also not appropriate. Thus, the information on patient characteristics is also considered insufficient for inclusion, irrespective of the source document reporting the results.

PICOs without comparator data

For Population 1 (PICOs 1 to 4), Population 2 (PICO 6) and Population 3 (PICO 8), the HTD provides single-arm descriptive outcomes on treatment with tovorafenib from the pivotal study FIREFLY-1. As described in Section 3, these data are not discussed in the JCA report.

Detailed references to the sections of the dossier presenting the single-arm data are provided in Sections 4.3.1 and 4.3.3.

4.2 Characteristics of included studies and RoB

4.2.1 Included studies

This section shall provide for the studies included in the assessment:

— information on the study design (e.g. on randomisation, blinding, or parallel observation studies, and the key inclusion and exclusion criteria),

— information on enrolled study populations (e.g. diagnosis, general severity of medical condition, and line of therapy),

— characteristics of the study interventions,

— information on the course of the study (e.g. planned and actual follow-up times per outcome),

— information on the study duration.

Table 15 and Table 16 describe the studies included in the assessment.

Table 15: Characteristics of the included studies

Study reference/ID	Study type and design	Study population	Study arms (number of randomised/ included patients)	Study duration, data cut-off(s) and locations	Study endpoints[a]
FIREFLY-1	Phase 2, non-randomised, non-comparative, open-label, multi-centre, multi-arm study	**Arm 1 (pivotal, LGG)** Patients aged 6 months to 25 years, inclusive, with R/R[b] LGG harbouring an activating BRAF alteration, including BRAF V600 mutations and KIAA1549:BRAF fusions. **Arm 2 (expansion cohort, LGG)** Patients aged 6 months to 25 years, inclusive, with R/R[b] LGG harbouring an activating or expected to be activating RAF alteration (e.g., BRAF or CRAF/RAF1 fusion or BRAF V600 mutations). **Arm 3c (advanced solid tumour)** Patients aged 6 months to 25 years, inclusive, with advanced solid tumours harbouring an activating or expected to be activating RAF fusion.	**Arm 1** ▪ Tovorafenib monotherapy: N = 77 **Arm 2** ▪ Tovorafenib monotherapy: N = 60 **Arm 3**[c] ▪ Tovorafenib monotherapy: N = 6	**Study duration:** ▪ Screening: 4 weeks ▪ Treatment & follow-up: Ongoing: Patients in Arms 1 and 2 entered a 36-month long-term extension phase after the last patient enrolled in Arm 1 completed 26 cycles of treatment[d]. **Period of study:** ▪ 4/2021 – ongoing (estimated completion: 5/2027) **Data cut-offs**[e]**:** ▪ Primary (9-month data cut-off): 22 Dec 2022 ▪ 15-month data cut-off: 5 June 2023 (reported in Kilburn et al., 2024 [12]) ▪ 2-year data cut-off: 10 May 2024 ▪ 3-year data cut-off: 21 May 2025 ▪ Final analysis: Expected Q2 2027 **Locations:** ▪ 32 centres ▫ Asia: 6 ▫ North America: 14 ▫ Europe: 7 ▫ Oceania: 5	**Primary:** ▪ Arm 1: ORR ▪ Arm 2: Safety and tolerability, including evaluation of AEs and laboratory abnormalities **Key secondary/other**[f]**:** ▪ Arm 1: Safety and tolerability, including evaluation of AEs and laboratory abnormalities ▪ Arm 2: ORR ▪ PFS ▪ DOR ▪ TTR ▪ Composite outcome comprising CR, PR or SD ▪ OS ▪ Visual acuity (as measured using BCVA) ▪ HRQoL (PedsQL Core Generic Scales, PedsQL Cancer Module, PROMIS)

Study reference/ID	Study type and design	Study population	Study arms (number of randomised/ included patients)	Study duration, data cut-off(s) and locations	Study endpoints[a]
Bouffet 2023	Phase 1/2, non-randomised, non-comparative, open-label, multi-centre, multi-arm study	**Part A** Children and adolescents (aged ≥ 1 month to < 18 years) with R/R[g] histologically confirmed solid tumours **Part B** Children and adolescents (aged ≥ 1 month to < 18 years) with the following R/R[g] tumours: ▪ Histologically confirmed neuroblastoma with MAPK/RAS/MEK pathway activation ▪ Gliomas or primary brain tumours with BRAF fusion/duplication or NF-1 gliomas not eligible for the NF-1 with PN cohort ▪ NF-1 patients with progressive or morbid PN that is not suitable for resection ▪ BRAF V600-mutant solid tumours	**Part A** ▪ Trametinib monotherapy): N = 50 **Part B** ▪ Trametinib monotherapy: N = 41 **Part C** ▪ Dabrafenib + trametinib: N = 18 **Part D** ▪ Dabrafenib + trametinib: N = 30 **Total (all tumours):** ▪ Trametinib monotherapy: N = 91 ▪ Dabrafenib + trametinib: N = 48 **Total (BRAF V600E-mutant LGG only):** ▪ Trametinib monotherapy: N = 13 ▪ Dabrafenib + trametinib: N = 36[h]	**Study duration:** ▪ Screening: 14 days ▪ Treatment: until disease progression, death or unacceptable toxicity[i] ▪ Follow-up: ongoing in a separate long-term extension study[j] **Period of study:** 1/2015–12/2020 **Data cut-off:** Not reported **Location:** ▪ 16 study centres ▫ Europe: 4 ▫ North America: 11 ▫ Australia: 1	**Primary:** ▪ Safety and tolerability, including evaluation of AEs and laboratory abnormalities, to determine the trametinib dose for paediatric patients achieving similar exposure to the recommended adult dose **Key secondary/other**[f]: ▪ Safety and tolerability, including evaluation of AEs and laboratory abnormalities, to determine the dabrafenib dose for paediatric patients achieving similar exposure to the recommended adult dose ▪ ORR[k] ▪ Composite outcome comprising CR, PR or SD[k] ▪ PFS[k] ▪ DOR[k] ▪ BOR[k]

Study reference/ID	Study type and design	Study population	Study arms (number of randomised/ included patients)	Study duration, data cut-off(s) and locations	Study endpoints[a]
		Part C Children and adolescents (aged ≥ 12 months to < 18 years) with R/R[f] tumours harbouring the BRAF V600 mutation **Part D** Children and adolescents (aged ≥ 12 months to < 18 years) with the following recurrent, refractory or unresectable tumours: ▪ BRAF V600-mutant LGG ▪ BRAF V600-mutant LCH			

a: Study endpoints are only listed if included in at least 1 PICO.
b: Patients must have received at least 1 line of prior systemic therapy and have documented evidence of radiographic progression.
c: This study part/arm is not relevant for the JCA.
d: During the long-term extension phase, patients could either continue on tovorafenib or opt to enter a 'drug holiday' treatment discontinuation period. Patients on a 'drug holiday' could be re-treated with tovorafenib if there was clinical or radiographic evidence of disease progression as documented by the investigator.
e: information on prespecification of data cut-offs is not available from the dossier or the CSR; in the dossier, only data up to the 2-year data cut-off are available for the assessment. The date of the 3-year data cut-off refers to the date reported in the dossier. In the HTD's response to a request for additional information by the assessors, the 6 June 2025 is reported as the data cut-off date for the 3-year follow-up.
f: The secondary endpoint listing does not focus specifically on key secondary endpoints controlled for multiplicity, in the absence of such endpoints in all the studies.
g: Patients must have a disease that is relapsed/refractory to all potentially curative standard treatment regimens or must have a current disease for which there is no known curative therapy or therapy proven to prolong survival with an acceptable quality of life. Patients with prior treatment with dabrafenib, trametinib or other BRAF/MEK/ERK inhibitors were only allowed to be enrolled in Parts C or D if they had documented prior benefit from the received treatment as assessed by the investigator.

Study reference/ID	Study type and design	Study population	Study arms (number of randomised/ included patients)	Study duration, data cut-off(s) and locations	Study endpoints[a]
h: Two patients with BRAF V600E-mutant HGG were also included in the LGG cohort. i: Initially the study was planned to be considered completed when the last patient enrolled had been in the study for a minimum of 12 months AND, for those patients still benefiting from treatment, the rollover protocols are open to enrol paediatric patients. With Amendment 8 of the study protocol (January 2020), the minimum treatment duration was reduced from 12 months to 6 months. j. Paediatric Long-Term Follow-up and Rollover Study (NCT03975829) [22]. k: Preliminary assessment of anti-tumour activity of trametinib monotherapy and dabrafenib + trametinib combination therapy. The ORR and a composite outcome comprising CR, PR or SD, PFS, and DOR are presented in Bouffet 2023 [15]; only the ORR and a composite outcome comprising CR, PR or SD are listed as anti-tumour activity endpoints in publicly available study documentation. According to the statistical analysis plan, only the ORR and BOR are prespecified.					
AE: adverse event; BCVA: best corrected visual acuity; BRAF: v-Raf murine sarcoma viral oncogene homologue B; BOR: best overall response; CR: complete response; CRAF/RAF1: proto-oncogene serine/threonine-protein kinase; DOR: duration of response; EMA: European Medicines Agency; ERK: extracellular signal-regulated kinase; HGG: high-grade glioma; HRQoL: health-related quality of life; ID: identification; INV: investigator; IRC: independent review committee; JCA: joint clinical assessment; LCH: Langerhans cell histiocytosis; LGG: low-grade glioma; MAPK: mitogen-activated protein kinase; MR: minor response; NF-1: neurofibromatosis 1; ORR: objective response rate; OS: overall survival; PedsQL: Pediatric Quality of Life Inventory; PFS: progression-free survival; PK: pharmacokinetics; PN: plexiform neurofibroma; PR: partial response; PROMIS: Patient-Reported Outcomes Measurement Information System; Q: quarter; RAF: rapidly accelerated fibrosarcoma; RANO-HGG: Response Assessment in Neuro-Oncology criteria for high-grade gliomas; RAPNO: Response Assessment in Paediatric Neuro-Oncology; R/R: relapsed/refractory; SAE: serious adverse event; SD: stable disease; TTP: time to progression; TTR: time to response					

Table 16: Characterisation of the interventions of included studies

Study reference/ID	Study intervention	Study comparator
FIREFLY-1	▪ Oral tovorafenib at the RP2D of 420 mg/m^2 QW, according to the patient's baseline BSA (Days 1, 8, 15, and 22 of a 28-day cycle). BSA was determined by the Mösteller Formula [V(height (cm) × weight (kg)/3600)] and was calculated on Day 1 of each cycle to confirm dose. ▪ The maximum dose was to be no higher than the adult RP2D of 600 mg QW	NA (no comparator arm)
	Premedication required ▪ At least 1 line of prior systemic therapy **Prohibited premedication treatment** ▪ Major surgery within 14 days prior to initiation of tovorafenib **Permitted concomitant treatment** ▪ Steroids for tumour-associated symptoms, must be on a stable dose for 14 days prior to initiation of tovorafenib **Prohibited concomitant treatments** ▪ Strong cytochrome P450 2C8 inhibitors or inducers within 14 days before initiation of tovorafenib ▪ Substrates of breast cancer resistance protein with a narrow therapeutic index within 14 days before initiation of tovorafenib	
Bouffet 2023	**Trametinib monotherapy**[a] ▪ Part A: Oral trametinib at a dose of 0.0125, 0.025, 0.032 or 0.04 mg/kg/day ▪ Part B: Oral trametinib at a dose of 0.025 or 0.032 mg/kg/day ▪ A dosing nomogram based on weight and dose level was used to prescribe trametinib to minimise inter-patient dosing variability **Dabrafenib + trametinib** ▪ Part C: ▫ Oral dabrafenib at a dose of 50% RP2D: < 12 years: 2.63 mg/kg/day and ≥ 12 years: 2.25 mg/kg/day or 100% RP2D: < 12 years: 5.25 mg/kg/day and ≥ 12 years: 4.5 mg/kg/day ▫ Oral trametinib at a dose of 0.025 or 0.032 mg/kg/day (0.032 mg/kg/day dose only evaluated with 100% RP2D of dabrafenib) ▫ A dosing nomogram based on weight and dose level was used to prescribe dabrafenib/trametinib to minimise inter-patient dosing variability ▪ Part D: ▫ Oral dabrafenib at a dose of 5.25 mg/kg/day < 12 years and 4.5 mg/kg/day ≥ 12 years ▫ Oral trametinib at a dose of 0.032 mg/kg/day < 6 years and 0.025 mg/kg/day ≥ 6 years	NA (the monotherapy and combination therapy groups were not statistically compared in any way)

Study reference/ID	Study intervention	Study comparator
	▫ A dosing nomogram based on weight and dose level was used to prescribe dabrafenib/trametinib to minimise inter-patient dosing variability	
	Premedication required ▪ Must have had potentially curative frontline therapy (surgery, radiation, chemotherapy or a combination thereof) **Prohibited premedication treatment** ▪ Part C & D: MAPK inhibitors (including BRAF and MEK inhibitors), except in patients who have previously responded to treatment **Permitted concomitant treatment** ▪ Birth control treatments **Prohibited concomitant treatments** ▪ Medication for treatment of left ventricular systolic dysfunction ▪ Part C & D: ▫ Other anti-cancer treatments ▫ Other investigational treatments ▫ Anti-retroviral treatments ▫ Herbal remedies ▫ Strong inhibitors of cytochrome P450 2C8 and 3A4 (only in special circumstances)	
a: This study part/arm is not relevant for the JCA.		
BRAF: v-Raf murine sarcoma viral oncogene homologue B; BSA: body surface area; ID: identification; MAPK: mitogen-activated protein kinase; NA: not applicable; QW: every week; RP2D: recommended phase 2 dose; SD: stable disease		

4.2.1.1 Study design – FIREFLY-1

FIREFLY-1 is an ongoing, phase-2, open-label, multi-centre, multi-arm, non-comparative, non-randomised study evaluating tovorafenib monotherapy for the treatment of patients with LGG and advanced solid tumours. The study is pivotal for the EMA submission and consists of 3 study arms including patients aged 6 months to 25 years, inclusive, with:

- Arm 1 (LGG): relapsed or progressive LGG harbouring an activating BRAF alteration, including BRAF V600 mutations and KIAA1549:BRAF fusions,
- Arm 2 (LGG, expansion cohort): relapsed or progressive LGG harbouring an activating or expected to be activating BRAF mutation or rapidly accelerated fibrosarcoma (RAF) fusion, e.g., BRAF or proto-oncogene serine/threonine-protein kinase (CRAF/RAF1) fusion or BRAF V600 mutations,
- Arm 3 (advanced solid tumour): locally advanced or metastatic solid tumour harbouring an activating or expected to be activating RAF fusion, e.g., BRAF or CRAF/RAF1 fusion.

Arm 3 does not include any patients within the scope of the assessment at the time of the 2-year data cut-off (according to the clinical study report [CSR], 6 patients with non-LGG malignancies were enrolled in Arm 3 as of 10 May 2024). Thus, Arm 3 is not relevant for the JCA and hence not described in more detail.

Patients in Arms 1 and 2 must have received at least 1 line of prior systemic therapy and have had documented evidence of radiographic progression. At enrolment, patients must have demonstrated adequate renal and hepatic function and a Karnofsky (aged ≥ 16 years) or Lansky (aged < 16 years) performance score of ≥ 50.

A total of 137 patients were included in Arm 1 (n = 77) and Arm 2 (n = 60) of the study. According to the study protocol, oral tovorafenib was administered depending on the patient's baseline body surface area (BSA) at a starting dose of 420 mg/m^2 at Days 1, 8, 15 and 22 of 28-day cycles. The maximum dose was to be no higher than 600 mg weekly irrespective of the patient's BSA. According to the Summary of Product Characteristics (SmPC) [23,24], the intended, recommended dose based on BSA is 380 mg/m^2 orally once weekly and the maximum, recommended dose is 600 mg orally once weekly. It is further mentioned in the SmPC that the tovorafenib doses administered in FIREFLY-1 (Arm 1) were actually between 290 mg/m^2 to 476 mg/m^2. Thus, the actual dose used in FIREFLY-1 varied between the individual patients and some patients received higher or lower doses compared to the intended, recommended dose.

In FIREFLY-1, patients continue treatment with tovorafenib until radiographic evidence of disease progression as determined by the treating investigator, unacceptable toxicity, decision to enter a "drug holiday" discontinuation period (during the long-term extension phase), patient withdrawal of consent or death. Patients in Arms 1 and 2 enter the 36-month long-term extension phase after the last patient enrolled in Arm 1 completes 26 cycles of treatment (approximately 24 months). There are no restrictions regarding the administration of follow-up therapies.

The primary endpoint is the objective response rate (ORR) in Arm 1 and safety and tolerability in Arm 2, including the evaluation of adverse events (AEs) and laboratory abnormalities. Other endpoints included OS, progression-free survival (PFS), duration of response (DOR), time to response (TTR), health-related quality of life, visual acuity as well as a composite outcome comprising complete response (CR), partial response (PR) or stable disease (SD).

Data cut-offs

The dossier contains CSRs for 3 data cut-offs:

- 9-month data cut-off: 22 December 2022
- 15-month data cut-off: 5 June 2023

- 2-year data cut-off: 10 June 2024

Furthermore, the HTD indicates in the dossier that a 3-year data cut-off was performed (21 May 2025) and a final analysis (data cut-off expected Q2 2027) is planned. The HTD provides results from the 2-year data cut in the dossier. No data cut-offs were prespecified in the study protocol.

On 21 November 2025, the HTD informed that the sponsor of FIREFLY-1 had recently shared clinical data from the 3-year data cut (see Appendix D.1.2). The cut-off date mentioned by the HTD (6 June 2025) deviates from the information provided in the dossier (21 May 2025). The HTD did not include clinical data from this data cut in the submission. As for the 2-year data cut, the HTD describes that the 3-year data cut was not prespecified, and it was not part of any formal commitment to the Food and Drug Administration.

Relevant subpopulation to inform the current assessment

Population 2 is the only population, with PICO 5 being the only PICO, with data informing relative effectiveness and safety in the current assessment. The subpopulation relevant for PICO 5 comprises only patients > 1 year of age with LGG harbouring a BRAF V600E mutation. In FIREFLY-1, a subpopulation of 13 patients in Arm 1 (17%) and 9 patients in Arm 2 (15%) fulfils these criteria. The vast majority of the other patients with LGG included in the study had a BRAF fusion, most commonly the KIAA1549:BRAF fusion (56 patients [73%] in Arm 1 and 45 patients [75%] in Arm 2). These patients are the subject of Population 1 (PICOs 1 to 4), and Population 3 (PICOs 7 and 8) of the assessment scope. However, no data informing these PICOs are included in the current assessment (see Section 4.1.1 for details). For details on the data submitted by the HTD for the subpopulation of FIREFLY-1 to address PICO 5, see Section 4.3.2.

4.2.1.2 Study design – Bouffet 2023

Bouffet 2023 is a phase I/IIa, non-randomised, open-label, multi-centre, non-comparative study in paediatric patients with refractory or recurrent tumours. It is divided into 4 parts: Parts A and B evaluate trametinib monotherapy, while Parts C and D evaluate a combination therapy of dabrafenib + trametinib.

Parts A and B may comprise patients which are within the assessment scope of PICO 7 (where trametinib monotherapy was the comparator). However, no data for patients with the corresponding genetic alterations of population 3 (BRAF fusion, rearrangement or V600 [non-E] mutation) are available from these parts of the study. Thus, these parts are not relevant for the JCA and hence not described in more detail.

Parts C and D enrolled patients aged 1 to < 18 years. Part C included patients with recurrent or refractory malignant solid tumours harbouring a BRAF V600 mutation and aimed to

determine the safety, tolerability and preliminary activity of the recommended phase 2 dose (RP2D) of trametinib in combination with a limited dose escalation of dabrafenib (50% of paediatric RP2D and 100% RP2D). Part D evaluates dabrafenib + trametinib in 2 disease-specific cohorts in paediatric populations with BRAF V600-mutant relapsed/refractory LGG or Langerhans cell histiocytosis (LCH). Patients were not allowed to enrol in more than one part of the study.

A total of 18 patients were included in Part C, mostly patients with BRAF V600E LGG (n = 14) but also with BRAF V600E HGG (n = 2) and BRAF V600E LCH (n = 2). Part D included 30 patients (n = 20 with BRAF V600E LGG, n = 10 with BRAF V600E LCH).

The administration of trametinib and dabrafenib differed from the recommended dosing in the respective SmPCs of trametinib [7] and dabrafenib [6]. According to the SmPCs, the approved dosage for both active substances is based on body weight. In contrast, in Bouffet 2023, dosages were based on body weight and age. The dosages varied between individual patients depending on which cohort of the study they were enrolled in and based on their age. In Parts C and D of the study, trametinib was administered at a dose that was determined as the RP2D based on Part A of the study (0.025 mg/kg/day for patients ≥ 6 years or 0.032 mg/kg/day for patients < 6 years). Dabrafenib was administered at a dose of 5.25 mg/kg/day for patients < 12 years and 4.5 mg/kg/day for patients ≥ 12 years. Furthermore, a cohort in Part C of the study only received 50% of the RP2D of dabrafenib, but this applied to 3 patients only. The dabrafenib dose was divided into 2 equal daily doses, and trametinib was administered once daily, corresponding to the recommendation in the SmPCs. Age for daily dose determination in any period during the study was based on the age at enrolment, and the dose was to remain constant throughout the study. There were no restrictions regarding the administration of follow-up therapies.

Treatment was administered until disease progression, death or unacceptable toxicity. Patients were observed for up to 2 years after treatment discontinuation or the end of the study. The end of the study was defined as the last patient's last visit. The study was considered completed when the last patient enrolled had been in the study for a minimum of 6 months or prematurely discontinued the study and, for patients still benefiting from treatment, the rollover protocols for the separate extension study [22] were open to enrol paediatric patients for long-term follow-up.

The primary endpoint of the study was safety and tolerability, including the evaluation of AEs and laboratory abnormalities. Other endpoints include ORR, best overall response (BOR), PFS, DOR and a composite outcome comprising CR, PR or SD.

Data cut-offs

Data cut-off dates were not reported for Bouffet 2023.

Relevant subpopulation to inform the current assessment

A subpopulation from Bouffet 2023 was used for an indirect comparison in PICO 5. The population relevant for PICO 5 comprises patients > 1 year of age with LGG harbouring a BRAF V600E mutation. In Bouffet 2023, results are reported for a cohort of patients with BRAF V600E-mutant LGG treated with dabrafenib + trametinib from Parts C and D of the study (n = 36; including 2 patients [5.6%] with BRAF V600E-mutant HGG).

4.2.1.3 Tumour assessment criteria used in FIREFLY-1 and Bouffet 2023

In FIREFLY-1 and Bouffet 2023, tumour response was assessed using different sets of criteria, among these:

- the original set of criteria developed by a RANO working group in 2010: RANO-HGG [25],
- a version specifically adopted by a RANO working group in 2017 for LGG: RANO-LGG [26] and
- a version specifically adopted by a Response Assessment in RAPNO working group in 2020 for paediatric LGG: RAPNO [27].

These 3 standardised sets of criteria were developed to improve and standardise tumour response assessment in clinical studies including patients with HGG and subsequently also (paediatric) LGG. However, when the studies FIREFLY-1 and Bouffet 2023 were conducted, there was no global consensus on imaging standards or the appropriate method to measure tumour response using magnetic resonance imaging (MRI) across paediatric LGG studies. Thus, the criteria used differed between the studies.

In the following sections, the criteria and their specific characteristics are described in detail. Subsequently, the use of the criteria in FIREFLY -1 and Bouffet 2023 and the comparability of the definitions used in the studies are discussed in detail.

RANO-HGG

RANO-HGG criteria for response assessment in high-grade gliomas distinguish between four categories of tumour response: CR, PR, SD and progressive disease (PD). A CR requires a complete disappearance of all enhancing measurable and non-measurable disease sustained for at least 4 weeks. Furthermore, patients must be off corticosteroids (or on physiologic replacement doses only) and their clinical status has to be stable or improved. The primary imaging focus is on contrast-enhanced T1-weighted (T1) imaging. Additionally, the RANO-HGG criteria consider qualitative alterations in non-enhancing lesions, depicted on the MRI T2-weighted and fluid-attenuated inversion recovery (FLAIR) image sequences. For a CR, PR or SD, patients must show stable or improved non-enhancing (T2/FLAIR) lesions [25]. Because of the difficulties in quantifying non-enhancing disease progression, the RANO-HGG criteria suggest that any qualitative increase in non-enhancing disease constitutes progression. This

lack of a quantifiable measure of non-enhancing progression is seen as one limitation of the RANO-HGG criteria [28].

RANO-LGG

LGGs have been considered historically to be slower growing and radiographically distinct from HGGs. They are characterized by minimal, if any, evidence of contrast enhancement and nonspecific T2/FLAIR hyperintense lesions with often indistinct margins [28]. The RANO working group therefore proposed response criteria specific for LGG that are similar to those for HGG but measure T2/FLAIR rather than contrast enhancement as these tumours rarely enhance. RANO-LGG utilise a percent change in T2/FLAIR signal rather than contrast enhancement for determination of response and progression (unlike RANO-HGG). To have a CR according to RANO LGG, all non-enhancing (T2/FLAIR) lesions must disappear. For a PR, a decrease in T2/FLAIR tumour of ≥ 50%, sustained for a minimum of 4 weeks is required. As responses are often relatively modest, the new tumour response category minor response (MR) was introduced. MR is characterized by a decrease in T2/FLAIR tumour of 25% to 50%. As with RANO-HGG, corticosteroid use and clinical status are considered in the determination of response and progression. Additionally, the RANO working group notes that, while T2/FLAIR provides the clearest and most reproducible definition of LGG, distinguishing tumour from radiation-induced changes, postsurgical changes, demyelination, ischaemic injury and other comorbid events can be difficult [26]. As many of these features are present in brain tumours, particularly those that have undergone multiple lines of therapy, a CR to treatment of a non-enhancing tumour is not expected, as residual T2/FLAIR changes are likely to remain present [29]. Therefore, compared to tumour assessment based on RANO-HGG criteria, achieving a CR according to RANO-LGG criteria is more challenging. However, the introduction of the new response category, MR, allows even small reductions in tumour size to be recorded more sensitively.

RAPNO

More recently, the RAPNO working group developed specific recommendations for the tumour assessment of patients with paediatric LGG. According to the RAPNO group, T2-weighted and T2/FLAIR imaging appear to be the best measures of tumour size. Furthermore, postcontrast imaging (T1-weighted) is described as necessary for paediatric LGG assessment and as part of evaluating response and only in the rare situation that a paediatric LGG is completely non-enhancing, imaging without contrast can be considered in follow-up surveillance. Besides CR, PR, MR, SD and PD, RAPNO criteria include major response as a sixth tumour response category. To have a CR, as is the case with RANO LGG criteria, all non-enhancing (T2/FLAIR) lesions must disappear. A major response requires a 75% or greater reduction in the target lesion but insufficient response to qualify as a CR. The cut-off values for the other response categories are comparable to those of the RANO-LGG criteria. For all

response categories (except for PD), patients should be clinically stable or have improved on physical examination and functional or neurological assessment [27].

Criteria assessed in FIREFLY-1

In FIREFLY-1 tumour response was assessed by either an independent review committee (IRC) and/or the treating investigator (INV) using three different response assessments:

- RANO-HGG (IRC, INV)
- RANO-LGG (IRC)
- RAPNO (IRC)

Different analysis sets were predefined in FIREFLY-1 using specific criteria for the tumour response assessment:

- Full analysis set: The full analysis set (FAS) included all patients who were enrolled in FIREFLY-1, received at least one dose of study treatment and had measurable disease at baseline per RANO-HGG criteria, as determined by the IRC.
- RAPNO evaluable analysis set: The RAPNO evaluable analysis set included all patients enrolled in FIREFLY-1, who received at least one dose of study treatment and had measurable disease at baseline per RAPNO criteria, as determined by the IRC.
- RANO-LGG evaluable analysis set: The RANO-LGG evaluable analysis set included all patients enrolled in FIREFLY-1, who received at least one dose of study treatment and had measurable disease at baseline per RANO-LGG criteria, as determined by the IRC.

Criteria assessed in Bouffet 2023

According to the study protocol, in Bouffet 2023 tumour response was assessed by either an independent review committee (IRC) and/or the treating investigator (INV) using two different response assessments:

- RANO-HGG (IRC, INV)
- Modified RANO-LGG (IRC) (without MR category)

The additional response assessment with RANO-LGG criteria by IRC was implemented with a protocol amendment (9th amendment dated 21 August 2020) and the response category of MR was omitted. Therefore, patients that were meeting RANO-LGG-criteria for an MR were considered to have SD in Bouffet 2023.

Comparability of the tumour assessment criteria used in FIREFLY-1 and Bouffet 2023

Both studies, FIREFLY-1 and Bouffet 2023, used RANO-LGG and RANO-HGG criteria for assessing the tumour response. However, Bouffet 2023 used modified RANO-LGG criteria

omitting the MR category (including these cases in the SD category). This difference must be considered for the analyses on relative effectiveness of the respective treatments. Similarity of outcome definitions from both studies should be ensured for the outcomes based on tumour responses assessment criteria.

4.2.1.4 Information on the course of included studies – planned follow up times

Table 17 shows the planned duration of follow-up in FIREFLY-1 and Bouffet 2023 for the outcomes with data on comparative effectiveness and safety (ORR, PFS and safety outcomes).

Table 17: Information on the course of included studies – planned follow up times

Comparison **Study reference/ID** **Outcome**	**Planned follow-up**
Tovorafenib monotherapy vs. dabrafenib + trametinib	
FIREFLY-1	
ORR	Until study completion/withdrawal from the study, the start of any subsequent anti-cancer therapy, disease progression/recurrence or death
PFS	Until disease progression, death or the start of any subsequent anti-cancer therapy
Safety	Until 30 days after last study drug dose[a]
Bouffet 2023	
ORR	Until 30 days after last study drug dose[b]
PFS	Until 30 days after last study drug dose[b]
Safety	Until 30 days after last study drug dose[b]
a. In the dossier, the HTD describes a planned follow up until death or end of data collection which, according to the study protocol, only applies to certain AEs considered to be related to the study treatment. Data on safety outcomes for all treatment-emergent AEs were only collected until 30 days after the last study drug dose. b. In the dossier, the HTD describes a planned follow up until study completion/withdrawal from the study. According to the study protocol, a patient was considered to have completed the study if they discontinued study treatment and completed a post-treatment follow-up visit or died while receiving study treatment.	
AE: adverse event; ID: identification; JCA: joint clinical assessment; ORR: objective response rate; PFS: progression-free survival	

4.2.1.5 Information on the course of included studies – actual treatment duration and observation periods

Table 18 describes the actual treatment duration and observation periods. The HTD presents this information only for those outcomes where information is available for at least one included study. For FIREFLY-1, information on the full study population in Arm 1 and Arms 1 + 2 are presented, whereas for Bouffet 2023, information is limited to the relevant subpopulation for PICO 5. Information on the actual treatment duration and observation

periods from FIREFLY-1 for the relevant subpopulation for PICO 5 are reported in Appendix C.1.

Table 18: Information on the course of included studies – actual treatment duration and observation periods

Study reference/ID[a] **Outcome category**	**FIREFLY-1[b]**		**Bouffet 2023**
Intervention	**Tovorafenib** **N = 77 (Arm 1)**	**Tovorafenib** **N = 137 (Arm 1 + 2)**	**Dabrafenib + trametinib** **N = 36**
Treatment duration [days/months]			
Median [min; max]	722.0 [22; 975] days	638.0 [22; 975] days	24 (range: 2.1 -52.5) months
Mean (SD)	581.0 (259.31) days	559.0 (235.25) days	NR
Study observation period [months]			
Median [min; max]	28.1 [1.1; 36.5]	24.4 [1.1; 36.5]	NR
Mean (SD)	27.45 (6.79)	24.49 (6.49)	NR
OS observation period [months][c]			
Median [min; max]	27.8[c] [1.1; 36.2]	NR	NR
Mean (SD)	27.15[c] (6.11)	NR	NR
PFS observation period [months]			
Median [min; max]	22.34[d] [16.53; 25.07]	NR	NR
Mean (SD)	NR	NR	NR
ORR observation period [months]			
Median [min; max]	NR	NR	NR
Mean (SD)	NR	NR	NR
Safety observation period [months]			
Median [min; max]	28.1 [1.1; 36.5]	24.4 [1.1; 36.5]	NR
Mean (SD)	27.45 (6.79)	24.49 (6.49)	NR
a: Only outcomes with the requested information available for at least one included study are presented in the dossier. b: Efficacy analyses were conducted for Arm 1 and safety and HRQoL analyses were conducted for Arm 1 + Arm 2. c: There is no information available on how the observation period was calculated. d: Median PFS/OS follow-up time only reported for the Full Analysis Set, N = 69.			
ID: identification; max: maximum; min: minimum; NR: not reported; ORR: overall response rate; OS: overall survival; PFS: progression-free survival; SD: standard deviation			

In both studies, the patients were treated for a similar duration of about 2 years. Information on observation periods is available for FIREFLY-1 only.

4.2.1.6 Subsequent therapies

Table 19 shows the number of patients who received subsequent therapy.

Table 19: Subsequent therapy after withdrawal of the study medication

Subsequent therapy	**Study reference/ID Intervention Patients with follow-up therapy n (%)**	
	FIREFLY-1 (Arm 1): tovorafenib N = 77	**Bouffet 2023: dabrafenib + trametinib N = 36**
Anti-cancer therapy	20[a] (26.0%)	NR
a: Three of these patients received tovorafenib re-treatment.		
ID: identification; N: number of randomised patients; n: number of patients in the category; NR: not reported		

Only limited information is available for both studies. In Arm 1 of FIREFLY-1, 20 (26%) patients received a subsequent anti-cancer therapy. Detailed information on the active substances the patients received is not available. Furthermore, no information on subsequent treatment is available for the subpopulation relevant for PICO 5. For Bouffet 2023, no information on subsequent therapies is reported.

4.2.2 RoB

This section shall describe the assessment of RoB at the study level taking into account, if available, the methodological guidance adopted by the HTACG pursuant to Article 3(7), point (d), of the HTAR.

For the current assessment, no data from comparative studies are available for any of the PICOs. For the uncontrolled studies, an assessment of risk of bias (RoB), at the study level, is not required.

In the dossier, the HTD uses unanchored indirect comparison methods that combine data across different studies to estimate relative effects specified in the assessment scope. These methods rely on strong assumptions that are challenging to meet in practice and the resulting effect estimates are likely to be biased if these assumptions are violated. While no RoB assessment tool is recommended for these types of analysis, the internal and external validity of the indirect comparisons submitted by the HTD have been assessed by the assessors according to the HTACG guidelines. The assessors have identified a number of major uncertainties in the relative effectiveness results, which are detailed in Section 4.3.2.

4.3 Study results on relative effectiveness and relative safety

The results on relative effectiveness and relative safety shall be presented according to the assessment scope as set out pursuant to Article 8(6) of the HTAR, per PICO.
An assessment of the degree of certainty of the relative effectiveness and relative safety, considering the strengths and limitations of the available evidence shall be performed taking into account, if available, the methodological guidance adopted by the HTACG pursuant to Article 3(7), point (d), of the HTAR.

4.3.1 Results for patient population 1 (full claimed indication)

This section shall discuss to which extent the included patient populations and/or comparator(s) per study cover the relevant patient population/comparator(s) according to the assessment scope as set out pursuant to Article 8(6) of the HTAR.
For each patient population specified in PICO(s), a separate section shall be provided. Within this section, the results for all PICO(s) addressing this patient population shall be presented in sub-sections.

As described in Section 3, the HTD does not include comparative evidence addressing patients in the full claimed indication but provides single-arm data on treatment with tovorafenib from FIREFLY-1. These data are not included in the JCA report because they do not provide information on relative effectiveness or safety for PICOs 1-4 in the context of the current assessment.

In the HTD dossier, the information on single-arm data is available in the following sections:

- Methods: Sections 4.3.1 and 4.3.2
- Patient characteristics: Section 5.3.1.1 (Table 41)
- Outcomes: Section 5.3.1.2.4 (Tables 44-49)

4.3.2 Results for patient population 2 (BRAF V600E mutation in patients > 1 year)

This section shall discuss to which extent the included patient populations and/or comparator(s) per study cover the relevant patient population/comparator(s) according to the assessment scope as set out pursuant to Article 8(6) of the HTAR.
For each patient population specified in PICO(s), a separate section shall be provided. Within this section, the results for all PICO(s) addressing this patient population shall be presented in sub-sections.

The studies included in the assessment for patients > 1 year of age with BRAF V600E mutation (Population 2) are shown in Table 20.

Table 20: Studies included in the assessment of patients > 1 year of age with BRAF V600E mutation (Population 2) per PICO question

Study reference/ID **Relevant study arms/parts (number of randomised/included patients)**	**Analysed population** **(number of randomised/included patients)**
PICO 5	
Indirect comparison (unanchored MAIC): tovorafenib vs. dabrafenib + trametinib	
FIREFLY-1 Arm 1 (N = 77) Arm 2 (N = 60)	Only patients ▪ > 1 year of age with LGG harbouring a BRAF V600E mutation Relevant subpopulation: Arm 1 (n = 13) Arm 2 (n = 9)[a]
Bouffet 2023 Part C (N = 18) Part D (N = 30)	Only patients ▪ > 1 year of age with glioma harbouring a BRAF V600E mutation Relevant subpopulation: Part C (N = 16)[b] Part D (N = 20)
PICO 6	
No relevant evidence informing this PICO provided by the HTD[c]	
a: In the dossier, the HTD includes the relevant subpopulation from Arm 2 in analyses on safety outcomes only. b: The subpopulation of Part C of Bouffet 2023 includes 2 patients with HGG who are not included in the assessment scope. c: For PICO 6 the HTD refers to single-arm data from FIREFLY-1 shown for PICO 5, without providing relative effects versus the comparator of the PICO.	
BRAF: v-Raf murine sarcoma viral oncogene homologue B; HGG: high grade glioma; HTD: health technology developer; LGG: low grade glioma; MAIC: matching-adjusted indirect comparison; N: number of randomised patients; n: number of patients;	

To address PICO 5, the HTD provides an unanchored MAIC of tovorafenib with dabrafenib + trametinib for patients > 1 year of age with BRAF V600E mutation (Population 2) which is based on subpopulations from the studies FIREFLY-1 and Bouffet 2023. This comparison is discussed in the subsequent sections.

Regarding PICO 6, the HTD did not submit data separately, but refers to the presentation of descriptive outcomes for tovorafenib relevant to this patient population, already shown for PICO 5, without providing relative effects versus the comparator of the PICO. These data are not discussed in the assessment because they do not provide information on relative effectiveness and safety.

Data sources for patients > 1 year of age with BRAF V600E mutation (Population 2) to address PICO 5

The unanchored MAIC the HTD submits for PICO 5 is based on a relevant subpopulation from FIREFLY-1 for tovorafenib and on a relevant subpopulation from Bouffet 2023 for dabrafenib + trametinib. For FIREFLY-1, the analyses use the 2-year data cut-off (10 May 2024); for Bouffet 2023, no information on the data cut-off is available in the dossier.

For tovorafenib, the HTD includes data from different analysis sets of FIREFLY-1 depending on the outcome assessed:

- for relative safety analyses, the relevant subpopulation of both arms of FIREFLY-1 is considered (Arms 1 and 2; n = 22)
- for relative effectiveness analyses, the relevant subpopulation of Arm 1 of FIREFLY-1 is considered only, and limited to the patients evaluable considering specific response assessment criteria (Arm 1 RANO-LGG evaluable set: n = 12; Arm 1 RANO-HGG evaluable set: n = 10)

In all cases, the subpopulation of patients with BRAF V600E mutation and with age > 1 year, was selected for inclusion in the relevant analysis set for FIREFLY-1. Other analysis sets were used for non-comparative analyses only and are described in Section 4.3.2.3 of the dossier. All patients in the relevant tovorafenib analysis sets also had at least 1 line of prior systemic therapy and had documented evidence of radiographic progression.

For dabrafenib + trametinib, the analysis sets for relative effectiveness and safety consisted of a subpopulation of patients enrolled into Parts C and D of Bouffet 2023. Irrespective of the outcome assessed, patients enrolled in these parts aged > 1 year, with a diagnosis of BRAF V600E mutated glioma were included in the subpopulation (N = 36). This also includes 2 patients with HGG and three patients who had not received prior systemic therapy (see Table 21). These patient characteristics do not align with Population 2 or with those of the

patients included in the analysis sets for tovorafenib which exclusively include a diagnosis of LGG and prior systemic therapy.

Both studies enrolled patients with relapsed/progressive (FIREFLY-1) or relapsed/refractory (Bouffet 2023) glioma. However, Population 2 does not specify a requirement for patients to have relapsed/progressive/refractory disease, but rather to have received prior systemic therapy only. In clinical practice, a number of factors are considered in a decision to start second- (or later) line systemic therapy and the presence of progression or refractory disease is not a prerequisite. The datasets in the analysis, therefore, represent a subset of patients within Population 2, who have relapsed/progressive/refractory disease. No data were provided for patients without relapsed/progressive/refractory disease. The final therapeutic indication differs from the claimed indication that was the basis for the JCA report (see Table 7). It covers only the subset of patients who have progressed after one or more prior systemic therapies. Therefore, the patients in both studies correspond to the final therapeutic indication, with regard to disease progression.

Reporting of outcomes based on tumour response assessment in FIREFLY 1 Arm 1 and 2

For relative effectiveness analyses for the outcomes PFS and ORR, the HTD considers results from Arm 1 of FIREFLY-1 only, although similar analyses were also prespecified for Arm 2 of the study. According to the study protocol and statistical analysis plan, ORR and PFS were planned to be assessed by IRC based on RANO-HGG, RANO-LGG and RAPNO criteria as well as by INV-assessment based on RANO-HGG criteria in Arms 1 and 2 of FIREFLY-1. However, efficacy analyses were not reported for Arm 2, neither in the dossier nor in the CSR.

The assessors requested additional relative effectiveness analyses from the HTD, in which the tovorafenib arm is informed by pooled response data from Arm 1 and Arm 2 of the FIREFLY-1 study but this was not provided by the HTD. In response to the request, the HTD describes that the sponsor of FIREFLY-1 (Day One Biopharmaceuticals, Inc.) did not conduct formal efficacy analyses for Arm 2 and no Analysis Data Model datasets are available. For patients in Arm 2 of FIREFLY-1, analyses for the same outcomes as prespecified for Arm 1 were predefined in the study protocol. The HTD provides no explanation as to why the sponsor of the study (Day One Biopharmaceuticals) did not conduct the efficacy analyses for Arm 2. However, the HTD refers to Arm 2 RANO-HGG (INV) data provided in responses to the EMA Day 120 list of questions (see Appendix D.2.2). Nevertheless, relative effectiveness analyses including Arm 2 RANO-HGG (INV) data were not provided by the HTD for Population 2. The HTD did not provide sufficient justification for not carrying out the requested analyses themselves.

4.3.2.1 Patient characteristics

This section shall present patient characteristics from all studies covering the relevant patient population included in any of PICO(s) addressing this patient population.

A summary of baseline characteristics for the included studies, prior to any statistical adjustments made for the purposes of evidence synthesis, is shown in Table 21. Standardised differences for the characteristics were not reported in the dossier and therefore not included in Table 21.

Table 21: Patient baseline characteristics for Population 2 (BRAF V600E mutation in patients > 1 year)

Study reference/ID **Characteristics** **Category**	**FIREFLY-1** **tovorafenib**			**Bouffet 2023** **dabrafenib and trametinib**
	RANO-LGG evaluable set[a]	**RANO-HGG evaluable set[b]**	**Safety analysis set[c]**	**Efficacy and safety analysis set[d]**
	Unadjusted	**Unadjusted**	**Unadjusted**	**Unadjusted**
	N = 12	**N = 10**	**N = 22**	**N = 36**
Age [years], Median	8.2	9.9	10	10
Age group, n (%)				
< 2 years	0 (0)	0 (0)	NR	1 (2.8)
2 years to < 6 years	2 (16.7)	2 (20.0)	3 (13.6)	7 (19.4)
6 years to < 12 years	7 (58.3)	5 (50.0)	11 (50.0)	12 (33.3)
12 years to < 16 years	2 (16.7)	2 (20.0)	5 (22.7)	16 (44.4)[e]
16 years to ≤ 25 years	1 (8.3)	1 (10.0)	2 (9.1)	NR
Sex, n (%)				
Male	5 (41.7)	3 (30.0)	11 (50)	18 (50)
Female	7 (58.3)	7 (70.0)	11 (50)	18 (50)
Race, n (%)				
Asian	1 (8.3)	1 (10.0)	3 (13.6)	NR
White	6 (50.0)	4 (40.0)	11 (50.0)	NR
Other	1 (8.4)	1 (10.0)	1 (4.5)	NR
Not reported	4 (33.3)	4 (40.0)	7 (31.8)	NR
Ethnicity, n (%)				
Not Hispanic or Latino	7 (58.3)	5 (50.0)	14 (63.6)	NR
Not stated	4 (33.3)	4 (40.0)	7 (31.8)	NR
Missing	1 (8.4)	1 (10.0)	1 (4.5)	NR
Height percentile at baseline[f], mean (SD)	46.3 (32.7)	45.1 (28.3)	44.5 (35.2)	NR

Study reference/ID **Characteristics** **Category**	**FIREFLY-1** **tovorafenib**			**Bouffet 2023** **dabrafenib and trametinib**
	RANO-LGG evaluable set[a]	**RANO-HGG evaluable set[b]**	**Safety analysis set[c]**	**Efficacy and safety analysis set[d]**
	Unadjusted	**Unadjusted**	**Unadjusted**	**Unadjusted**
	N = 12	**N = 10**	**N = 22**	**N = 36**
Weight percentile at baseline[f], mean (SD)	62.4 (36.1)	65.8 (32.2)	59.9 (37.0)	NR
BSA at baseline, mean (SD)	1.2 (0.3)	1.20 (0.3)	1.2 (0.4)	NR
Primary tumour location, n (%)				
Cerebral hemisphere	2 (16.7)	2 (20.0)	3 (13.6)	NR
Deep midline structures	2 (16.7)	2 (20.0)	5 (22.7)	NR
Optic pathway	4 (33.3)	3 (30.0)	7 (31.8)	NR
Brain stem	1 (8.3)	1 (10.0)	1 (4.5)	NR
Other	3 (25.0)	2 (20.0)	6 (27.3)	NR
Histology, n (%)				
Astrocytic	9 (75.0)	7 (70.0)	17 (77.3)	NR
Oligodendroglial	NA	NA	NA	NA
Mixed glial-neuronal	3 (25.0)	3 (30.0)	5 (22.7)	NR
Prior systemic therapy, n(%)	12 (100)	10 (100)	22 (100)	33 (91.7)
Any prior tumour directed surgery for treatment of primary disease, n (%)	7 (58.3)	6 (60.0)	12 (54.5)	30 (83.3)
Prior chemotherapy, n (%)	12 (100)	10 (100)	NR	31 (86.1)
Prior targeted therapy, n (%)	NR	NR	NR	4 (11.1)
Prior radiotherapy, n (%)	NR	NR	NR	1 (2.8)
Pre-operative staging, n (%)				
Localised disease	10 (83.3)	9 (90.0)	16 (72.7)	NR
Disseminated/metastatic disease	1 (8.3)	NA	4 (18.2)	NR
Leptomeningeal spread	1 (8.3)	1 (10.0)	2 (9.1)	NR
Post-operative staging[g], n (%)				
Sub-total resection	6 (50.0)	6 (60.0)	9 (40.9)	NR
Biopsy only, not attempted	6 (50.0)	4 (40.0)	13 (59.1)	NR
Baseline IRC SPPD (mm^2) by RANO-HGG, mean (SD)	842.7 (734.7)	842.7 (734.7)	842.7 (734.7)	NR
BRAF alteration, n (%)				
BRAF V600E mutation	12 (100)	10 (100)	22 (100)	36 (100)
Baseline Karnofsky/ Lansky performance score[h], n (%)				
50–70	1 (8.3)	1 (10.0)	1 (4.5)	1 (2.8)[i]

Study reference/ID Characteristics Category	FIREFLY-1 tovorafenib			Bouffet 2023 dabrafenib and trametinib
	RANO-LGG evaluable set[a]	RANO-HGG evaluable set[b]	Safety analysis set[c]	Efficacy and safety analysis set[d]
	Unadjusted	Unadjusted	Unadjusted	Unadjusted
	N = 12	N = 10	N = 22	N = 36
80–100	11 (91.7)	9 (90.0)	21 (95.5)	35 (97.2)
Time since diagnosis [months], median [min; max]	NR	NR	NR	40.2 [3.4; 123.8]
LGG, n (%)	12 (100)	10 (100)	22 (100)	34 (94.4)
HGG, n (%)	0 (0)	0 (0)	0 (0)	2 (5.6)
Treatment discontinuation, n (%)	10 (83.3)	8 (80.0)	15 (68.2)	NR
Study discontinuation, n (%)	2 (16.7)	2 (20.0)	5 (22.7)	NR

a: Analysis set for: PFS based on RANO LGG 2017 criteria (IRC)
b: Analysis set for: ORR based on RANO HGG 2010 criteria (INV, IRC)
c: Analysis set for: All submitted safety outcomes
d: Analysis set for: All submitted efficacy and safety outcomes
e: numbers refer to the age group ≥ 12 years
f: Only patients with height and weight percentile available per Centre for Disease Control and Prevention standard growth chart are included in the summary. The HTD did not specify the exact number of patients included here.
g: At the time of the initial diagnosis.
h: Denominators for Karnofsky performance score and Lansky performance score summaries are the number of patients whose ages at study enrolment were ≥ 16 years and < 16 years, respectively
i: This only includes score of 70 in Bouffet 2023, no patients had lower scores

BSA: body surface area; BRAF: v-Raf murine sarcoma viral oncogene homolog B; IRC: independent radiology review committee; HGG: high-grade glioma; LGG: low grade glioma; max: maximum; min: minimum; N: number of treated patients; n: number of patients in category; NA: not applicable; NR: not reported; PFS: progression-free survival; RANO: Response Assessment in Neuro-Oncology; SPPD: sum of products of the perpendicular diameters; SD: standard deviation

The HTD has submitted unanchored matching-adjusted indirect treatment comparisons (ITCs) to estimate relative effectiveness and safety of tovorafenib compared with dabrafenib + trametinib, using data from FIREFLY-1 and Bouffet 2023. Reliable effect estimation using these methods requires adjustment for all baseline characteristics that are prognostic or effect-modifying for the relevant outcome(s) (see Section 4.3.2.2.1). This is only possible if the relevant characteristics are reported consistently across both studies, and if there is sufficient overlap between study populations in terms of these characteristics. A major limitation of this analysis, therefore, arises from the fact that FIREFLY-1 and Bouffet 2023 reported few baseline characteristics in common. Further description of the identification of prognostic and effect-modifying variables is provided in Section 4.3.2.2.2.1.

The proportion of patients who received prior systemic therapy differs across the studies, whereby 100% of patients in FIREFLY-1 received prior systemic therapy compared with 91.7% of patients in the Bouffet 2023 study. Prior treatment including "small molecule-targeted therapy" was received by four patients (11.1%) in the Bouffet 2023 study. Patients previously treated with prior BRAF inhibitor (except for sorafenib) were only permitted to enrol in Parts C and D of the Bouffet 2023 study if they had prior benefit as determined by the investigator. No such restriction was imposed on patients receiving tovorafenib in FIREFLY-1. The proportion of patients in Population 2 receiving prior BRAF inhibitor in the FIREFLY-1 study was not reported by the HTD. According to patient data listings in the CSR of FIREFLY-1, 13 patients in Population 2 (59.1%) received prior BRAF inhibitors and 7 of these patients in Arm 1 and 2 showed a CR or PR. The remaining 6 patients did not show a response to prior treatment with BRAF inhibitors, 2 of these were included in Arm 1. 5 of the 13 patients with prior BRAF inhibitor in Arm 1 and 2 showed a SD, 1 had a PD. Prior surgery for the primary disease was less frequent in the FIREFLY-1 study than in Bouffet 2023.

Patients in the Bouffet 2023 study were slightly older than those in the FIREFLY-1 study, with a higher proportion of patients aged ≥ 12 years in the former.

There are differences in the proportions of patients with paediatric LGG across datasets, whereby 100% of patients in FIREFLY-1 were diagnosed with paediatric LGG compared with 94.4% of patients in the Bouffet study. The implications of this discrepancy for the analysis of relative effectiveness are uncertain.

Overall, it is not possible to comprehensively compare patient characteristics between the two studies due to limited reporting in Bouffet 2023. A description of the HTD's reweighting of the FIREFLY-1 population and the resulting patient baseline characteristics after reweighting, is provided in Section 4.3.2.2.

4.3.2.2 Evidence synthesis methods

This section shall briefly describe, when applicable, the evidence synthesis methods used by the HTD, including the associated strengths and limitations, and any factors arising from these methods and their application which may affect the certainty of the evidence taking into account, if available, the methodological guidance adopted by the HTACG pursuant to Article 3(7), point (d), of the HTAR.

4.3.2.2.1 Choice of evidence synthesis method

In the absence of direct comparative data for Population 2 or a connected network of evidence which would allow anchored indirect comparative methods, unanchored ITCs were carried out to estimate relative effects.

For Population 2, the assessment scope included two PICOs. However, comparative data were submitted for only one of these (PICO 5). Therefore, the assessment of evidence synthesis methods covers only the data submitted for this PICO. For PICO 5, data from FIREFLY-1 was used for the intervention and from Bouffet 2023 for the comparator.

Unanchored ITCs are a type of non-randomised evidence and, as such, are associated with an inherently higher risk of bias than comparisons informed by randomised evidence. Reliable estimation of relative treatment effects in unanchored ITCs requires analysis methods that adequately adjust for confounding and other sources of bias. In general, these methods require access to full individual participant data (IPD) for both the intervention and comparator studies, including detailed information on baseline characteristics and outcomes of interest. As IPD from Bouffet 2023 were not available to the HTD, population-adjusted indirect comparisons like matching-adjusted indirect comparison (MAIC) and simulated treatment comparison (STC) are the only potential methods to estimate treatment effects for PICO 5. The MAIC method was selected by the HTD for the indirect comparison. This method re-weights IPD from FIREFLY-1 to create a weighted pseudo-population, whose baseline characteristics align with those from Bouffet 2023, based on reported summary statistics.

The validity of relative effect estimates obtained from an unanchored MAIC relies on the following key assumptions, which are challenging to meet in practice:

- All prognostic and effect-modifying variables (PVs/EMs) are included in the weighting model.
- Either (i) the covariance structure of the included variables is the same in both studies or (ii) the included variables in the comparator study are mutually independent.
- There is sufficient overlap between the trial populations in terms of the included variables.
- Any other factors likely to affect observed outcomes (e.g., outcome definitions, assessment methods, intercurrent event handling etc.) are sufficiently similar between the studies being compared to allow for a valid comparison to be made.

If any of these assumptions do not hold then the effect estimates from the MAIC will likely be biased. In the case of the MAIC submitted by the HTD, the assessors have identified specific concerns suggesting that the first, third and fourth of these assumptions may not hold, which are described in subsequent sections. The second assumption cannot be assessed as the comparator study publication does not provide information on the covariance structure of the baseline variables, which must therefore be considered an additional risk of bias in the MAIC results.

The HTD's choice of MAIC over STC in the base case was based on their claim that STC tends to produce more bias when mis-specified or when important PVs/EMs are omitted from the model, because it targets a conditional treatment effect whereas MAIC targets a marginal treatment effect. However, the HTD does not provide evidence to support this claim and has furthermore not specified a preferred target estimand for the analysis. While the underlying approaches and assumptions of MAIC and STC differ, both methods require that all relevant PVs/EMs be adjusted for in the unanchored setting. Furthermore, concerns regarding the performance of the MAIC method (specifically, bias and coverage) in situations of poor covariate overlap have been raised in the published literature, based on the results of simulation studies and other observations [30-32]. The HTD reported that unanchored STC would be carried out as an alternative to MAIC if the reduction in the effective sample size (ESS) relative to the original sample size after MAIC weighting was over 75%. However, as this did not occur, STC was not carried out. Regardless of the choice of MAIC or STC, the small sample sizes of the FIREFLY-1 analysis sets limit the number of variables that can be adjusted for in the analysis, making it difficult to meet the requirement of adjusting for all PVs/EMs and leading to statistical imprecision (i.e., wide confidence intervals) of the resulting effect estimates.

Specific concerns that threaten the internal and external validity of the submitted MAICs and compromise the reliability of the results are discussed in subsequent sections.

4.3.2.2.2 Assessment of evidence synthesis method

The assessment of evidence synthesis methods identified specific concerns with regard to the following points:

- Identification and completeness of PVs/EMs
- Similarity of outcome definitions
- Statistical methods

These concerns are discussed in detail in the following sections.

4.3.2.2.2.1 Identification and completeness of prognostic and effect-modifying variables (PVs/EMs)

As unanchored MAIC requires adjustment for all PVs/EMs, identification of all such variables is a prerequisite for the analysis. Therefore, the process used to identify these variables has to be comprehensive and transparently reported.

This first step in this process was a literature review, conducted by the HTD, focusing on subgroup analyses presented in forest plots from FIREFLY-1 and relevant comparator studies. According to the HTD, this review included studies incorporated into the feasibility analysis

conducted for the ITC as well as all clinical studies identified in the systematic literature review (SLR). The HTD describes that publications were screened for reported subgroups and information on potential effect modifiers or prognostic variables, including a targeted, grey literature search and an examination of referenced sources cited within these studies. Internal clinical input from the HTD was also obtained to suggest potentially relevant variables. A list of identified variables was then ranked in terms of relevance by clinical and HTA experts at an advisory board. This list of variables, together with details of which variables were included in the HTD's base-case analysis, is shown in Table 22. In accordance with the ITC statistical analysis plan (SAP) (see Appendix D.5 of the dossier), univariate statistical tests were carried out by the HTD on FIREFLY-1 data to test the strength of association between each candidate variable and the outcomes of ORR and PFS. As planned (and as is appropriate), results of these tests were not used as part of the variable selection process for the primary analysis but were used to select variables for sensitivity analysis (see Section 4.3.2.2.3).

Table 22: Variables identified as prognostic or effect-modifying by the HTD

Variable	**Included in the HTD's base-case MAIC analysis**	**Details**
Age	Yes	Continuous, assuming median and mean age from Bouffet 2023 are equal
Gender	Yes (for analyses of PFS and safety outcomes) No (for analyses of ORR based on RANO-HGG) Unclear, due to ambiguous reporting in the dossier (for analyses of ORR based on RANO-LGG)	Categorical: male versus female
Karnofsky/Lansky score (combined)	Yes	Categorical: < 80 versus 80-100
Histological type	No	Not reported
Primary Tumour location	No	Not reported
BRAF V600E Mutation	No	Applies to all patients in Population 2
BRAF fusion	No	Not relevant for this analysis, as it applies to no patients in Population 2
Prior surgery	Yes	Categorical: yes versus no
Prior chemotherapy	No	Applies to all patients in FIREFLY-1 but not in Bouffet 2023
Prior radiation therapy	Yes (for analyses of PFS and safety outcomes) No (for analyses of ORR)	Categorical – yes versus no
Prior MAPK inhibitor treatment	No	Not reported

Variable	Included in the HTD's base-case MAIC analysis	Details
HGG: high-grade glioma; HTD: health technology developer; LGG: low-grade glioma; MAIC: matching adjusted indirect comparison; MAPK: mitogen-activated protein kinase; ORR: objective response rate; RANO: Response Assessment in Neuro-oncology; PFS: progression-free survival		

For all the steps described by the HTD, no further information was provided on the variables initially considered nor the criteria used to identify these variables as potentially prognostic and/or effect-modifying. The HTD did not provide any further information on the specific expert input received or on the methods by which the potential PVs/EMs were presented to, considered or confirmed by the advisory board. Based on the available information, it is therefore not possible to establish whether the identification process was comprehensive or whether the list is likely to be complete.

Regardless of whether the list of potentially prognostic and/or effect-modifying variables is likely to be complete, not all identified variables were taken into account in the analyses. Some of the variables identified were not relevant for the analysis in Population 2 (e.g., BRAF fusion versus V600E mutation), but other variables could not be adjusted for because they apply to either all patients in FIREFLY-1 but not to all patients in Bouffet 2023 (e.g., prior chemotherapy) or were not reported in Bouffet 2023 and can therefore not be included (e.g., histological type, prior tumour location). Of the remaining variables, five variables were mutually reported in the relevant datasets and adjusted for in the MAIC analysis for PFS and safety outcomes (age, prior surgery, Karnofsky/Lansky score, prior radiotherapy and gender). For the MAIC analysis of ORR based on RANO-HGG criteria, three variables were adjusted for (age, prior surgery, Karnofsky/Lansky score), with the HTD reporting that gender was not included as the weights did not converge until gender was removed from the model. No justification for the exclusion of prior radiotherapy in the MAIC analysis of ORR based on RANO-HGG criteria was provided by the HTD. For the analysis of ORR based on RANO-LGG criteria, it is unclear from the dossier which variables have been adjusted for due to inconsistent reporting. In Table 60 of the main dossier, the HTD has reported that three variables (age, prior surgery, Karnofsky/Lansky score) were adjusted for; however, Table 36 of Appendix B.2 of the dossier reports that a fourth (gender) was also included. Again, "prior radiotherapy" was reported as having been excluded, without justification by the HTD. Furthermore, the results of the MAIC for ORR and PFS by RANO-LGG (described further in Section 4.3.2.3) report the same ESS, indicating that the adjustment is likely to have been based on the same set of variables. It is unclear why different sets of variables would be used for the analysis of ORR and PFS by RANO-LGG as these comparisons were based on the same analysis sets.

Summary

As the methods used to identify the final list of potential prognostic and effect-modifying variables was not transparently reported, it is not possible to establish whether the identification process was comprehensive or whether this list is likely to be complete. Regardless of this, histological type, primary tumour location and prior chemotherapy were all identified as likely to be prognostic and/or effect-modifying but could not be adjusted for in the MAIC. Furthermore, having a diagnosis of HGG versus LGG is also likely to be prognostic and/or effect-modifying and applies to a small number of patients in Bouffet 2023, who could not be removed from the analysis due to the absence of IPD or suitable subgroup data from Bouffet 2023. Therefore, the requirement that all relevant prognostic and effect-modifying variables be adjusted for in the MAIC has not been met, leading to bias of unknown magnitude and direction in the resulting treatment effect estimates. In addition, the list of variables which were actually used by the HTD to conduct the adjustment are unclear due to inconsistent reporting and ambiguous results.

4.3.2.2.2.2 Operationalisation and comparability of outcomes

Results for the majority of outcomes, for which relative effect estimates were requested in the assessment scope, are not provided in the dossier. This is primarily due to the absence of comparator data for these outcomes or to differences between the definitions of the outcomes used in the studies which limit their comparability. The study definitions of each of the outcomes that were reported in both FIREFLY-1 and Bouffet 2023 and an assessment on their comparability and the implications for feasibility of indirect comparisons are presented in Table 23. An overview of all outcomes for which relative effect estimates were requested in the assessment scope and for which results are included in assessment is presented in Table 25.

Although Bouffet 2023 reported the outcomes of CR, PR and SD by both RANO-LGG and RANO-HGG, the number of subjects with SD of minimum 6 and/or 12 months' duration was not reported. For this reason, the "Composite outcome consisting of CR, PR and SD lasting a minimum of (i) 6 and (ii) 12 months" requested in the assessment scope was not included in the dossier.

Table 23: Outcome definitions, handling of missing data/intercurrent events, and assessment of comparability

Outcome characteristic	**FIREFLY-1**	**Bouffet 2023**	**Assessment of comparability**
Objective response based on RANO-LGG 2017 criteria			
Disease assessment criteria	RANO-LGG 2017	Modified RANO-LGG 2017	Definitions of ORR, based on different assessment criteria, are not comparable due to the inclusion of the MR
Definition of	Binary outcome defined as	Binary outcome defined as	

Outcome characteristic	FIREFLY-1	Bouffet 2023	Assessment of comparability
objective response	confirmed BOR of either CR, PR or MR.	confirmed BOR of either CR or PR. Subjects with BOR of MR were classified as non-responders.	category in the objective responses in FIREFLY-1 but not in Bouffet 2023. This inconsistency is expected to bias the comparison of ORR in favour of tovorafenib. *Despite these differences, a comparison of ORR using these definitions was provided by the HTD.* *An alternative definition of objective response in FIREFLY-1 including just CR and PR, could have been derived from FIREFLY-1 data, and would have been aligned with the Bouffet 2023 definition. This comparison was requested by the assessors but was not provided by the HTD.*
Missing or non-evaluable response data	Subjects with missing or non-evaluable post-baseline response data were classified as non-responders.	Subjects with missing or unknown BOR were classified as non-responders.	
Assessments following subsequent anti-cancer therapy	Subjects with initial CR, PR or MR after the start of subsequent anti-cancer therapy were classified as non-responders.	Handling of subsequent anti-cancer therapy not reported.	
IRC- or INV-assessment	IRC-assessed response available only.	IRC-assessed response available only.	
Objective response based on RANO-HGG 2010 criteria			
Disease assessment criteria	RANO-HGG 2010	RANO-HGG 2010	Definitions of ORR, based on RANO-HGG assessment criteria, are comparable. *Analyses based on this comparison were provided by the HTD, for INV-assessment in the dossier and for IRC-assessment in the response to an additional information request of the assessors (see Appendix D.2.2).*
Definition of objective response	Binary outcome defined as confirmed BOR of either CR or PR per RANO-HGG 2010 criteria.	Binary outcome defined as confirmed BOR of either CR or PR per RANO-HGG 2010 criteria.	
Missing data, non-evaluable response, and subsequent anti-cancer therapy	As per analysis by RANO-LGG criteria described above.	As per analysis by (modified) RANO-LGG criteria described above.	
IRC- or INV-assessment	IRC- and INV-assessed response available.	IRC- and INV-assessed response available.	
PFS based on RANO-LGG 2017 criteria			
Disease assessment criteria	*RANO-LGG 2017*	*RANO-LGG 2017*	Definitions, censoring rules and assessment schedules for IRC-assessed PFS, based on RANO-LGG 2017 criteria, appear to be broadly comparable overall. *An analysis based on this comparison was provided by the HTD.*
Definition of PFS	Time from initiation of treatment until PD or death.	Time from the start of treatment with the study medication until the first documented progression or death from any cause.	
Censoring rules	Censoring rules are shown in Table 27 of the HTD dossier.	The following information was provided: If no event	

Outcome characteristic	FIREFLY-1	Bouffet 2023	Assessment of comparability
		occurred, the time was censored at the time of the last adequate tumour evaluation. For patients who initiated further antineoplastic therapy, the follow-up time was censored at the time of the last adequate tumour evaluation before the start of the new antineoplastic therapy.	
Assessment schedule	Radiographic disease assessments were performed at baseline, at the end of every third cycle (each cycle: 28 days) of treatment starting with the end of cycle 3, at the EOT visit and every 3 (+/-1) months during the LTFU period.	Radiographic disease assessments were performed at baseline, every 8 weeks for the first 24 weeks, and then every 12 weeks thereafter.	
INV- or IRC-assessment	Outcome definition based on IRC -assessed PD available only.	Outcome definition based on IRC -assessed PD available only.	
PFS based on RANO-HGG 2010 criteria			
Disease assessment criteria	***RANO-HGG 2010***	***RANO-HGG 2010***	Definitions and assessment schedules for INV-assessed PFS, based on RANO-HGG. 2010 criteria, appear to be broadly comparable overall. *An analysis based on this comparison was not provided by the HTD.*
Definition, censoring rules and assessment schedule	As per analysis of PFS by RANO-LGG 2017 criteria.	As per analysis of PFS by RANO-LGG 2017 criteria.	
INV- or IRC-assessment	INV- and IRC -assessed PFS available.	INV-assessed PFS available.	
Safety outcomes: Any AEs, Serious AEs, AEs by CTCAE grade ≥ 3, Deaths due to AEs, Treatment discontinuation due to AEs, Treatment interruption due to AEs			
Definition	Binary outcomes measuring the number of subjects in the safety analysis set experiencing the relevant event.	Binary outcomes measuring the number of subjects in the safety analysis set experiencing the relevant event.	Definitions of safety outcomes are comparable.

AE: adverse event; BOR: best overall response; CR: complete response; CBR: clinical benefit rate; CTCAE: common terminology criteria for adverse events; EOT: end of treatment; HGG: high-grade glioma; HTD: health technology developer; INV: investigator assessed; IRC: independent radiology review committee; LGG: low-grade glioma; LTFU: long-term follow-up; MAIC: matching adjusted indirect comparison; MR: minor response; ORR: objective response rate; OS: overall survival; PD: progressed disease; PFS: progression-free survival; PR: partial response; RANO: Response Assessment in Neuro-oncology; RAPNO: Response Assessment in Paediatric Neuro-oncology; SAE: serious adverse event; SD: stable disease

ORR

Differences in the specific criteria used to measure and report response in the FIREFLY-1 and Bouffet 2023 studies and in the data selected by the HTD for inclusion in the comparative analysis compromise the reliability of the analysis of ORR based on RANO-LGG criteria. For the analysis of ORR based on RANO-LGG (IRC) criteria, the definitions of ORR used in the HTD's analysis differed between FIREFLY-1 and Bouffet 2023. ORR was defined in FIREFLY-1 as the composite of CR, PR and MR according to RANO-LGG criteria in FIREFLY-1. In Bouffet 2023, modified RANO-LGG criteria were used with ORR defined as a composite of CR and PR, and patients with MR were considered to have SD, and were not separately reported. The inclusion of the MR category in the definition of ORR in the tovorafenib arm but not the dabrafenib + trametinib arm is methodologically inappropriate and leads to an ambiguous definition of the target estimand. If the relative effect estimate is interpreted as the relative effect on the outcome of ORR by RANO-LGG including MR, and if some of the patients classified as SD in the Bouffet study did in fact meet the criteria for MR, then the exclusion of MR from the dabrafenib + trametinib arm would underestimate ORR in this group and thus bias the relative effect estimate in favour of tovorafenib. If the relative effect estimate is interpreted as the relative effect on ORR by RANO-LGG excluding MR, then the inclusion of MR in the tovorafenib arm will overestimate ORR in this group and again bias the relative effect estimate in favour of tovorafenib. The impact of this is likely to be substantial as 2 of the 7 (28.6%) objective responses by RANO-LGG in the relevant subgroup of FIREFLY-1 were classed as MR. A methodologically appropriate comparative analysis of ORR based on modified RANO-LGG criteria is possible, using a consistent definition of objective response for FIREFLY-1 as the composite of PR and CR only, as reported in the Bouffet 2023 study. This was requested of the HTD by the assessor and co-assessor but was not provided. Due to the clear bias arising from the methodological flaws and inappropriate analysis methods used by the HTD, the results of the analysis of ORR based on RANO-LGG criteria are not presented in the assessment.

For the analysis of ORR based on RANO-HGG (INV) criteria, the definitions are comparable between the studies. This analysis was also requested by the assessors and provided by the HTD after clarification of the criteria used in Bouffet 2023. In the response to the request for additional analysis, the HTD also included analyses of ORR based on RANO-HGG (IRC) criteria using data on RANO-HGG assessed by IRC for Bouffet 2023. For details see Appendix D.2.2.

PFS

The outcome of PFS in both FIREFLY-1 and Bouffet 2023 was defined as time from initiation of treatment until progression or death, with PD defined according to RANO-LGG criteria and assessed by IRC. To obtain PFS data, published Kaplan-Meier (KM) curves from Bouffet 2023 were digitised using Engauge Digitizer software and pseudo-IPD was constructed using the Guyot algorithm [33]. As IPD on PFS times from FIREFLY-1 were available to the HTD no digitisation was required for this group.

The HTD has argued throughout the dossier that the results of the ITC for PFS are unreliable and highly likely to be biased as a result of differences in PFS censoring between studies as well as other factors (see the section on additional discussion below). The censoring rules for PFS for FIREFLY-1 are presented in the dossier (Section 4.3.2.6, Table 27). No details of PFS censoring rules nor numbers of patients censored for different reasons were provided in the journal publication of Bouffet 2023 (this is incorrectly described as "non-informative censoring" in the dossier), though additional information referred to in Appendix D.2.2 indicates that the rules were broadly similar to FIREFLY-1. Nonetheless, validity of the comparison of PFS relies on the assumption of non-informative censoring, which may be violated if censoring mechanisms differ between treatment groups. As the numbers of patients censored for different reasons are not provided in Bouffet 2023, this assumption cannot be reliably assessed. The HTD further argues in the dossier that differences in the rates of PFS censoring between studies limits their comparability, with PFS censoring rates (per IRC-assessed RANO-LGG criteria) of 8.3% in FIREFLY-1 and 75% in Bouffet 2023 (estimated from the digitised KM curve). The assessors disagree with this conclusion and note that differences in overall rates of censoring do not provide evidence that the mechanisms of censoring differ between studies, since the higher rate of censoring in Bouffet 2023 may have arisen as a result of longer underlying PFS observed in this group (i.e., more patients reaching the end of study follow-up without prior progression).

Additional discussion of PFS

In addition, the HTD discusses issues regarding the assessment of PFS by IRC or INV as well as the clinical plausibility of the PFS results from Bouffet 2023. Firstly, the HTD describes differences in IRC- and INV-assessed PFS results as evidence of bias in Bouffet 2023. The assessors do not agree with this conclusion since IRC- and INV-assessment of progression in Bouffet 2023 were based on different criteria (RANO-LGG and RANO-HGG respectively), therefore differences are expected. Furthermore, bias arising from "optimistic" INV-assessment, if present, would not affect IRC-assessed PFS which is the operationalisation of PFS used in the ITC. Secondly, the HTD argues, based on clinical expert input, that PFS reported in Bouffet 2023 is clinically implausible for a population of patients with second and later line (2+L) paediatric LGG, as it exceeds PFS reported in the TADPOLE study which enrolled patients with first-line (1L) paediatric LGG. This conclusion of the HTD is based on a naïve, indirect comparison of PFS between Bouffet 2023 and TADPOLE, which does not account for differences in baseline characteristics or study design. Other issues, cited by the HTD as evidence of bias for the ITC of PFS, relate to the population of Bouffet 2023 and have been discussed previously in Section 4.3.2.1.

Safety

Safety outcomes were defined as binary outcomes measuring the number of patients in the safety analysis set experiencing the relevant event at least once. No adjustment could be

made for differences in duration of treatment exposure between patients as these data were unavailable from Bouffet 2023. This may limit the comparability of safety data between treatment groups, though it is not possible to predict the magnitude and direction of any bias that may arise as a result of this.

4.3.2.2.2.3 Statistical methods

The MAIC weights were derived from a propensity-score type model estimated using logistic regression, which models the odds of assignment to the comparator study as a function of the selected baseline variables. The model was fitted using the method of moments, as described in [34].

In the case of missing baseline covariate data from FIREFLY-1, the HTD stated that the extent and patterns of missingness would be evaluated and that multiple imputation of missing data would be implemented where appropriate. According to Section 4.3.4.2 of the HTD dossier, no missing data were identified in the IPD from FIREFLY-1. The HTD reported that subjects with missing baseline covariate data in comparator studies would be excluded for the purposes of MAIC-weighting, i.e., that aggregate summary percentages would be calculated based on those subjects with available data only. While the HTD states that missing data were identified in the aggregate data for the comparators, this statement appears to refer to the non-reporting of aggregate summary data on covariates identified as potential PVs/EMs, as opposed to missing observations of reported covariates, which is a separate issue. Based on the reported summaries in Bouffet 2023, there were no missing observations for the variables age, Karnofsky/Lansky performance status and gender, while it is unclear whether or not there were missing observations for prior surgery and prior radiotherapy. Given the small sample sizes, the MAIC results may be sensitive to missing baseline covariate data and therefore any such missing data, if present, would further limit the internal and external validity of the results.

The ESS after weighting was estimated in order to assess the expected loss of precision arising as a result of weighting, and the covariate overlap of the study populations. The distribution of weights was also examined in order to identify highly influential observations, which may provide additional evidence of poor covariate overlap. Additional methods for the assessments of positivity and overlap assumptions were described in the SAP but the results of these assessments are not reported in the dossier. Therefore, this important requirement for the validity of the MAIC cannot be fully assessed. This results in additional uncertainty of any effect estimates provided in the dossier by the HTD.

Binary outcomes (ORR and Safety) were analysed using a (weighted) logistic regression model with a single covariate for treatment group, with treatment effects for these outcomes estimated and reported on the odds ratio scale. Confidence intervals and p-values for relative

effect estimates were obtained from the logistic regression model. The robust sandwich estimator was used to account for the additional variance that arises from the MAIC re-weighting process, as recommended in the HTACG guidelines and elsewhere [35,36].

Time-to-event outcomes (PFS) were analysed using a (weighted) Cox model with a single covariate for treatment group, with treatment effects measured on the hazard ratio scale. Confidence intervals for HRs were derived from the Cox model and p-values were calculated using the associated Wald test, again using the robust sandwich variance estimator. To account for possible violations of the proportional hazards (PH) assumption (see below), an additional analysis of PFS based on restricted mean survival time (RMST) was also carried out, with treatment effects measured as the difference in RMST between arms. The follow-up timepoint of 28.43 months for the analysis of RMST was defined as the midpoint of (i) the minimum of the longest observed PFS times (including censored observations) for the intervention and control arms and (ii) the smaller of the maximum observed PFS *event* times for the intervention and control arms. The reason for this choice of timepoint is not explained in the dossier, though the assessors note that it is sufficient to capture all PFS events in the treatment groups with the shorter maximum PFS event time. Confidence intervals and p-values for the comparison of the absolute difference in RMST were reported, though no details were provided on how these were calculated. The HTD stated that piecewise Cox regression would also be carried out if the PH assumption was violated; however, no results of this analysis were reported.

No pre-planned hypothesis testing for the ITCs was described in the relevant sections of the dossier nor in Appendices B.2 and D.5. However, footnotes to the relative effectiveness results tables (Tables 60 and 61 of the HTD dossier) state that 95%-CIs excluding the null and a p-value < 0.05 correspond to a statistically significant result. The assessors therefore understand that all reported p-values are nominal, no hypothesis tests were prespecified and that no control for multiplicity was implemented. Furthermore, although multiple operationalisations of the ORR and PFS outcomes were available (RANO-HGG and RANO-LGG, INV- and IRC-assessment), no preferred “primary” analysis was clearly specified in the ITC SAP.

The HTD stated in the dossier that the PH assumption was “visually violated based on the Schoenfeld Residuals and Log Cumulative Hazard plots and statistically based on the Grambsch-Therneau test (i.e., p-value < 0.05)” and that the comparison of PFS on the absolute difference in RMST scale was therefore the preferred analysis. No further details of the PH assessment were provided; therefore, the assessors cannot assess the appropriateness of this conclusion.

Relative effect measures for the outcome of PFS measured as a dichotomous outcome at 6 and 12 months were requested in the assessment scope. Rates of PFS at of 6 and 12 months were estimated using the Kaplan-Meier method and were reported separately for each

treatment group; however, the corresponding relative effect measures were not provided in the dossier. These relative effects were therefore calculated by the assessors, measured as relative risks of the underlying dichotomous outcomes, using the approach described in Appendix C.2.4.

Statistical analysis was conducted in R (version 4.5.1) and MAIC weighting was carried out using the "maic" package (a reference for this package was not provided by the HTD).

4.3.2.2.3 Subgroup and sensitivity analysis

The HTD provides different types of sensitivity analyses. However, given the limited information on confounding variables available for the unanchored MAIC presented in the dossier, the sensitivity analyses provide insufficient information to understand the remaining uncertainties.

Sensitivity of the MAIC results to variable selection

Sensitivity of the MAIC results to variable selection was investigated via a scenario analysis in which only those variables ranked of median or high importance by clinicians were included in the adjustment; for ORR, one variable (age) and for PFS, two variables (age and prior radiotherapy) were included in this analysis. In the case of PFS, additional sensitivity analysis was carried out in which only those variables demonstrating a statistically significant association with PFS in univariate analysis of FIREFLY-1 data (based on the overall study population) were included in the adjustment (age, Karnofsky/Lansky scores, and prior radiotherapy). As these analyses include only subsets of variables adjusted for in the primary analysis, they are also associated with a risk of bias due to missing PVs/EMs and do not provide any meaningful information on the sensitivity of the MAIC results to this major uncertainty.

Sensitivity analyses for possible unmeasured confounding

The sensitivity of the results to possible unmeasured confounding was investigated using shifted null hypothesis tests and by calculating E-values associated with the effect estimates. The results have been reported for the outcome of ORR only, although this was not specified in the ITC methods section and this analysis would have been possible for the comparisons of PFS and safety outcomes.

Shifted null hypotheses

Shifted null hypothesis testing involved testing treatment effect estimates for statistical significance against a null hypothesis that is shifted away from the "conventional" null of equivalence between treatment groups. In the context of non-randomised studies, such as unanchored MAICs, the value of the shifted null hypothesis is typically pre-specified to be of sufficient magnitude to ensure that if a statistically significant effect is observed, it is very unlikely to have arisen as a result of unmeasured confounding and/or other biases [37]. For

the shifted null hypothesis tests, effect-estimates were tested against a range of shifted null hypotheses, with values ranging from -0.5 to 0.5 in increments of 0.01 on the log-odds ratio scale (binary outcomes), which corresponds to a range of approximately 0.61 to 1.65 on the un-transformed scale. The range of shifted null hypotheses tested by the HTD was selected based on Horita et al. (2024) and Fernainy and Sourial (2022) [38,39], which discuss minimally clinically important differences and do not provide information that can be used to estimate the possible magnitude of unmeasured confounding in the MAIC results. The HTD then calculated the "critical" value of the shifted null hypothesis for which the conclusion regarding statistical significance would change, if such a value could be found in the range of -0.5 to 0.5. In cases where no such value could be found, then it was reported as either -0.5 or 0.5, regardless of whether or not this value was contained within the 95% confidence interval, which differs from the approach prespecified in the ITC SAP (see Appendix D.5 of the dossier). The reason for this approach is unclear since the critical value, if it can be found within the prespecified range, is by definition equal to the (log-transformed) limit of the 95% confidence interval which is closest to 1 (0, if log-transformed); therefore this analysis provides no additional information beyond what is provided by the confidence interval itself. Contrary to what is claimed by the HTD in the dossier, a large (absolute) critical value indicates that the results are robust to unmeasured confounding. For these reasons, the results of the HTD's shifted null hypothesis testing have not been presented in the JCA.

E-values

The E-value, which measures the minimum strength of association between unmeasured confounding and the outcome that would be required to fully explain the treatment effect was calculated for binary outcomes. Large E-values indicate that a greater degree of unmeasured confounding would be required to overturn or to reverse the conclusion regarding treatment effectiveness. Particularly in the present case, however, with a high likelihood of residual confounding due to incomplete PVs/EMs identification and inclusion in the MAIC (Section 4.3.2.2.2.1), the reported E-Values must be viewed critically. Indeed, the more PVs/EMs are missing, the higher E-values must be to provide convincing evidence that results are robust to unmeasured confounding. Without information about the strength of association between missing, relevant PVs/EMs and variables already taken into account, an E-value is difficult to interpret.

In the dossier, the HTD presented a table of proposed interpretations of ranges of E-value thresholds (Table 28 of Appendix B.2 in the dossier); however, the assessors could not identify a source for this table and note that the cited source [40] explicitly states that "we do not propose any threshold cut-off for the E-value." Furthermore, E-values were calculated for the point-estimates of treatment effects but not for the limits of the confidence intervals; therefore, these values represent only the degree of unmeasured confounding that would be required to nullify this point-estimate. A lesser degree of unmeasured confounding would be

required to alter the conclusion regarding statistical significance of the treatment effect estimate, but the corresponding E-value has not been reported by the HTD. Finally, the formula for the E-value used by the HTD relies on an approximation of the relative risk by the odds ratio which is only recommended in VanderWeele and Ding (2017) [40] when the outcome is uncommon (< 15%), which is not the case for ORR in the current study (an alternative formula to approximate the E-value for more common outcomes is recommended in VanderWeele and Ding (2017) [40] but was not used). As the E-values have been calculated incorrectly and only for the point estimates, they are not reported in the JCA.

Naïve indirect comparison

A naïve indirect comparison (i.e., a comparison of outcomes between treatment groups without adjustment for imbalances in baseline covariates) is also provided. Such naïve comparisons are prone to bias and should not be relied upon for effect estimation but may nonetheless provide useful information on the impact of population-adjustment on study outcomes.

Subgroup analyses

Subgroup analyses as requested per assessment scope on age and prior treatment lines were not available for the JCA as no data on these subgroups are reported in Bouffet 2023.

4.3.2.2.4 Patient characteristics before and after weighting

Baseline characteristics after MAIC weighting in the datasets, reported by the HTD as having been used for the analysis of efficacy outcomes, are shown in Table 24. The assessors highlight the following concerns which suggest that the presented data may contain errors:

- In the case of the RANO-LGG evaluable set, the HTD reports that different sets of covariates were adjusted for in the base case analyses of ORR and PFS (see Section 4.3.2.2.2.1); however, baseline characteristics for a single “adjusted base case” are reported in the dossier (Table 51) for this analysis set. Furthermore, the same ESS is reported for both outcomes in the presentation of the relative effectiveness results in the dossier (Tables 60 and 61), which is unlikely to be correct if the set of adjusted variables are different.

For all categorical variables, the sum of the weights across the different categories should be the same (in the absence of missing data). This is not the case for a number of variables in Table 24: for example, in the “adjusted base case” for the RANO-LGG evaluable set, the sum of the weights for the “Karnofsky/Lansky score” variable is 5.8 but the sum for the “Post-operative staging” variable is 8.41. In both cases, no missing observations are reported for the unadjusted data and the sum of the patient numbers is 12; therefore this discrepancy cannot be explained by missing data.

The weighting process achieved exact balance on all included covariates for the RANO-LGG and RANO-HGG analysis sets.

For the MAIC in the RANO-LGG evaluable set, the ESS was reduced by over 50%, compared with the original sample size (from 12 to 5.81) as a result of the weighting process, indicating somewhat poor covariate overlap, resulting in statistical imprecision (i.e., wide confidence intervals) and potentially increased risk of bias in the resulting effect estimates, based on previous simulation studies [30,31]. Reductions in ESS of this magnitude also indicate differences between study populations and a corresponding risk that the effect estimates, which correspond to the population of Bouffet 2023, may not be applicable to the wider FIREFLY-1 subpopulation corresponding to this PICO. Plots of the weight distributions, shown in Appendix C.2.1, showed that the majority of subjects in FIREFLY-1 were down-weighted, and the largest weight was between 2.5 and 3. In the context of an underlying sample size of n = 12 and many participants being down-weighted, a single observation of this weight is likely to be highly influential on the results.

For the MAIC in the RANO-HGG evaluable set and the safety evaluable set, the relative reductions in ESS were smaller at approximately 34% and 35% respectively. In the case of the RANO-HGG evaluable set, this is likely as a result of fewer variables being included in the re-weighting process and should not be interpreted as evidence of greater comparability between treatment groups, but rather that “comparability” is being measured on the basis of fewer baseline characteristics. Plots of the weight distribution for the RANO-HGG evaluable analysis set were not provided by the HTD.

The HTD does not provide a comparison of baseline characteristics after MAIC weighting in the safety dataset (Table 24). Therefore, the level of adjustment cannot be assessed for the analyses of safety outcomes.

Similarly, baseline characteristics after MAIC weighting are not reported for the second scenario analysis (adjustment for statistically significant variables only) in the comparison of PFS.

Table 24: Patient baseline characteristics for Population 2 (BRAF V600E mutation in patients > 1 year) for PICO 5, before and after matching

Study reference/ID **Characteristics** **Category**	**FIREFLY-1 tovorafenib**								**Bouffet 2023 dabrafenib and trametinib**
	RANO-LGG evaluable set[a] Informs MAIC analysis of PFS			**RANO-HGG evaluable set[b] Informs MAIC analysis of ORR**			**Safety analysis set[c] Informs MAIC analysis of all safety outcomes**		**Efficacy and safety analysis set[d]**
	Unadjusted	**Adjusted Base case[e]**	**Adjusted Scenario 1[f]**	**Unadjusted**	**Adjusted Base case[g]**	**Adjusted Scenario 1**	**Unadjusted**	**Adjusted Base case[e]**	**Unadjusted**
	N = 12	**ESS = 5.81**	**ESS = 11.54**	**N = 10**	**ESS = 6.64**	**ESS = 10**	**N = 22**	**ESS = 14.26**	**N = 36**
Age [years], Median	8.2	10.0	10.0	9.9	10.0	10.0	10	NR	10
Age group, n (%)									
< 2 years	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	NR	NR	1 (2.8)
2 years to < 6 years	2 (16.7)	0.739 (9.1)	1.66 (14.1)	2 (20.0)	1.04 (13.4)	2.04 (20.4)	3 (13.6)	NR	7 (19.4)
6 years to < 12 years	7 (58.3)	5.55 (68.1)	7.25 (61.6)	5 (50.0)	4.02 (51.6)	4.97 (49.7)	11 (50.0)	NR	12 (33.3)
12 years to < 16 years	2 (16.7)	1.63 (20.0)	1.53 (13.0)	2 (20.0)	2.51 (32.2)	1.95 (19.5)	5 (22.7)	NR	16 (44.4)[h]
16 years to ≤ 25 years	1 (8.3)	0.228 (2.8)	1.33 (11.3)	1 (10.0)	0.22 (2.8)	1.04 (10.4)	2 (9.1)	NR	NR
Sex, n (%)									
Male	5 (41.7)	4.07 (50)	5.05 (39.8)	3 (30.0)	3.89 (50)	2.98 (29.8)	11 (50)	NR	18 (50)
Female	7 (58.3)	4.07 (50)	7.64 (60.2)	7 (70.0)	3.89 (50)	7.02 (70.2)	11 (50)	NR	18 (50)
Race, n (%)									
Asian	1 (8.3)	0.099 (1.21)	1.04 (8.9)	1 (10.0)	0.33 (4.3)	1 (10.0)	3 (13.6)	NR	NR
White	6 (50.0)	5.91 (72.6)	5.76 (48.9)	4 (40.0)	4.41 (56.7)	4.01 (40.1)	11 (50.0)	NR	NR
Other	1 (8.4)	0.268 (3.29)	1.07 (9.1)	1 (10.0)	0.255 (3.3)	9.89 (9.89)	1 (4.5)	NR	NR
Not reported	4 (33.3)	1.86 (22.9)	3.90 (33.1)	4 (40.0)	2.78 (3.6)	4.01 (40.1)	7 (31.8)	NR	NR
Ethnicity, n (%)									
Not Hispanic or Latino	7 (58.3)	5.73 (70.4)	6.96 (59.1)	5 (50.0)	4.14 (53.2)	5 (50.0)	14 (63.6)	NR	NR
Not stated	4 (33.3)	2.18 (26.8)	3.49 (29.6)	4 (40.0)	3.43 (44.0)	4 (40.0)	7 (31.8)	NR	NR

Study reference/ID Characteristics Category	FIREFLY-1 tovorafenib								Bouffet 2023 dabrafenib and trametinib
	RANO-LGG evaluable set[a] Informs MAIC analysis of PFS			RANO-HGG evaluable set[b] Informs MAIC analysis of ORR			Safety analysis set[c] Informs MAIC analysis of all safety outcomes		Efficacy and safety analysis set[d]
	Unadjusted	Adjusted Base case[e]	Adjusted Scenario 1[f]	Unadjusted	Adjusted Base case[g]	Adjusted Scenario 1	Unadjusted	Adjusted Base case[e]	Unadjusted
	N = 12	ESS = 5.81	ESS = 11.54	N = 10	ESS = 6.64	ESS = 10	N = 22	ESS = 14.26	N = 36
Missing	1 (8.4)	0.228 (2.8)	1.33 (11.3)	1 (10.0)	0.21 (2.8)	1 (10.0)	1 (4.5)	NR	NR
Height percentile at baseline[i], mean (SD)	46.3 (32.7)	62.5 (NR)	47.1 (NR)	45.1 (28.3)	45.8 (27.6)	44.9 (28.1)	44.5 (35.2)	NR	NR
Weight percentile at baseline[i], mean (SD)	62.4 (36.1)	72.1 (NR)	62.1 (NR)	65.8 (32.2)	65.6 (30.1)	65.5 (32.3)	59.9 (37.0)	NR	NR
BSA at baseline, mean (SD)	1.2 (0.3)	1.26 (NR)	1.18 (NR)	1.20 (0.3)	1.20 (0.2)	1.19 (0.3)	1.2 (0.4)	NR	NR
Primary tumour location, n (%)									
Cerebral hemisphere	2 (16.7)	1.31 (16.1)	2.38 (18.7)	2 (20.0)	2.15 (27.7)	1.99 (19.9)	3 (13.6)	NR	NR
Deep midline structures	2 (16.7)	0.978 (1.2)	2.23 (17.5)	2 (20.0)	1.47 (18.8)	2.04 (20.4)	5 (22.7)	NR	NR
Optic pathway	4 (33.3)	1.89 (23.2)	3.97 (31.3)	3 (30.0)	2.43 (31.2)	2.97 (29.7)	7 (31.8)	NR	NR
Brain stem	1 (8.3)	0.098 (1.21)	1.03 (8.11)	1 (10.0)	0.33 (4.4)	1 (10.0)	1 (4.5)	NR	NR
Other	3 (25.0)	3.86 (47.4)	3.09 (24.4)	2 (20.0)	1.4 (17.9)	2 (20.0)	6 (27.3)	NR	NR
Histology, n (%)									
Astrocytic	9 (75.0)	6.99 (85.9)	9.59 (75.6)	7 (70.0)	6.06 (77.7)	7 (70.0)	17 (77.3)	NR	NR
Oligodendroglial	NA	NA	NA	NA	NA	NA	NA	NR	NA
Mixed glial-neuronal	3 (25.0)	1.15 (14.1)	3.09 (24.4)	3 (30.0)	1.73 (22.3)	3 (30.0)	5 (22.7)	NR	NR

Study reference/ID Characteristics Category	FIREFLY-1 tovorafenib								Bouffet 2023 dabrafenib and trametinib
	RANO-LGG evaluable set[a] Informs MAIC analysis of PFS			RANO-HGG evaluable set[b] Informs MAIC analysis of ORR			Safety analysis set[c] Informs MAIC analysis of all safety outcomes		Efficacy and safety analysis set[d]
	Unadjusted	Adjusted Base case[e]	Adjusted Scenario 1[f]	Unadjusted	Adjusted Base case[g]	Adjusted Scenario 1	Unadjusted	Adjusted Base case[e]	Unadjusted
	N = 12	ESS = 5.81	ESS = 11.54	N = 10	ESS = 6.64	ESS = 10	N = 22	ESS = 14.26	N = 36
Any prior tumour directed surgery for treatment of primary disease, n (%)	7 (58.3)	6.83 (83.3)	NR (57.8)	6 (60.0)	6.48 (83.3)	6 (60.0)	12 (54.5)	NR	30 (83.3)
Prior chemotherapy, n (%)	12 (100)	NR	NR	10 (100)	NR	NR	NR	NR	31 (86.1)
Prior targeted therapy, n (%)	NR	NR	NR	NR	NR	NR	NR	NR	4 (11.1)
Prior radiotherapy, n (%)	NR	NR	NR	NR	NR	NR	NR	NR	1 (2.8)
Pre-operative staging, n (%)									
Localised disease	10 (83.3)	7.54 (87.6)	9.53 (81.0)	9 (90.0)	6.92 (88.9)	9 (90.0)	16 (72.7)	NR	NR
Disseminated/ metastatic disease	1 (8.3)	0.29 (3.57)	1.33 (11.3)	NA	NA	NA	4 (18.2)	NR	NR
Leptomeningeal spread	1 (8.3)	0.545 (6.7)	0.913 (7.7)	1 (10.0)	0.865 (11.1)	1 (10.0)	2 (9.1)	NR	NR
Post-operative staging[j], n (%)									
Sub-total resection	6 (50.0)	3.20 (37.1)	5.82 (49.4)	6 (60.0)	4.82 (61.9)	6 (60.0)	9 (40.9)	NR	NR
Biopsy only, not attempted	6 (50.0)	5.21 (64.0)	5.95 (50.6)	4 (40.0)	2.97 (38.1)	4 (40.0)	13 (59.1)	NR	NR

Study reference/ID Characteristics Category	FIREFLY-1 tovorafenib								Bouffet 2023 dabrafenib and trametinib
	RANO-LGG evaluable set[a] Informs MAIC analysis of PFS			RANO-HGG evaluable set[b] Informs MAIC analysis of ORR			Safety analysis set[c] Informs MAIC analysis of all safety outcomes		Efficacy and safety analysis set[d]
	Unadjusted	Adjusted Base case[e]	Adjusted Scenario 1[f]	Unadjusted	Adjusted Base case[g]	Adjusted Scenario 1	Unadjusted	Adjusted Base case[e]	Unadjusted
	N = 12	ESS = 5.81	ESS = 11.54	N = 10	ESS = 6.64	ESS = 10	N = 22	ESS = 14.26	N = 36
Baseline IRC SPPD (mm^2) by RANO-HGG, mean (SD)	842.7 (734.7)	702.0 (NR)	934.0 (NR)	842.7 (734.7)	629 (728.5)	851 (732.7)	842.7 (734.7)	NR	NR
BRAF alteration, n (%)									
BRAF V600E mutation	12 (100)	NR[k]	NR[k]	10 (100)	NR[k]	NR[k]	22 (100)	NR	36 (100)
Baseline Karnofsky/ Lansky performance score[l], n (%)									
50–70	1 (8.3)	0.16 (2.8)	1.33 (11.3)	1 (10.0)	0.218 (2.8)	1 (10.0)	1 (4.5)	NR	1 (2.8)[m]
80–100	11 (91.7)	5.64 (97.2)	10.4 (88.7)	9 (90.0)	7.57 (97.2)	9 (90.0)	21 (95.5)	NR	35 (97.2)
Time since diagnosis [months], median [min; max]	NR	NR	NR	NR	NR	NR	NR	NR	40.2 [3.4; 123.8]
LGG, n (%)	12 (100)	NR	NR	10 (100)	NR	NR	22 (100)	NR	34 (94.4)
HGG, n (%)	0 (0)	NR	NR	0 (0)	NR	NR	0 (0)	NR	2 (5.6)
Treatment discontinuation, n (%)	10 (83.3)	7.5 (92.1)	9.81 (83.4)	8 (80.0)	7.5 (92.1)	8 (80.0)	15 (68.2)	NR	NR
Study discontinuation, n (%)	2 (16.7)	0.867 (10.0)	2.40 (20.4)	2 (20.0)	0.867 (10)	2.03 (20.3)	5 (22.7)	NR	NR

Study reference/ID Characteristics Category	FIREFLY-1 tovorafenib								Bouffet 2023 dabrafenib and trametinib
	RANO-LGG evaluable set[a] Informs MAIC analysis of PFS			RANO-HGG evaluable set[b] Informs MAIC analysis of ORR			Safety analysis set[c] Informs MAIC analysis of all safety outcomes		Efficacy and safety analysis set[d]
	Unadjusted	Adjusted Base case[e]	Adjusted Scenario 1[f]	Unadjusted	Adjusted Base case[g]	Adjusted Scenario 1	Unadjusted	Adjusted Base case[e]	Unadjusted
	N = 12	ESS = 5.81	ESS = 11.54	N = 10	ESS = 6.64	ESS = 10	N = 22	ESS = 14.26	N = 36

a: Analysis set for: PFS based on RANO LGG 2017 criteria (IRC)
b: Analysis set for: ORR based on RANO HGG 2010 criteria (INV, IRC)
c: Analysis set for: All submitted safety outcomes
d: Analysis set for: All submitted efficacy and safety outcomes
e: Reported by the HTD as adjusted for age, prior surgery, prior radiotherapy, Karnofsky/Lansky score, gender (male). Prior radiotherapy and gender are reported as being adjusted for in the analysis of PFS, but not ORR, though the validity of this is unclear due to the reporting of just one ESS for this dataset
f: Reported by the HTD as adjusted for age only in the case of ORR, but as adjusted for age and prior radiotherapy in the case of PFS. The validity of this is unclear due to the reporting of the same ESS for both analyses, which is implausible if different variables have been adjusted for.
g: Reported by the HTD as adjusted for age, prior surgery, Karnofsky/Lansky score.
h: numbers refer to the age group ≥ 12 years
i: Only patients with height and weight percentile available per Centre for Disease Control and Prevention standard growth chart are included in the summary. The HTD did not specify the exact number of patients included here.
j: At the time of the initial diagnosis.
k: The HTD reported identical (weighted) patient numbers for the unadjusted and adjusted analysis, which is implausible. The weighted patient numbers are therefore not presented in the table. However, as the proportion of participants with a BRAF V600E mutation was 100% in the unadjusted analysis, the same must also be true in the adjusted analysis.
l: Denominators for Karnofsky performance score and Lansky performance score summaries are the number of patients whose ages at study enrolment were ≥ 16 years and < 16 years, respectively
m: This only includes score of 70 in Bouffet 2023, no patients had lower scores

BSA: body surface area; BRAF: v-Raf murine sarcoma viral oncogene homolog B; INV: investigator assessed; IRC: independent radiology review committee; HGG: high-grade glioma; LGG: low grade glioma; max: maximum; min: minimum; N: number of treated patients; n: number of patients in category; NA: not applicable; NR: not reported; ORR: objective response rate; PFS: progression-free survival; RANO-HGG: Response Assessment in Neuro-Oncology for High-Grade Glioma; SPPD: sum of products of the perpendicular diameters; SD: standard deviation

4.3.2.3 Health outcome results for PICO 5 and uncertainties in the results

Within the given patient population, results on health outcomes describing relative effectiveness and relative safety shall be described by PICO. The section shall start from describing and justifying the choice of evidence (type of comparison) submitted to address the given PICO <1>.
The section shall present an overview of the available outcomes requested in the assessment scope per study.
Results on relative effectiveness and relative safety (i.e. the relative effects of the medicinal product versus the comparator) shall include the results from all individual studies, as well as any quantitative syntheses of results, e.g. from meta-analyses.
The results of the analyses of each of the presented outcomes shall be described briefly.
The description shall address any issues affecting the degree of uncertainty of the relative effects taking into account, if available, the methodological guidance adopted by the HTACG pursuant to Article 3(7), point (d), of the HTAR.
For any additional PICO question related to a given patient population, a new sub-section presenting the results in terms of health outcomes for this PICO question shall be added.

In the absence of direct comparative data to inform PICO 5 or a connected network of evidence which would allow anchored indirect comparative methods, an unanchored MAIC is used to estimate relative effectiveness and relative safety.

4.3.2.3.1 Overview of available outcomes

An overview of the available outcomes for which relative effect estimates were requested in the assessment scope is presented in Table 25. Table 25 further indicates for which outcomes results of relative effectiveness and safety analysis are presented in the assessment report. Results for the majority of the requested outcomes are not provided by the HTD, due to the absence of available data for dabrafenib + trametinib. Results for some outcomes provided in the HTD dossier are not presented in the assessment report due to methodological flaws and inappropriate analysis methods. Outcomes, for which descriptive summary data were requested in the assessment scope, are included in Appendix C.2.3.

Table 25: Matrix of outcomes in the included studies for Population 2 (BRAF V600E mutation in patients > 1 year) – PICO 5 – unanchored matching-adjusted indirect comparison: tovorafenib vs. dabrafenib + trametinib

Outcomes requested **Operationalisation**	**Outcome reported in the study (Yes/No)**		**Results included in the JCA report (Yes/No)** ***All results are based on unanchored MAIC methods***
	Tovorafenib FIREFLY-1	**Dabrafenib + trametinib Bouffet 2023**	
Safety outcomes			
Any AE	Yes	Yes	Yes[a]
Serious AE	Yes	Yes	Yes
Severe AE [CTCAE Grade≥ 3]	Yes	Yes	Yes
Death related to AE	Yes	Yes	Yes[b]
Treatment discontinuation due to AE	Yes	Yes	Yes
Treatment interruption due to AE	Yes	Yes	Yes
OS	Yes	No	No
PFS			
RANO-LGG (IRC)	Yes	Yes	Yes[c]
RAPNO (IRC)	Yes	No	No
RANO-HGG (IRC)	Yes	No	No
RANO-HGG (INV)	Yes	Yes	No[d]
Objective response			
RANO-LGG (IRC)	Yes	No[e]	No[f]
Modified RANO-LGG (IRC)[g]	No[h]	Yes	No
RAPNO (IRC)	Yes	No	No
RANO-HGG (IRC)	Yes	Yes	Yes
RANO-HGG (INV)	Yes	Yes	Yes
HRQoL			
Generic	No[i]	No	No
Disease-specific	No[i]	No	No
Symptoms of the disease			
impaired vision	Yes[j]	No	No
motor function	No[k]	No	No
neurological function; impaired balance and hearing; epileptic seizures; severe haemorrhagic events	No	No	No
General and cognitive fatigue	No[l]	No	No
Symptomatic disease control[m]	No	No	No

Outcomes requested **Operationalisation**	**Outcome reported in the study (Yes/No)**		**Results included in the JCA report (Yes/No)** ***All results are based on unanchored MAIC methods***
	Tovorafenib FIREFLY-1	**Dabrafenib + trametinib Bouffet 2023**	
Composite of CR, PR, or SD lasting a minimum of (i) 6 and (ii) 12 months			
RANO-LGG (IRC)	Yes[n]	No[e]	No
RAPNO (IRC)	Yes[n]	No	No
RANO-HGG (IRC)	Yes[n]	No	No
RANO-HGG (INV)	Yes[n]	No[o]	No
General and cognitive fatigue	No[j]	No	No

a: Descriptive results are included for 'Any AE', a relative effect estimate was not feasible due to the incidence rate of 100% in both study groups
b: Descriptive results are included for 'Death related to AE', a relative effect estimate was not provided due to zero events in Bouffet 2023
c: PFS, as a time-to-event outcome, was provided by the HTD and relative effectiveness results are included in the JCA. Rates of PFS at 6- and 12- months are also requested in the assessment scope, and provided by the HTD, but relative effectiveness results were not provided by the HTD. Relative effectiveness results were calculated by the assessors.
d: Not provided in the dossier
e: This outcome is identified in the relevant table of relative effectiveness results in the HTD dossier as being reported in the Bouffet 2023 study. However, the Bouffet 2023 study used modified RANO-LGG (IRC) criteria that excluded the MR category, classifying such patients as SD.
f: Relative effectiveness results for IRC-assessed ORR based on RANO-LGG criteria are not included in the JCA report due to clear bias arising from methodological flaws and inappropriate analysis methods used by the HTD
g: Modified RANO-LGG (IRC) criteria, in the Bouffet 2023 study, excluded the MR category, classifying such patients as SD
h: ORR based on modified RANO-LGG (IRC) criteria (including CR and PR, excluding MR) was not explicitly reported in the dossier, but the individual components of this outcome (i.e. CR and PR) are available from FIREFLY-1 and were reported in the dossier.
i: Generic and disease-specific HRQoL were assessed in FIREFLY-1 by PedsQL 4.0 Generic Core Scales and the PedsQL 3.0 Cancer Module domains and global scale as well as the PROMIS Pediatric/Parent Proxy Profile 49 (or 57 for patients greater than 17 years of age). According to the HTD results for these outcomes are not available for Population 2.
j: In FIREFLY-1 visual acuity worsening was assessed as an increase of ≥ 0.2 Logarithm of the Minimum Angle of Resolution (logMAR) (corrected for age).
k: In FIREFLY-1, the Vineland-3 Motor Skills questionnaire was planned to be administered to patients with baseline motor function deficits at US sites only. Results for this outcome are not available in the dossier for Population 2.
l: In FIREFLY-1, fatigue was assessed as one domain of the PROMIS Pediatric/Parent Proxy Profile 49 (or 57 for patients greater than 17 years of age). In addition, a PROMIS cognitive function short form was also administered. Results for these outcomes are not available in the dossier for Population 2.
m. Composite outcomes, comprising stable and improved symptoms are not assessed in FIREFLY-1 or Bouffet 2023.
n: The HTD provided data for a composite outcome of CR+PR, +MR (where applicable) in the dossier, labelling this data as "CBR (CR+PR+SD)". However, according to the CSR none of the patients in Population 2 had an SD with a duration of more than 12 month.

Outcomes requested Operationalisation	Outcome reported in the study (Yes/No)		Results included in the JCA report (Yes/No) *All results are based on unanchored MAIC methods*
	Tovorafenib FIREFLY-1	Dabrafenib + trametinib Bouffet 2023	
o: Although the outcomes of CR, PR and SD were reported in Bouffet 2023, the required component 'SD of at least 6/12 months' duration' of the specified composite outcome was not.			
AE: adverse event; CR: complete response; CTCAE: common terminology criteria for adverse events; HGG: high grade glioma; HRQoL: health-related quality of life; INV investigator assessed; IRC: independent radiology review committee; JCA: Joint Clinical Assessment; LGG: low-grade glioma; OS: overall survival; PR: partial response; PFS: progression-free survival; PROMIS: Patient-Reported Outcomes Measurement Information System; PedsQL: Paediatric Quality of Life Inventory; RANO: Response Assessment in Neuro-oncology; RAPNO: Response Assessment in Paediatric Neuro-oncology; SAE: serious adverse event; SD: stable disease			

4.3.2.3.2 Results on relative effectiveness and safety

Results on relative effectiveness are shown in Table 26 and Table 27. Results on relative safety are shown in Appendix C.3. All reported effect-estimates target the population-average treatment effect in the population of the Bouffet 2023 study (as the MAIC method targets a marginal estimand in this population).

For all the results on relative effectiveness and safety included in the dossier, major limitations of the available evidence mean that the reported effect estimates may not represent true causal effects of treatment and may not be sufficiently applicable to the target population of PICO 5. In particular, the following major, general uncertainties should be considered when interpreting these results:

- As described in Section 4.3.2.2, the absence of direct comparative data for Population 2 or a connected network of evidence which would allow anchored indirect comparative methods necessitated the use of unanchored ITCs to estimate relative effects for all outcomes. These methods are a type of non-randomised evidence and, as such, are associated with an inherently greater risk of bias than comparisons based on randomised evidence.
- As described in Section 4.3.2.1, there are some differences between the Bouffet 2023 population and that of Population 2 specified in the assessment scope (and reflected in the subpopulation of FIREFLY-1 in PICO 5), including small proportions of patients in the Bouffet 2023 study with a diagnosis of HGG (rather than paediatric LGG) or who had not received prior systemic therapy. These differences could not be adjusted for in the MAIC due to the absence of IPD or appropriate subgroup data from Bouffet 2023. The resulting effect-estimates may be biased as a result of these differences in population.

- As described in Section 4.3.2.2, all treatment effect estimates are associated with a risk of bias of unknown magnitude and direction due to residual confounding. This is likely to arise from identified confounders that were not adjusted for in the MAIC, including histological type, primary tumour location and prior chemotherapy and may also arise from other confounders that were not identified due to shortcomings in the HTD's confounder identification process.
- As described in Section 4.3.2.2, the ESS for the analyses of all outcomes is small, leading to statistical imprecision (i.e., wide confidence intervals) for the resulting effect estimates. Furthermore, the ESS in the base case analysis represents a reduction of 34%-52%, depending on the outcome, from the original sample size. This may indicate poor covariate overlap between studies, potentially leading to bias and invalid standard errors (and thus invalid confidence intervals) [30-32]. These issues also threaten the applicability of the results to the target population of PICO 5.
- Sensitivity analysis conducted by the HTD to investigate the possible impact of unmeasured confounding (shifted null hypothesis testing and calculation of E-values) included a number of errors and the reported conclusions were inaccurate. For this reason, the results of this sensitivity analysis are not presented in the JCA report.
- As described in Section 4.3.2.2.2.3, all reported p-values are nominal, do not represent the result of prespecified hypothesis testing and are not controlled for multiplicity.

Table 26: Relative effectiveness results (dichotomous outcomes) for Population 2 (BRAF V600E mutation in patients > 1 year) – PICO 5 – unanchored matching-adjusted indirect comparison: tovorafenib vs. dabrafenib + trametinib

Intervention	**Tovorafenib**		**Dabrafenib + trametinib**		**Indirect comparison method**	**Tovorafenib vs. dabrafenib + trametinib**	
Study	**FIREFLY-1**		**Bouffet 2023**				
Data-cut	**2 year**		**NR**				
Outcome	**Events (sample size)**	**Rate (%) [95%-CI]**	**Events (sample size)**	**Rate (%) [95%-CI]**		**ESS (FIREFLY-1)**	**OR[a]/RR[b] [95%-CI] p-value[c]**
OS							
Relative effectiveness results not available							
PFS							
6-month PFS rate[d] by RANO-LGG criteria, IRC-assessment	NR	94 [85, 100]	NR (36)	86 [75, 98]	MAIC base case: adjusted for age, prior surgery, prior radiotherapy, Karnofsky/Lansky score, gender	5.81	RR: 1.09 [0.92; 1.30] –
		80 [61, 100]			MAIC scenario 1: adjusted for age and prior radiotherapy	11.54	RR: 0.93 [0.69, 1.26] –
		87 [73, 100]			MAIC scenario 2: adjusted for age, Karnofsky/Lanky score and prior radiotherapy	10.66	RR: 1.01 [0.81, 1.26] –
		83 [65, 100]			Naïve comparison	12	RR: 0.97 [0.73, 1.28] –
12-month PFS rate[d] by RANO-LGG criteria, IRC-assessment	NR	76 [52, 100]	NR (36)	83 [71, 96]	MAIC base case: adjusted for age, prior surgery, prior radiotherapy, Karnofsky/Lansky score, gender	5.81	RR: 0.92 [0.61, 1.38] –
		56 [34, 92]			MAIC scenario 1: adjusted for age and prior radiotherapy	11.54	RR: 0.67 [0.40, 1.14] –
		63 [41, 97]			MAIC scenario 2: adjusted for age, Karnofsky/Lanky score and prior radiotherapy	10.66	RR: 0.76 [0.48, 1.20] –

Intervention	**Tovorafenib**		**Dabrafenib + trametinib**		**Indirect comparison method**	**Tovorafenib vs. dabrafenib + trametinib**	
Study	**FIREFLY-1**		**Bouffet 2023**				
Data-cut	**2 year**		**NR**				
Outcome	**Events (sample size)**	**Rate (%) [95%-CI]**	**Events (sample size)**	**Rate (%) [95%-CI]**		**ESS (FIREFLY-1)**	**OR[a]/RR[b] [95%-CI] p-value[c]**
		58 [36, 94]			Naïve comparison	12	RR: 0.70 [0.42, 1.15] –
Objective response							
Composite of CR+PR by RANO-HGG criteria, INV-assessment	NR	NR	19 (36)	52.8 [NR, NR]	MAIC base case: adjusted for age, prior surgery, Karnofsky/Lansky score	6.64	OR: 0.56 [0.08, 3.84] p = 0.551
	NR	NR			MAIC scenario 1: adjusted for age only	10	OR: 0.60 [0.13, 2.83] p = 0.515
	4 (10)	40.0 [12.2, 73.8]			Naïve comparison	10	OR: 0.60 [0.14, 2.48] p = 0.477
Composite of CR+PR by RANO-HGG criteria, IRC-assessment[e]	NR	NR	7 (36)	19.4 [8.2, 36.0]	MAIC base case: adjusted for age, prior surgery, Karnofsky/Lansky score.	6.64	OR: 7.26 [0.98, 53.68] p = 0.052
	NR	NR			MAIC scenario 1: adjusted for age only	10	OR: 4.16 [0.83, 21.01] p = 0.084
	5 (10)	50.0 [18.7, 81.3]			Naïve comparison	10	OR: 4.41 [0.93, 18.37] p = 0.061
HRQoL, generic							
Relative effectiveness results not available							
HRQoL, disease-specific							
Relative effectiveness results not available							
Symptoms of the disease							
Relative effectiveness results not available							

Intervention	Tovorafenib		Dabrafenib + trametinib		Indirect comparison method	Tovorafenib vs. dabrafenib + trametinib	
Study	FIREFLY-1		Bouffet 2023				
Data-cut	2 year		NR				
Outcome	Events (sample size)	Rate (%) [95%-CI]	Events (sample size)	Rate (%) [95%-CI]		ESS (FIREFLY-1)	OR[a]/RR[b] [95%-CI] p-value[c]
Symptomatic disease control							
Relative effectiveness results not available							
Composite outcome comprising of CR, PR, or SD lasting a minimum of (i) 6 and (ii) 12 months							
Relative effectiveness results not available							
General and cognitive fatigue							
Relative effectiveness results not available							

a: Values of OR > 1 indicate higher ORR with tovorafenib
b: Calculated by assessors from Kaplan-Meier estimated probabilities per arm given in the dossier (see Appendix C.2.4). Values of RR > 1 indicate higher probabilities of remaining alive and progression-free with tovorafenib.
c: No formal hypothesis testing was described in the evidence synthesis SAP nor in the relevant sections of the dossier. Therefore, all p-values should be interpreted as nominal.
d: Calculated as the Kaplan-Meier probability estimate of PFS (probability of remaining alive and progression-free) at the relevant timepoint.
e. additional information on this assessment in Appendix D.2.2

CI: confidence interval; CR: complete response; HRQoL: health-related quality of life; INV: investigator; IRC: independent review committee; MAIC: matching-adjusted indirect comparison; NR: not reported; OR: odds ratio; ORR: objective response rate; PFS: progression-free survival; PR: partial response; RANO-HGG: Response Assessment in Neuro-Oncology criteria for high-grade gliomas; RANO-LGG: Response Assessment in Neuro-Oncology criteria for low-grade gliomas; RR: relative risk; SD: stable disease

Table 27: Relative effectiveness results (time-to-event outcomes) for Population 2 (BRAF V600E mutation in patients > 1 year) – PICO 5 – unanchored matching-adjusted indirect comparison: tovorafenib vs. dabrafenib + trametinib

Intervention	**Tovorafenib**		**Dabrafenib + trametinib**		**Indirect comparison method**	**Tovorafenib vs. dabrafenib + trametinib**		
Study	**FIREFLY-1**		**Bouffet 2023**					
Data-cut	**2 year**		**Study completion**					
Outcome	**Sample size (number censored)**	**Median TTE in months [95%-CI]**	**Sample size (number censored)**	**Median TTE in months [95%-CI]**		**ESS (FIREFLY-1)**	**HR[a] [95%-CI] p-value[c]**	**RMST difference[b] [95%-CI] p-value[c]**
PFS								
PFS by RANO-LGG criteria, IRC-assessment	NR	NR	36 (27)	36.9 [36.0, NE]	MAIC base case: adjusted for age, prior surgery, prior radiotherapy, Karnofsky/Lansky score, gender	5.81	4.88 [2.14, 11.14] p = 0.011	-6.64 [-13.16, -0.12] p = 0.046
	NR	NR			MAIC scenario 1: adjusted for age and prior radiotherapy	11.54	5.44 [2.26, 13.05] P = 0.001	-9.51 [-15.10, -3.92] p = 0.002
	NR	NR			MAIC scenario 2: adjusted for age, Karnofsky/Lanky score and prior radiotherapy	10.66	5.22 [2.24, 12.18] p < 0.001	-7.77 [-13.25, -2.29] p = 0.004
	12 (1)	13.67 [4.60, 24.87]			Naïve comparison	12	5.09 [2.03, 12.80] P < 0.001	-8.35 [-13.98, -2.72] p = 0.004

a: Hazard ratios less than 1 indicate longer PFS with tovorafenib
b: Positive values indicate greater RMST with tovorafenib
c: No formal hypothesis testing was described in the dossier nor in the evidence synthesis SAP. All p-values are nominal, have not be obtained through prespecified analysis and are not controlled for multiplicity.

CI: confidence interval; ESS: effective sample size; HR: hazard ratio; IRC: independent review committee; MAIC: matching-adjusted indirect comparison.NE: not estimable; NR: not reported; PFS: progression-free survival; RANO-LGG: Response Assessment in Neuro-Oncology criteria for low-grade gliomas; RMST: restricted mean survival time; TTE: time-to-event

ORR

Relative effectiveness results for the outcome of ORR are shown in Table 26. Results showed numerically higher ORR with dabrafenib+trametinib in the INV-assessed analysis, and with tovorafenib in the IRC-assessed analysis, though in both cases, confidence intervals for the effect estimates were wide and contained the null. While results of the comparisons were reported by the HTD as not statistically significant based on a nominal 5% alpha level, no formal hypothesis testing was described and analyses were not prespecified nor controlled for multiplicity, which limits the interpretability of the reported p-values. Additional issues affecting certainty of evidence for the ORR outcome specifically include the following:

- Relative effectiveness results for IRC-assessed ORR based on RANO-LGG criteria are not included in Table 26 due to clear bias arising from methodological flaws and inappropriate analysis methods used by the HTD, as described in 4.3.2.2.2.2. In the analysis of IRC-assessed ORR by RANO-LGG presented by the HTD, different definitions of objective response are used for the intervention and comparator arms, with the MR category classified as an objective response in FIREFLY-1 but as a non-response in Bouffet 2023. This results in an unclear target estimand for the comparison of ORR, with two possible interpretations (i.e., the effect of treatment on ORR with and without MR). Regardless of which interpretation is adopted, the inconsistency in definitions between studies leads to bias in favour of tovorafenib.
- Relative effectiveness results for ORR presented in Table 26 are based on criteria designed for HGG and not specifically for LGG (see Section 4.2.1.3).
- There is a substantial difference between INV- and IRC-assessed ORR by RANO-HGG in Bouffet 2023. The reason for this difference is unclear. The corresponding difference in the direction and magnitude of relative treatment effects for these outcomes make the interpretation of the ORR results highly uncertain.
- Data for the outcome of INV-assessed ORR by RANO-HGG are available from Arm 2 of FIREFLY-1 including n = 9 patients meeting the inclusion criteria for Population 2, but the HTD declined to include these in the indirect comparison. Inadequate justification for the exclusion of these data was provided by the HTD.
- Additional identified PVs/EMs were excluded from the adjustment for the MAIC of ORR by RANO-HGG, specifically gender (excluded due to non-convergence of the weighting model) and prior radiotherapy (reason for exclusion not reported by the HTD). This further increases the risk of bias due to unmeasured confounding.

PFS

Relative effectiveness results for the outcome of PFS are shown in Table 26 and Table 27. The results for the time-to-event outcome show increased PFS in favour of dabrafenib + trametinib

over tovorafenib when measured both on the HR scale and on the RMST scale. While this comparison was reported by the HTD as statistically significant based on a nominal 5% alpha level, no formal hypothesis testing was described and analyses were not prespecified nor controlled for multiplicity, which limits the interpretability of the reported p-values. Rates of PFS at 6 and 12 months were similar between treatment groups in the base case MAIC. In addition to the general uncertainties in the evidence listed previously, issues affecting certainty of evidence for the PFS outcome specifically include the following:

- The HTD has reported that the PH assumption is unlikely to hold; however, results of the assessment of PH were not provided and therefore the assessors cannot comment on the appropriateness of this conclusion. If the PH assumption is violated, then the HR may not be an appropriate measure of treatment effect and the difference in RMST may be a more reliable summary of the corresponding relative effect.
- The definitions of PFS (including censoring rules) and assessment schedules were broadly similar between FIREFLY-1 and Bouffet 2023, although a breakdown of number of censored observations by censoring reason was not provided in Bouffet 2023. The possibility of bias due to informative censoring and/or differences in censoring mechanisms between studies cannot be entirely excluded.
- Data on PFS by INV-assessed RANO-HGG criteria were available from both FIREFLY-1 and, in the form of published Kaplan-Meier curves with numbers at-risk, also from Bouffet 2023, but results of the corresponding indirect comparison were not provided by the HTD. There is, therefore, a risk of bias due to selective reporting of the PFS outcome. In addition, analyses for rates of PFS at 6 and 12 months (binary outcomes) were also requested in the assessment scope. Relative effectiveness analyses were not provided by the HTD, although data on 6- and 12-month PFS rates in both treatment groups were reported. However, for completeness, the assessors calculated relative risks for PFS at these time points (for details see Appendix C.2.4).

Safety

Descriptive safety outcomes are shown in Table 28, while results for the analysis of relative safety are presented in Appendix C.3. It was not possible to generate a relative effect estimate for the outcomes of “Any AE” due to the incidence rate of 100% in both study groups. Direct calculation of a relative effect estimate (i.e., without applying a continuity correction or other adjustment) for “death related to AE” was not possible due to zero events in Bouffet 2023 and is not deemed necessary due to low overall event numbers and sample size. The incidence of safety outcomes may be affected by differences in the duration of drug exposure between studies, which could not be adjusted for in the analysis and may negatively affect certainty of evidence for safety outcomes.

Table 28: Descriptive safety outcomes for Population 2 (BRAF V600E mutation in patients > 1 year) – PICO 5 – tovorafenib vs. dabrafenib + trametinib

Study reference/ID **Time point** **Outcome**	**Tovorafenib** **FIREFLY-1[a]**		**Dabrafenib + trametinib** **Bouffet 2023**	
	N	**Patients with event n (%)**	**N**	**Patients with event n (%)**
Data-cut	**2-year**		**Study Completion**	
Any AE	22	22 (100.0)	36	36 (100.0)
Serious AE	22	13 (59.1)	36	15 (41.7)
Severe AE [CTCAE]				
Grade ≥ 3	22	20 (90.9)	36	22 (61.1)
Grade 3	22	17 (77.3)	36	NR
Grade 4	22	1 (4.5)	36	NR
Grade 5	22	2 (9.1)	36	NR
Death related to AE	22	2 (9.1)	36	0 (0)
Treatment discontinuation due to AE	22	3 (13.6)	36	8 (22.2)
Treatment interruption due to AE	22	15 (68.2)	36	26 (72.2)
a: Safety data is based on Arm 1 + 2 of FIREFLY-1				
AE: adverse event; BRAF: v-Raf murine sarcoma viral oncogene homolog B; CTCAE: Common Terminology Criteria for Adverse Events; ID: identification; N: number of patients in the analysis; n: number of patients with at least 1 event; NR: not reported				

4.3.2.4 Health outcome results for PICO 6 and uncertainties in the results

An analysis of the relative effectiveness and safety of tovorafenib for PICO 6 was not provided by the HTD. See Section 3 for further details on reasons for the omission of results for this PICO.

4.3.3 Results for patient population 3 (BRAF fusion, rearrangement, or V600 [non-E] mutation)

This section shall discuss to which extent the included patient populations and/or comparator(s) per study cover the relevant patient population/comparator(s) according to the assessment scope as set out pursuant to Article 8(6) of the HTAR.
For each patient population specified in PICO(s), a separate section shall be provided. Within this section, the results for all PICO(s) addressing this patient population shall be presented in sub-sections.

As described in Section 3 and 4.1.1, the HTD includes evidence addressing the subpopulation of patients with a BRAF fusion, rearrangement or V600 [non-E] mutation (Population 3) and provides an unanchored MAIC of tovorafenib with trametinib for PICO 7. However, this comparison based on the studies FIREFLY-1 and TRAM-01 is not included in the JCA report due to insufficient information available for the assessment of TRAM-01.

In the results shown by the HTD for PICO 7, single-arm data on treatment with tovorafenib from FIREFLY-1 for Population 3 are included. These data are not included in the JCA report because they do not provide information on relative effectiveness or safety for PICOs 7 or 8 in the context of the current assessment.

In the HTD dossier, the information on these data is available in the following sections:

- Methods: Sections 4.3.1 and 4.3.2
- Patient characteristics: Section 5.3.3.1 (Table 63-65)
- Outcomes: Section 5.3.3.2.4 (Tables 68-70)

4.3.4 Results of the main study from the clinical development programme of the medicinal product (if not addressed by any of PICO(s))

Results from the FIREFLY-1 study, which are not addressed by the PICOs, are presented in Section 5.3.1.2.4 of the HTD dossier.

5 List of references

1. Roka K, Scheinemann K, Avula S et al. European Standard Clinical Practice Recommendations for primary Pediatric Low-grade Gliomas. EJC Paediatric Oncology 2024; 4: 100169. https://doi.org/10.1016/j.ejcped.2024.100169.

2. Ryall S, Tabori U, Hawkins C. Pediatric low-grade glioma in the era of molecular diagnostics. Acta Neuropathol Commun 2020; 8(1): 30. https://doi.org/10.1186/s40478-020-00902-z.

3. Ipsen Pharma. Sponsor's report on the maintenance of designation criteria at the time of MA or type II variation application for a designated orphan medicinal product - Ojemda [unpublished]. 2025.

4. Gnekow AK, Kandels D, Tilburg CV et al. SIOP-E-BTG and GPOH Guidelines for Diagnosis and Treatment of Children and Adolescents with Low Grade Glioma. Klin Padiatr 2019; 231(3): 107-135. https://doi.org/10.1055/a-0889-8256.

5. Zelt S, Cooney T, Yu S et al. Disease burden and healthcare utilization in pediatric low-grade glioma: A United States retrospective study of linked claims and electronic health records. Neurooncol Pract 2024; 11(5): 583-592. https://doi.org/10.1093/nop/npae037.

6. Novartis Europharm. Finlee 10 mg dispersible tablets [online]. 11.2025 [Accessed: 05.11.2025]. URL: https://www.ema.europa.eu/en/documents/product-information/finlee-epar-product-information_en.pdf.

7. Novartis Europharm. Spexotras 0.05 mg/ml powder for oral solution [online]. 11.2025 [Accessed: 05.11.2025]. URL: https://www.ema.europa.eu/en/documents/product-information/spexotras-epar-product-information_en.pdf.

8. Member State Coordination Group on Health Technology Assessment. Questions and Answers on general methodological and procedural issues for joint clinical assessments [online]. 2025 [Accessed: 28.01.2026]. URL: https://health.ec.europa.eu/publications/questions-and-answers-general-methodological-and-procedural-issues-joint-clinical-assessments_en.

9. Ipsen Pharma. FIREFLY-1 (DCO 10 May 2024); Clinical Study Report [unpublished]. 2025.

10. Day One Biopharmaceuticals. A Study to Evaluate Tovorafenib in Pediatric and Young Adult Participants With Relapsed or Progressive Low-Grade Glioma and Advance Solid Tumors (FIREFLY-1) [online]. 2025 [Accessed: 18.11.2025]. URL: https://clinicaltrials.gov/study/NCT04775485.

11. DOT Therapeutics-1 (Day One). A Phase 2, Open-Label, Multicenter Study to Evaluate the Safety and Efficacy of the Oral Pan-RAF Inhibitor DAY101 in Pediatric Patients with BRAF-Altered, Recurrent or Progressive Low-Grade Glioma [online]. [Accessed: 18.11.2025]. URL: https://www.clinicaltrialsregister.eu/ctr-search/search?query=2020-003657-30.

12. Kilburn LB, Khuong-Quang DA, Hansford JR et al. The type II RAF inhibitor tovorafenib in relapsed/refractory pediatric low-grade glioma: the phase 2 FIREFLY-1 trial. Nat Med 2024; 30(1): 207-217. https://doi.org/10.1038/s41591-023-02668-y.

13. Novartis Pharmaceuticals. Study to Investigate Safety, Pharmacokinetic (PK), Pharmacodynamic (PD) and Clinical Activity of Trametinib in Subjects With Cancer or Plexiform Neurofibromas and Trametinib in Combination With Dabrafenib in Subjects With Cancers Harboring V600 Mutations [online]. 2021 [Accessed: 18.11.2025]. URL: https://clinicaltrials.gov/study/NCT02124772.

14. Novartis Pharma Services. An Open-Label, Dose-Escalation, Phase I/II Study to Investigate the Safety, Pharmacokinetics, Pharmacodynamics and Clinical Activity of the MEK Inhibitor Trametinib in Children and Adolescents Subjects with Cancer or Plexiform Neurofibromas and Trametinib in Combination with Dabrafenib in Children and Adolescents with Cancers Harboring V600 mutations [online]. [Accessed: 18.11.2025]. URL: https://www.clinicaltrialsregister.eu/ctr-search/search?query=2013-003596-35.

15. Bouffet E, Geoerger B, Moertel C et al. Efficacy and Safety of Trametinib Monotherapy or in Combination with Dabrafenib in Pediatric BRAF V600-Mutant Low-Grade Glioma. J Clin Oncol 2023; 41(3): 664-674. https://doi.org/10.1200/JCO.22.01000.

16. Witt O, Chi SN, Kang HJ et al. Post hoc analysis of rashes reported in patients (pts) with BRAF-altered relapsed/refractory (r/r) pediatric low-grade glioma (pLGG) treated with the type II RAF inhibitor tovorafenib in FIREFLY-1. J Clin Oncol 2025; 43(16 Suppl): 10037. https://doi.org/10.1200/JCO.2025.43.16_suppl.10037.

17. Perreault S, Larouche V, Tabori U et al. A phase 2 study of trametinib for patients with pediatric glioma or plexiform neurofibroma with refractory tumor and activation of the MAPK/ERK pathway: TRAM-01. BMC Cancer 2019; 19(1): 1250. https://doi.org/10.1186/s12885-019-6442-2.

18. Perreault S, Larouche V, Tabori U et al. A phase 2 study of trametinib for patients with pediatric glioma or plexiform neurofibroma with refractory tumor and activation of the MAPK/ERK pathway: TRAM-01. J Clin Oncol 2022; 40: 2042. https://doi.org/10.1200/JCO.2022.40.16_suppl.2042.

19. Perreault S, Larouche V, Tabori U et al. TRAM-01: A phase 2 study of trametinib for patients with pediatric glioma with activation of the MAPK/ERK pathway. Neuro-Oncology Advances 2021; 3(Suppl 4): iv9. https://doi.org/10.1093/noajnl/vdab112.032.

20. St. Justine's Hospital. Trametinib for Pediatric Neuro-oncology Patients With Refractory Tumor and Activation of the MAPK/ERK Pathway [online]. 2023 [Accessed: 18.11.2025]. URL: https://www.clinicaltrials.gov/study/NCT03363217.

21. Member State Coordination Group on Health Technology Assessment. Guidance on filling in the joint clinical assessment (JCA) dossier template – Medicinal products [online]. 2024 [Accessed: 18.12.2025]. URL: https://health.ec.europa.eu/publications/guidance-filling-joint-clinical-assessment-jca-dossier-template-medicinal-products_en.

22. Novartis Pharmaceuticals. Pediatric Long-Term Follow-up and Rollover Study [online]. 2025 [Accessed: 12.09.2025]. URL: https://clinicaltrials.gov/study/NCT03975829.

23. Ipsen Pharma. OJEMDA 25 mg/ml powder for oral suspension [unpublished]. 2025.

24. Ipsen Pharma. OJEMDA 100 mg film-coated tablets [unpublished]. 2025.

25. Wen PY, Macdonald DR, Reardon DA et al. Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. J Clin Oncol 2010; 28(11): 1963-1972. https://doi.org/10.1200/JCO.2009.26.3541.

26. Wen PY, Chang SM, Van den Bent MJ et al. Response Assessment in Neuro-Oncology Clinical Trials. J Clin Oncol 2017; 35(21): 2439-2449. https://doi.org/10.1200/JCO.2017.72.7511.

27. Fangusaro J, Witt O, Hernaiz Driever P et al. Response assessment in paediatric low-grade glioma: recommendations from the Response Assessment in Pediatric Neuro-Oncology (RAPNO) working group. Lancet Oncol 2020; 21(6): e305-e316. https://doi.org/10.1016/S1470-2045(20)30064-4.

28. Chukwueke UN, Wen PY. Use of the Response Assessment in Neuro-Oncology (RANO) criteria in clinical trials and clinical practice. CNS Oncol 2019; 8(1): CNS28. https://doi.org/10.2217/cns-2018-0007.

29. van den Bent MJ, Cloughesy TF, Ellingson BM et al. The Use of Minor Response, Volumetric Assessment and Growth Rate Kinetics as Endpoints in Grade 1-3 Glioma Clinical Trials: a RANO perspective. Neuro Oncol 2025. https://doi.org/10.1093/neuonc/noaf173.

30. Phillippo DM, Dias S, Ades AE et al. Assessing the performance of population adjustment methods for anchored indirect comparisons: A simulation study. Stat Med 2020; 39(30): 4885-4911. https://doi.org/10.1002/sim.8759.

31. Phillippo DM, Dias S, Ades AE et al. Target estimands for efficient decision making: Response to comments on "Assessing the performance of population adjustment methods for anchored indirect comparisons: A simulation study". Stat Med 2021; 40(11): 2759-2763. https://doi.org/10.1002/sim.8965.

32. Liu Y, He X, Liu J et al. Is the Use of Unanchored Matching-Adjusted Indirect Comparison Always Superior to Naive Indirect Comparison on Survival Outcomes? A Simulation Study. Appl Health Econ Health Policy 2025; 23(4): 693-704. https://doi.org/10.1007/s40258-025-00952-1.

33. Guyot P, Ades AE, Ouwens MJ et al. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol 2012; 12: 9. https://doi.org/10.1186/1471-2288-12-9.

34. Signorovitch JE, Wu EQ, Yu AP et al. Comparative effectiveness without head-to-head trials: a method for matching-adjusted indirect comparisons applied to psoriasis treatment with adalimumab or etanercept. Pharmacoeconomics 2010; 28(10): 935-945. https://doi.org/10.2165/11538370-000000000-00000.

35. Member State Coordination Group on Health Technology Assessment. Practical Guideline for Quantitative Evidence Synthesis: Direct and Indirect Comparisons [online]. 2024 [Accessed: 18.12.2025]. URL: https://health.ec.europa.eu/publications/practical-guideline-quantitative-evidence-synthesis-direct-and-indirect-comparisons_en.

36. Phillippo DM, Ades AE, Dias S et al. NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submission to NICE [online]. 2016 [Accessed: 18.12.2025]. URL: https://sheffield.ac.uk/nice-dsu/tsds/population-adjusted.

37. Glasziou P, Chalmers I, Rawlins M et al. When are randomised trials unnecessary? Picking signal from noise. BMJ 2007; 334(7589): 349-351. https://doi.org/10.1136/bmj.39070.527986.68.

38. Horita N, Yamamoto S, Mizuki Y et al. Minimal Clinically Important Difference (MCID) of Effect Sizes other than Mean Difference. Journal of Clinical Question 2024; 1(3): 116-127. https://doi.org/10.69854/jcq.2024.0016.

39. Fernainy P, Sourial N. Leveraging confidence intervals to inform the clinical significance of interventions. Fam Pract 2022; 39(6): 1196-1198. https://doi.org/10.1093/fampra/cmac065.

40. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. Ann Intern Med 2017; 167(4): 268-274. https://doi.org/10.7326/M16-2607.

41. Day One Biopharmaceuticals. Tovorafenib (DAY101) Monotherapy for Patients With Melanoma and Other Solid Tumors [online]. 2025 [Accessed: 29.10.2025]. URL: https://clinicaltrials.gov/study/NCT04985604.

42. Millennium Pharmaceuticals. Study of MLN2480 in Participants With Relapsed or Refractory Solid Tumors Followed by a Dose Expansion in Participants With Metastatic Melanoma (C28001) [online]. 2020 [Accessed: 04.11.2025]. URL: https://clinicaltrials.gov/study/NCT01425008.

43. Millennium Pharmaceuticals. A Study MLN2480 in Combination With MLN0128 or Alisertib, or Paclitaxel, or Cetuximab, or Irinotecan in Adult Participants With Advanced Nonhematologic Malignancies (C28002) [online]. 2020 [Accessed: 04.11.2025]. URL: https://clinicaltrials.gov/study/NCT02327169.

44. Wright KD. DAY101 In Gliomas and Other Tumors [online]. 2025 [Accessed: 04.11.2025]. URL: https://www.clinicaltrials.gov/study/NCT03429803.

45. Leclercq E, Leeflang MM, van Dalen EC et al. Validation of search filters for identifying pediatric studies in PubMed. J Pediatr 2013; 162(3): 629-634.e2. https://doi.org/10.1016/j.jpeds.2012.09.012.

46. Gemeinsamer Bundesausschuss. [Benefit assessment procedure for the active ingredient dabrafenib (malignant glioma, BRAF V600E mutation, ≥ 1 year, low-grade (LGG)/high-grade (HGG) after at least 1 prior therapy; combination with trametinib); Module 4A, Appendix 4-H] [online]. 2024 [Accessed: 15.12.2025]. URL: https://www.g-ba.de/downloads/92-975-7639/2024_04_25_Modul4A_Dabrafenib_Anhang_4_H.pdf.

47. Therneau T. A package for survival analysis in R [online]. 2024 [Accessed: 18.12.2025]. URL: https://cran.r-project.org/web/packages/survival/vignettes/survival.pdf.

Appendix A Input from experts and stakeholders

A.1 Carer input on the consolidated assessment scope proposal

JCA 202406

Consolidated assessment scope proposal

Pediatric low-grade glioma (LLG) is increasingly recognized as a chronic condition, often requiring children to undergo multiple lines of treatment over several years. This prolonged therapeutic journey exposes young patients to a wide range of side effects, which can significantly impact their quality of life. In this context, innovative therapies like Tovorafenib offer a promising alternative, potentially improving both treatment outcomes and the overall survival of affected children.

Tovorafenib presents advantages over other targeted therapies, such as MEK inhibitors, by acting upstream in the MAPK pathway. This upstream action may result in more effective inhibition of cell proliferation.

Given these factors, it is both timely and necessary to evaluate the efficacy of Tovorafenib in comparison to existing standard-of-care treatments for pediatric LLG. Traditional chemotherapies such as carboplatin, vincristine, and the TPCV regimen have been associated with long-term toxicities, including the risk of secondary malignancies. Therefore, the introduction of a targeted therapy with a potentially more favorable safety profile is a welcome development.

Within this framework:

- **PICO 1** is designed to compare Tovorafenib against all currently available treatment options for pediatric LLG, excluding proton therapy. This broad comparison is essential to assess its relative effectiveness and impact on health-related quality of life (HrQoL).
- **PICO 2** focuses on the first four lines of treatment. However, it raises the question of why combinations such as everolimus alone or bevacizumab with chemotherapy were excluded. In clinical practice, bevacizumab combined with chemotherapy is sometimes used before TPCV, making its omission worth reconsidering.
- **PICO 3** is particularly relevant, as it evaluates carboplatin and vincristine, which are commonly used as first-line therapy.
- **PICO 4** addresses vinblastine often used when patients do not tolerate carboplatin + vincristine or experience relapse.
- **PICO 5 and 6** are especially important for children with the BRAFV600E mutation. While the combination of Dabrafenib and Trametinib has shown efficacy in this subgroup, rapid disease rebound after discontinuation is a known issue. If Tovorafenib can offer more durable responses with fewer long-term side effects - especially given the uncertainties surrounding MEK inhibitor use in children - it could represent a significant advancement.
- **PICO 7**, in my view, is the most compelling, as it focuses on HrQoL outcomes when comparing the two available targeted therapies. This is particularly crucial for very young patients who, if they survive, must also be able to live well.

On a personal note, my daughter was diagnosed with LLG at just 9 months old and has undergone multiple lines of treatment, including carboplatin + vincristine, vinblastine, bevacizumab + irinotecan, TPCV, and ultimately Trametinib. Although it took time to find the right dose, Trametinib has been the most effective so far. Her quality of life has improved significantly—no more infections from chemotherapy, no need for a port, and the convenience of an oral tablet. However, the initial side

effects were challenging, including ingrown nails, mouth ulcers, and pneumonia. One concern raised by our oncologist regarding Tovorafenib is its potential to inhibit growth in children.

Another critical consideration is the optimal duration and discontinuation strategy for Tovorafenib and MEK inhibitors. While current guidance suggests a two-year treatment period, a gradual tapering approach should be explored. Slowly reducing the dose and maintaining low levels before complete cessation may help mitigate rebound effects and improve long-term outcomes.

A.2 Clinical expert input on the consolidated assessment scope proposal

JCA 202406

Consolidated assessment scope proposal

Reg PICO 1 and PICO 2.

Patients with a BRAF fusion could be treated with single agent trametinib

Patients with BRAF V600E mutation could be treated with single agent Dabrafenib.

Reg PICO 5

Patients with a BRAF V600E mutation could be treated with single agent Dabrafnib

Reg Outcome

Register growth (height and weight) as we see reduced height growth during treatment with Tovorafenib

Registration of tumor bleed during treatment as this seems to be a possible adverse advent to Tovorafenib

A.3 Clinical expert input during the preparation of the draft JCA and summary reports

A.3.1 Questions from the assessors to the clinical expert

JCA-MP-2024-06

INPUT REQUIRED FROM CLINICAL EXPERT

further to Article 14(1) of Commission Implementing Regulation (EU) 2024/1381

The assessor and co-assessor are seeking input from the clinical expert on the following questions:

1. Among patients who are starting a second (or later)-line systemic therapy, would you consider the populations of (i) patients with relapsed or refractory pLGG and (ii) patients with pLGG who have received one or more prior systemic therapies, to be different?

2. In your clinical practice, is it necessary for a patient to have relapsed disease or be refractory to their first-line systemic therapy, in order to consider starting second (or later) -line treatment?

3. In your clinical practice, what factors (either patient factors or disease factors) are considered in your decision to:
    a) stop first-line treatment
    b) start second (or later)-line treatment
    c) stop second (or later)-line treatment
    d) retreat with targeted systemic therapy

4. In your clinical practice, are lesions at baseline, disease progression, and response to treatment measured according to the Response Assessment in Paediatric Neuro-oncology (RAPNO) LGG (RAPNO-LGG) criteria? Are previously published neuro-oncology response assessment (RANO-LGG and RANO-HGG) criteria used to measure disease status and/or inform treatment decisions?

    For each of these questions, in your experience, does clinical practice vary across EU member states?

A.3.2 Answers from the clinical expert

JCA-MP-2024-06: Responses clinical expert

Among patients who are starting a second (or later)-line systemic therapy, would you consider the populations of (i) patients with relapsed or refractory pLGG and (ii) patients with pLGG who have received one or more prior systemic therapies, to be different?

In a clinical and research context, these populations are largely overlapping but may not be strictly identical
Relapsed or refractory pLGG refers to tumors that have shown disease progression either while on treatment or after stopping a prior therapy.

Patients who have received one or more prior systemic therapies is a broader group. While most patients in this category would have received subsequent therapy due to progression, some might have switched therapies for other reasons, such as severe toxicity, even if their disease was stable at that point

In clinical trials, inclusion criteria for second-line therapy often specifically require objective evidence of progression (radiological or clinical, typically per criteria like RAPNO-LGG) to ensure a homogenous population of growing tumors for efficacy assessment.

2. In your clinical practice, is it necessary for a patient to have relapsed disease or be refractory to their first-line systemic therapy, in order to consider starting second (or later) -line treatment?

No, it is not strictly necessary for a patient to have documented relapsed or refractory disease to consider starting second (or later)-line treatment in the context of pediatric low-grade gliomas. While disease progression is the most common trigger, clinical practice involves a comprehensive assessment of various factors in a multidisciplinary setting.
Decisions are individualized and balance the risks associated with the tumor and the potential morbidities of the treatment.
Key factors that may lead to starting a new line of therapy even without formal progression include:

- Unacceptable Toxicity**:** If the patient experiences severe or unmanageable side effects from the first-line treatment that significantly impact their quality of life, a switch to an alternative agent may be warranted.
- Clinical Deterioration**:** Worsening symptoms, such as vision loss or neurological deficits attributable to the tumor's location, may necessitate a change in therapy to prevent

permanent functional damage, even if formal radiographic progression criteria are not yet met.

- Molecular Profile: The identification of a specific "actionable" molecular alteration (e.g., a *BRAF* fusion or mutation) might prompt the use of a more effective targeted therapy earlier in the disease course, as these can offer better outcomes for specific tumor types.
- Tumor Location and Age**:** For tumors in critical locations or in very young children, where long-term management goals include minimizing morbidity, clinicians may opt for alternative therapies or earlier intervention with less toxic options.

The management of pLGG is often viewed as treating a chronic condition, focusing on disease control and quality of life.

3. In your clinical practice, what factors (either patient factors or disease factors) are considered in your decision to:

a) stop first-line treatment : Unacceptable or unmanageable toxicity that affects quality of life or disease progression (radiologically or clinically)

b) start second (or later)-line treatment : Clinical deterioration (neurological symptoms, vision loss), age, general condition or disease progression, tumor location (risk to vital structures), molecular profile.

c) stop second (or later)-line treatment Unacceptable toxicity, completed planned treatment duration with good response, patient preference or disease progression during treatment

d) retreat with targeted systemic therapy Good tolerance to previous targeted therapy, patient preference, unacceptable toxicity to ongoing treatment. Previous positive response to the same or similar treatment, presence of an actionable molecular alteration, documented relapse after treatment break.

4. In your clinical practice, are lesions at baseline, disease progression, and response to treatment measured according to the Response Assessment in Paediatric Neuro-oncology (RAPNO) LGG (RAPNO-LGG) criteria? Are previously published neuro-oncology response assessment (RANO-LGG and RANO-HGG) criteria used to measure disease status and/or inform treatment decisions?
Yes to both questions

For each of these questions, in your experience, does clinical practice vary across EU member states?

- Implementation of guidelines**:** Although organizations like SIOP Europe and PaedCan develop European Standard Clinical Practice Recommendations for pediatric tumors (including pLGG), implementation is often voluntary and depends on national and local systems.
- Access to treatment**:** There is great variability in access to new drugs, including targeted therapies, across Europe. Reimbursement rates and timelines vary widely, leading to up to

a 10-fold difference in clinical use between countries (especially between Western and Eastern Europe).

- Resources and infrastructure**:** Differences in healthcare expenditure, equipment (e.g., number of radiotherapy machines), and specialized personnel contribute to inequalities in the care provided.
- Although the RAPNO criteria are the established standard for clinical trials, actual clinical practice can vary across EU member states. Some institutions or countries that have fewer resources or are less involved in the latest international research networks may still use older or less standardized methods. Consensus reports from organizations such as SIOP Europe (International Society of Paediatric Oncology European branch) aim to standardize these procedures.

A.4 Carer input on the revised draft JCA and summary reports

JCA-MP-2024-06 / review of revised JCA reports by individual experts:
Comments should be submitted no later than 24 February 2026 EOB

Please provide your comments on the draft revised Joint Clinical Assessment (JCA) report and summary report.

As part of the initial phase of the JCA, you previously contributed to defining the assessment scope. This scope outlined the key research questions relevant to Member States, structured according to the standardised PICO framework (Population, Intervention, Comparator(s), Outcomes).

The final assessment scope was subsequently shared with the health technology developer, who prepared and submitted a JCA dossier addressing the specified information needs. The evidence in the dossier was organised in line with the defined PICOs.

Following submission, an assessor and co-assessor, appointed by Member State Representatives within the JCA subgroup, reviewed the dossier and prepared a draft JCA report and summary report. After receiving input from the JCA subgroup, the assessor and co-assessor revised these documents. The revised draft JCA report and summary report are now shared with you for review.

You are now kindly requested, in your capacity as an individual expert, to review the revised draft reports and provide your comments.

Please download this form for providing your comments. When completed, upload it to your subfolder. Your comments may be noted here for both the JCA report and the summary report of the JCA. Please be mindful that your input will be annexed to the publicly available JCA report.

1. Please put each new comment in a new row and insert additional rows if needed.
2. Please insert the page number, line number and section title on which your comment applies. If your comment relates to the document as a whole, please put ‘general’ in the “report section title” column.
3. Please provide a description of your comment as specific as possible.

JCA-MP-20XX / review of revised JCA report by individual experts:
Comments should be submitted no later than XX Month 20XX EOB

Comment from **Please indicate if you are 'a clinical expert' or 'a patient / carer'**	**For clinical expert only: Insert your name**	**Which document is the comment related to?** - **JCA report** - **the summary report of the JCA**	**Page number** ***Insert 'general' if your comment relates to the whole document.***	**Line number**	**Report section title**	**Comment** ***Insert each new comment in a new row.***
Patient carer		JCA report	General		General information on the JCA (page 12–14) and Appendix A – Input from experts and stakeholders	While carer input is documented in Appendix A.1, the report does not explain how this input was considered or whether it influenced the assessment scope or conclusions (e.g. PICO definitions, Comparator selection, Outcome prioritisation).
Patient carer		JCA report	Tables 10–12 Pages 19–24		Section 3 – Assessment scope	The exclusion of partial or mixed real-world comparator data limits the relevance of the assessment for families, as paediatric LGG treatment is often individualised in clinical practice.
Patient Carer		JCA report	Pages 74–76		Section 4.3.2 – Results for patient	No HRQoL results are available for Population 2, in FIREFLY-1 so none are assessed. The absence of quality-of-life data represents a major evidence

Of note, expert input is provided on the draft version of the report and summary. Page and line numbers might differ in the final version.

JCA-MP-20XX / review of revised JCA report by individual experts:
Comments should be submitted no later than XX Month 20XX EOB

Comment from **Please indicate if you are 'a clinical expert' or 'a patient / carer'**	**For clinical expert only: Insert your name**	**Which document is the comment related to?** - **JCA report** - **the summary report of the JCA**	**Page number** ***Insert 'general' if your comment relates to the whole document.***	**Line number**	**Report section title**	**Comment** ***Insert each new comment in a new row.***
					population 2 Table 25: Matrix of outcomes	gap for patients and carers, particularly given the chronic nature of paediatric LGG and the importance of functional and cognitive outcomes.
Patient carer		JCA report	Table 21 Pages 49–51		Section 4.3.2.1 – Patient characteristics	The report does not meaningfully discuss growth and developmental impacts as long-term patient-relevant safety outcomes.
Patient carer		JCA report	General			Significant variability in access across countries has major implications for families and should be addressed more explicitly from an equity perspective.
Patient carer		JCA report	General			No discussion of: - Discontinuation strategies - Tapering (how treatment is discontinued, whether gradual reduction is considered, or whether stopping strategies

Of note, expert input is provided on the draft version of the report and summary. Page and line numbers might differ in the final version.

JCA-MP-20XX / review of revised JCA report by individual experts:
Comments should be submitted no later than XX Month 20XX EOB

Comment from **Please indicate if you are 'a clinical expert' or 'a patient / carer'**	**For clinical expert only: Insert your name**	**Which document is the comment related to?** - **JCA report** - **the summary report of the JCA**	**Page number** ***Insert 'general' if your comment relates to the whole document.***	**Line number**	**Report section title**	**Comment** ***Insert each new comment in a new row.***
						influence outcomes) - Rebound effects
Patient Carer		The summary report of the JCA	General			While the Summary Report provides a clear overview of the assessment, it also highlights major uncertainties and missing evidence for patient-relevant outcomes, including quality of life, long-term treatment considerations, and real-world comparators which are particularly important in a paediatric setting.

Of note, expert input is provided on the draft version of the report and summary. Page and line numbers might differ in the final version.

A.5 Clinical expert input on the revised draft JCA and summary reports

As EMA is recommended granting a conditional marketing authorisation in the European Union (EU) for tovorafenib to treat patients aged 6 months and older with paediatric low-grade glioma when the tumour has certain changes in the BRAF gene in patients whose disease has worsened despite previous treatment with one or more systemic medicines, I think posibly Tovorafenib has to be prioritised higher as a treatment option in our recommendations? Paediatric low-grade glioma is the most common brain tumour in children. Although survival is generally high, many patients experience serious longterm complications affecting vision, movement, learning and overall development, with a significant impact on their quality of life. While chemotherapy can be effective for some patients, its benefits are often modest and it may cause substantial side effects.

Appendix B Assessment of information retrieval

The appropriateness of data sources and search strategies by the HTD with regard to the medicinal product under assessment was reviewed by the team of the assessor and co-assessor. Search strategies were checked for appropriateness, and the results of information retrieval included in the HTD's submission dossier were checked for completeness against a search of study registries and in PubMed (for details see Appendix B.1).

Concerns regarding the transparency and verifiability of the information retrieval include:

- For the study registry searches the HTD does not provide lists of excluded studies with reasons for exclusion.
- The search dates for studies performed or sponsored by the HTD and for the studies from submission files to EMA are missing.

Additional concerns were identified regarding the confounder identification the HTD describes in Appendix B.2 of the dossier (for details see Section 4.3.2.2.2.1). Concerns regarding information retrieval and study selection are described in more detail below.

Reasons for exclusion not transparent for all studies

Most importantly, for the searches in study registries, the HTD did not provide lists of excluded studies with the reasons for exclusion. The studies not considered in the dossier should be identified and listed for each of the information retrieval steps, and the reason for exclusion should be specified for each of them. This also includes information retrieval steps in study registries. As this information is missing from the dossier, the reason for exclusion of certain studies from this assessment is not transparent.

For example, the HTD gives no reason for exclusion for study DAY101-102 [41] which is listed in the overview of the tovorafenib clinical programme. Besides patients with melanoma, this study also planned to include patients ≥ 12 years of age with other solid tumours with alterations of the RAS/RAF/MEK/ERK pathway. This may include patients with paediatric LGG and BRAF alterations. As the HTD does not provide reasons for exclusion for the searches in study registries, it is unclear from the dossier why this study was excluded. However, because the inclusion criteria were not limited to LGG or brain tumours, it is unlikely that the study included a sufficient number of patients with paediatric LGG and BRAF alterations relevant for the JCA. Thus, the missing information on the reason for exclusion does not affect the current assessment.

Reasons for exclusion not suitable for all studies

The reasons for exclusion for studies C28001 [42] and C28002 [43] from the tovorafenib clinical programme as well as the reason for exclusion of the investigator-sponsored trial PNOC014 [44] are inappropriate.

Studies C28001 and C28002 were excluded by the HTD because adults with solid malignancies and solid tumours were included, respectively. This reason alone is not sufficient for exclusion because young adults with paediatric LGG (a form of solid tumours) are within the scope of the assessment. However, in study C28001 only one patient with a brain tumour who may potentially be relevant for the current assessment was included according to the information from the CSR. For study C28002 another reason for exclusion applies because tovorafenib is used in combination with other drugs and not as monotherapy. Thus, although adequate reasons for the exclusion of these studies were not provided, this does not affect the evidence base of the current assessment.

Study PNOC014 is a phase 1 dose-escalation, investigator-sponsored trial that included patients with recurrent/progressive mitogen-activated protein kinase (MAPK) pathway-altered paediatric tumours, including paediatric LGG. According to the HTD, neither Ipsen nor their partner Day One funded the study nor do they oversee or manage it.

In Section 5.1.3 of the dossier, the HTD mentions 2 reasons for exclusion of the study: (i) phase 1 dose-escalation study with no data submitted as part of the EMA submission, and (ii) investigator-sponsored study (limited data/information available). If relevant data matching the assessment scope were collected, the first reason alone may not justify exclusion of the study. In Appendix B.2 of the dossier, the HTD describes that Day One attempted to obtain data from PNOC014, but that the quantity of data was limited and no CSR was provided. In addition, no publication on the methods and results of the study is available for the current assessment. According to ClinicalTrials.gov, the study is ongoing with an estimated completion date of 31 December 2025. Therefore, the publication of the results is not expected within the time frame of the JCA [44].

In addition, the HTD generally excluded studies with fewer than 10 participants in the population of interest from data extraction. In general, using such a cut-off point as an exclusion criterion is not considered useful, and the decision on the inclusion of studies of a certain size is a matter of assessment.

B.1 Search strategies to verify completeness of studies

B.1.1 Study registries

1) ClinicalTrials.gov (last search on 31 July 2025)

Provider: U.S. National Institutes of Health

- URL: http://www.clinicaltrials.gov

Interface: Expert Search

Search strategy for tovorafenib and comparators
(tovorafenib OR MLN-2480 OR TAK-580) OR ((vinblastine OR vincristine OR everolimus OR SDZ-RAD OR RAD-001 OR bevacizumab OR trametinib OR GSK-1120212 OR JTP-74057) AND AREA[ConditionSearch](glioma OR astrocytoma OR glioblastoma OR oligodendroglioma OR ependymoma) AND AREA[StdAge](CHILD))

2) EU Clinical Trials Register (last search on 31 July 2025)

Provider: European Medicines Agency

- URL: https://www.clinicaltrialsregister.eu/ctr-search/search

Interface: Basic Search

Search strategy for tovorafenib
tovorafenib* OR MLN-2480 OR (MLN 2480) OR MLN2480 OR TAK-580 OR (TAK 580) OR TAK580

Search strategy for comparators
(vinblastine* OR vincristine* OR everolimus* OR SDZ-RAD OR (SDZ RAD) OR SDZRAD OR RAD-001 OR (RAD 001) OR RAD001 OR bevacizumab* OR trametinib* OR GSK-1120212 OR (GSK 1120212) OR GSK1120212 OR JTP-74057 OR (JTP 74057) OR JTP74057) AND (glioma* OR astrocytoma* OR glioblastoma* OR oligodendroglioma* OR ependymoma*) // Under 18

3) Clinical Trials Information System (CTIS) (last search on 31/07/2025)

Provider: European Medicines Agency

- URL: https://euclinicaltrials.eu/search-for-clinical-trials

Interface: Basic Criteria

Search strategy for tovorafenib
tovorafenib, MLN-2480, MLN2480, TAK-580, TAK580 [Contain any of these terms]

Search strategy for comparators
glioma [Contain all of these terms] / vinblastine, vincristine, everolimus, SDZ-RAD, SDZRAD, RAD-001, RAD001, bevacizumab, trametinib, GSK-1120212, GSK1120212, JTP-74057, JTP74057 [Contain any of these terms]
gliomas [Contain all of these terms] / vinblastine, vincristine, everolimus, SDZ-RAD, SDZRAD, RAD-001, RAD001, bevacizumab, trametinib, GSK-1120212, GSK1120212, JTP-74057, JTP74057 [Contain any of these terms]
glioblastoma [Contain all of these terms] / vinblastine, vincristine, everolimus, SDZ-RAD, SDZRAD, RAD-001, RAD001, bevacizumab, trametinib, GSK-1120212, GSK1120212, JTP-74057, JTP74057 [Contain any of these terms]
ependymoma [Contain all of these terms] / vinblastine, vincristine, everolimus, SDZ-RAD, SDZRAD, RAD-001, RAD001, bevacizumab, trametinib, GSK-1120212, GSK1120212, JTP-74057, JTP74057 [Contain any of these terms]

B.1.2 Bibliographic databases

MEDLINE (PubMed) (last search on 31 July 2025)

The following filter was applied:

Population search filter: Improved CCG child filter for PubMed [45]

#	Query
1	"tovorafenib"[Supplementary Concept] OR "tovorafenib"[All Fields]
2	"vinblastine"[Supplementary Concept] OR "vinblastine"[All Fields] OR "vinblastin"[All Fields] OR "vinblastine"[MeSH Terms] OR "vincristine"[Supplementary Concept] OR "vincristine"[All Fields] OR "vincristin"[All Fields] OR "vincristine"[MeSH Terms] OR "everolimus"[Supplementary Concept] OR "everolimus"[All Fields] OR "everolimus"[MeSH Terms] OR "bevacizumab"[Supplementary Concept] OR "bevacizumab"[All Fields] OR "bevacizumab"[MeSH Terms] OR "bevacizumab s"[All Fields] OR "trametinib"[Supplementary Concept] OR "trametinib"[All Fields]
3	"glioma"[MeSH Terms] OR "glioma"[All Fields] OR "gliomas"[All Fields] OR "glioma s"[All Fields] OR "astrocytoma"[MeSH Terms] OR "astrocytoma"[All Fields] OR "astrocytomas"[All Fields] OR "glioblastoma"[MeSH Terms] OR "glioblastoma"[All Fields] OR "glioblastomas"[All Fields] OR "oligodendroglioma"[MeSH Terms] OR "oligodendroglioma"[All Fields] OR "oligodendrogliomas"[All Fields] OR "ependymoma"[MeSH Terms] OR "ependymoma"[All Fields] OR "ependymomas"[All Fields]

#	Query
4	"infan*"[All Fields] OR "newborn*"[All Fields] OR "new-born"[All Fields] OR "perinat*"[All Fields] OR "neonat*"[All Fields] OR ("infant, newborn"[MeSH Terms] OR ("infant"[All Fields] AND "newborn"[All Fields]) OR "newborn infant"[All Fields] OR "baby"[All Fields] OR "infant"[MeSH Terms] OR "infant"[All Fields]) OR "baby*"[All Fields] OR ("baby s"[All Fields] OR "babys"[All Fields] OR "infant"[MeSH Terms] OR "infant"[All Fields] OR "babies"[All Fields]) OR "toddler*"[All Fields] OR ("minority groups"[MeSH Terms] OR ("minority"[All Fields] AND "groups"[All Fields]) OR "minority groups"[All Fields] OR "minorities"[All Fields] OR "minority"[All Fields] OR "minority s"[All Fields] OR "minors"[MeSH Terms] OR "minors"[All Fields] OR "minor"[All Fields]) OR "minors*"[All Fields] OR ("men"[MeSH Terms] OR "men"[All Fields] OR "boy"[All Fields]) OR ("men"[MeSH Terms] OR "men"[All Fields] OR "boys"[All Fields]) OR ("boyfriend"[All Fields] OR "boyfriend s"[All Fields] OR "boyfriends"[All Fields]) OR "boyhood"[All Fields] OR "girl*"[All Fields] OR "kid"[All Fields] OR "kids"[All Fields] OR ("child"[MeSH Terms] OR "child"[All Fields] OR "children"[All Fields] OR "child s"[All Fields] OR "children s"[All Fields] OR "childrens"[All Fields] OR "childs"[All Fields]) OR "child*"[All Fields] OR "children*"[All Fields] OR "schoolchild*"[All Fields] OR "schoolchild"[All Fields] OR "school child"[Title/Abstract] OR "school child*"[Title/Abstract] OR "adolescen*"[All Fields] OR "juvenil*"[All Fields] OR "youth*"[All Fields] OR "teen*"[All Fields] OR "under*age*"[All Fields] OR "pubescen*"[All Fields] OR "pediatrics"[MeSH Terms] OR "pediatric*"[All Fields] OR "paediatric*"[All Fields] OR "peadiatric*"[All Fields] OR "school"[Title/Abstract] OR "school*"[Title/Abstract] OR "prematur*"[All Fields] OR "preterm*"[All Fields]
5	"braf"[All Fields]
6	#2 AND #3 AND #4 AND #5
7	#1 OR #6
8	#7 AND Filters: English, German

Appendix C Additional study information and data, including uncertainties in the results

C.1 Additional characteristics of included studies for Population 2 (BRAF V600E mutation in patients > 1 year), PICO 5

Table 29: Information on the course of included studies – actual treatment duration and observation periods (Population 2)

Study reference/ID[a] **Outcome category**	**FIREFLY-1[b]**		**Bouffet 2023**
Intervention	**Tovorafenib** **N = 13 (Arm 1)**	**Tovorafenib** **N = 22 (Arm 1 + 2)**	**Dabrafenib + trametinib** **N = 36**
Treatment duration [days/months]			
Median [min; max]	722.0 [142.0; 958.0] days	612.0 [142.0; 958.0] days	24 (range: 2.1 -52.5) months
Mean (SD)	563.3 (278.1) days	545.7 (237.2) days	NR
Observation period [months]			
Median [min; max]	30.2 [6.5; 36.5]	20.4 [7.9; 23.7]	NR
Mean (SD)	27.42 (7.256)	24.08 (7.399)	NR
OS observation period [months][c]			
Median [min; max]	27.8[d] [10.0, 31.5]	NR	NR
Mean (SD)	25.86 (6.306)	NR	NR
PFS observation period[months]			
Median [min; max]	NR	NR	NR
Mean (SD)	NR	NR	NR
ORR observation period[months]			
Median [min; max]	NR	NR	NR
Mean (SD)	NR	NR	NR
Safety observation period [months]			
Median [min; max]	30.2 [6.5; 36.5]	24.9 [6.5; 36.5]	NR
Mean (SD)	27.42 (7.256)	24.08 (7.399)	NR

a: Only outcomes with the requested information available for at least one included study are presented in the dossier.
b: Efficacy analyses were conducted for Arm 1 and safety and HRQoL analyses were conducted for Arm 1 + Arm 2.
c: There is no information available on how the observation period was calculated.
d: Median OS only reported for the Full Analysis Set, N = 10.

ID: identification; max: maximum; min: minimum; NR: not reported; ORR: overall response rate; OS: overall survival; PFS: progression-free survival; SD: standard deviation

C.2 Additional study results for Population 2 (BRAF V600E mutation in patients > 1 year), PICO 5

C.2.1 Weight distribution plots

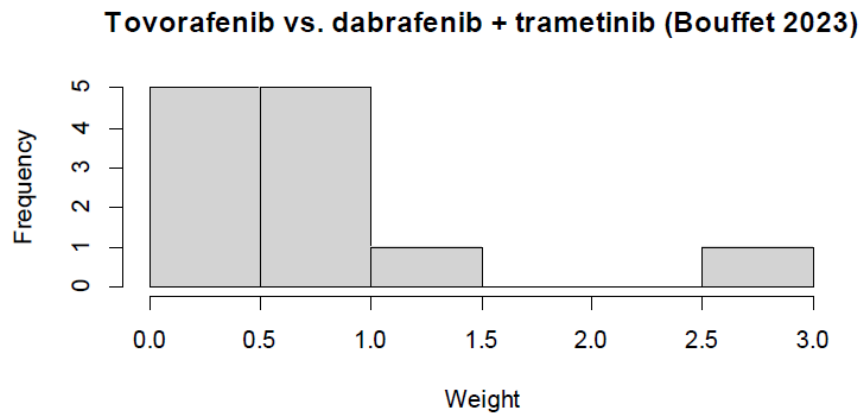

Figure 1: Weight distribution in FIREFLY-1 vs. Bouffet 2023, Base Case (ESS 5.82)

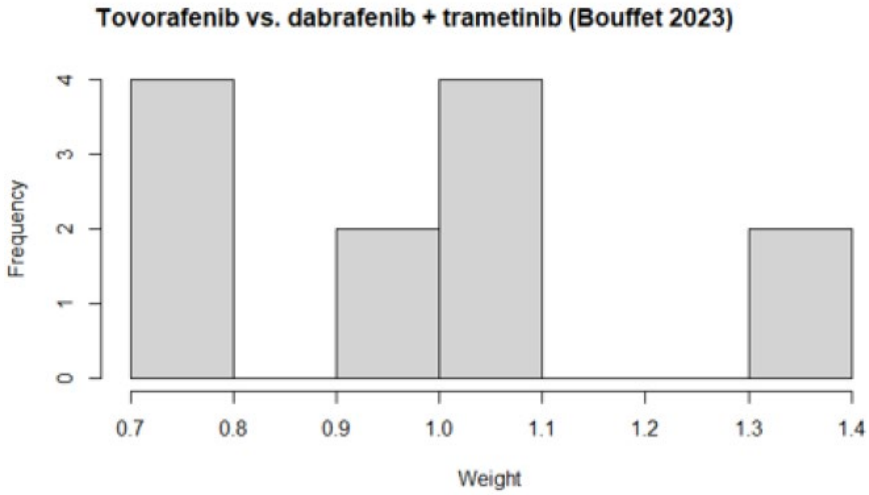

Figure 2: Weight distribution in FIREFLY-1 vs. Bouffet 2023, Scenario 1 (ESS 11.54)

C.2.2 Kaplan-Meier plots

The assessors have identified a number of major uncertainties with the analysis of PFS which are described in Section 4.3 and note in particular that the difference between the Kaplan-Meier plots (see Figure 3) for tovorafenib and dabrafenib + trametinib should not necessarily be interpreted as evidence of causal effects of treatment.

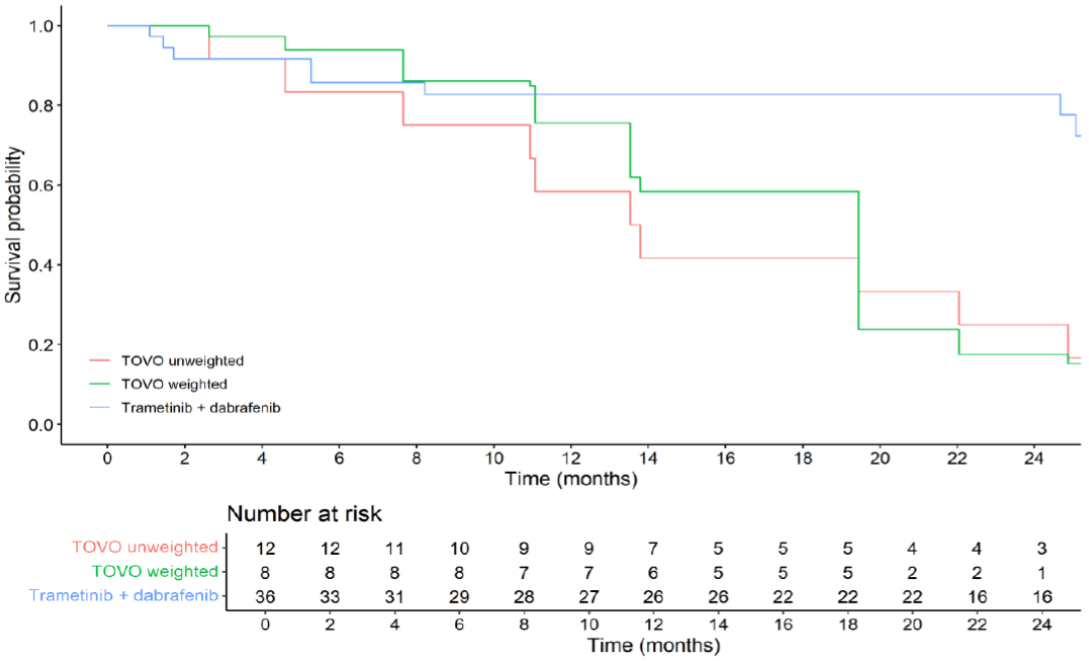

Figure 3: Kaplan-Meier plot of PFS – reconstructed IPD for dabrafenib + trametinib and MAIC-weighted/unweighted FIREFLY-1 data – RANO-LGG by IRC; for Population 2 (BRAF V600E mutation in patients > 1 year), PICO 5.

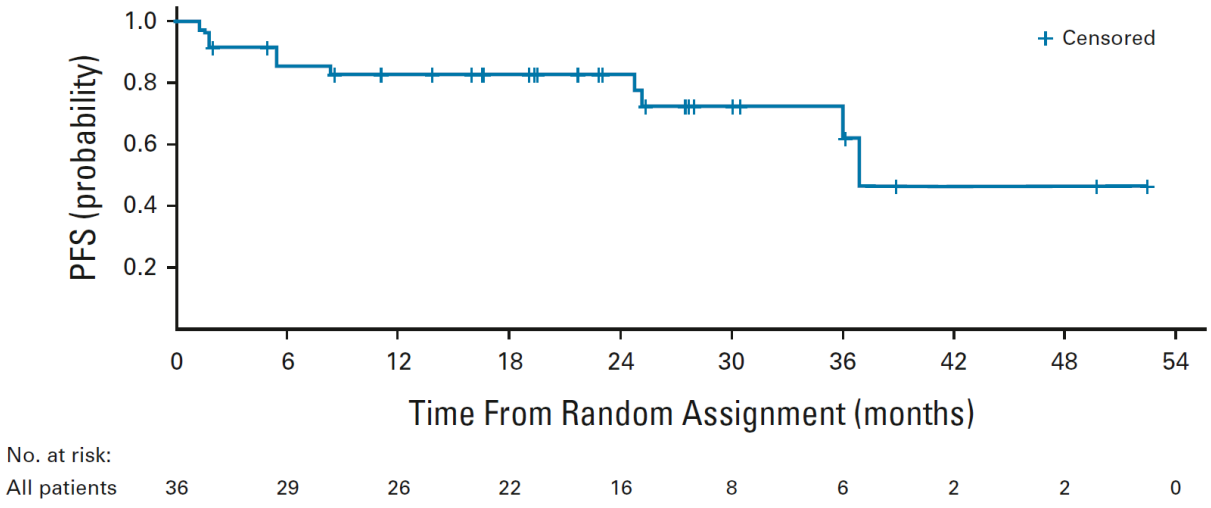

Figure 4: Kaplan-Meier plot of PFS – modified RANO-LGG by IRC (Bouffet 2023) for Population 2 (BRAF V600E mutation in patients > 1 year), PICO 5

C.2.3 Descriptive outcomes

Table 30: Matrix of descriptive outcomes in the included studies for Population 2 (BRAF V600E mutation in patients > 1 year) – PICO 5 – tovorafenib vs. dabrafenib + trametinib

Outcome **Operationalisation**	**Outcome reported in the study (Yes/No)**	
	FIREFLY-1 **tovorafenib**	**Bouffet 2023** **dabrafenib + trametinib**
Descriptive outcomes		
BOR		
RANO-HGG (INV)	yes	yes
RANO-HGG (IRC)	yes	no[a]
RANO-LGG (IRC)	yes	no[b]
modified RANO-LGG (IRC)	no[c]	yes
RAPNO (IRC)	yes	no
DOR		
RANO-HGG (INV)	yes	yes
RANO-HGG (IRC)	yes	no
RANO-LGG (IRC)	yes	no
modified RANO-LGG (IRC)	no	yes
RAPNO (IRC)	yes	no
TTR		
RANO-HGG (INV)	yes	yes
RANO-HGG (IRC)	yes	no
RANO-LGG (IRC)	yes	no
modified RANO-LGG (IRC)	no	yes
RAPNO (IRC)	yes	no

a: data is available in a submitted dossier from Novartis to the Federal Joint Committee [46]

b: This outcome is identified in the relevant table of descriptive results in the HTD dossier as being reported in the Bouffet 2023 study. However, the Bouffet 2023 study used modified RANO-LGG (IRC) criteria that excluded the MR category, classifying such patients as SD.

c: BOR based on modified RANO-LGG (IRC) criteria (including CR and PR, excluding MR) was not explicitly reported in the dossier, but the individual components of this outcome are available from FIREFLY-1 and were reported in the dossier.

BOR: best overall response; DOR: duration of response; HGG: high-grade glioma; INV: investigator assessed; IRC: independent radiology review committee; LGG: low-grade glioma; RANO: Response Assessment in Neuro-oncology; RAPNO: Response Assessment in Paediatric Neuro-oncology; TTR: time to response

Table 31: Descriptive outcomes (BOR) for Population 2 (BRAF V600E mutation in patients > 1 year) – PICO 5 – tovorafenib vs. dabrafenib + trametinib

Study reference/ID **Timepoint** **Outcome** **Tumour response assessment criteria and method**	**Tovorafenib** **FIREFLY-1**			**Dabrafenib + trametinib** **Bouffet 2023**		
	Response category	**N**	**Patients with event n (%)**	**Response category**	**N**	**Patients with event n (%)**
Data-cut	2-year			Study completion		
BOR						
RANO-HGG (INV)	CR	10	0	CR	36	3 (8.3)
	PR	10	4 (40.0)	PR	36	16 (44.4)
	SD	10	5 (50.0)	SD	36	15 (41.7)
	PD	10	1 (10.0)	PD	36	1 (2.8)
				unknown	36	1 (2.8)
RANO-HGG (IRC)	CR	10	2 (20.0)	CR	NR	NR
	PR	10	3 (30.0)	PR	NR	NR
	SD	10	4 (40.0)	SD	NR	NR
	PD	10	1 (10.0)	PD	NR	NR
				unknown	NR	NR
RANO-LGG (IRC)	CR	12	0	–	–	–
	PR	12	5 (41.7)	–	–	–
	MR	12	2 (16.7)	–	–	–
	SD	12	3 (25.0)	–	–	–
	PD	12	2 (16.7)	–	–	–

Study reference/ID **Timepoint** **Outcome** **Tumour response assessment criteria and method**	**Tovorafenib** **FIREFLY-1**			**Dabrafenib + trametinib** **Bouffet 2023**		
	Response category	**N**	**Patients with event n (%)**	**Response category**	**N**	**Patients with event n (%)**
modified RANO-LGG (IRC)	NR	NR	NR	CR	36	0
	NR	NR	NR	PR	36	9 (25.0)
	NR	NR	NR	SD	36	23 (63.9)
	NR	NR	NR	PD	36	3 (8.3)
	NR	NR	NR	unknown	36	1 (2.8)
RAPNO (IRC)	CR	12	0	–	–	–
	PR	12	5 (41.7)	–	–	–
	MR	12	1 (8.3)	–	–	–
	SD	12	3 (25.0)	–	–	–
	PD	12	3 (25.0)	–	–	–

BOR: best overall response; CR: complete response; HGG: high-grade glioma; INV: investigator; IRC: independent review committee; LGG: low-grade glioma; MR: minor response; N: number of patients in the analysis; NR: not reported; PD: progressed disease; PR: partial response; RANO: Response Assessment in Neuro-Oncology criteria; RAPNO: Response Assessment in Paediatric Neuro-Oncology; SD: stable disease

Table 32: Descriptive outcomes (DOR, TTR) for Population 2 (BRAF V600E mutation in patients > 1 year) – PICO 5 – tovorafenib vs. dabrafenib + trametinib

Study reference/ID **Time point** **Outcome** **Tumour response assessment criteria and method**	**Tovorafenib** **FIREFLY-1**		**Dabrafenib + trametinib** **Bouffet 2023**	
	N/N^{Cen}	**Median time to event in months [95%-CI]** **patients with event n (%)**	**N/N^{Cen}**	**Median time to event in months [95%-CI]** **patients with event n (%)**
Data-cut		2-year		Study completion
DOR				
RANO-LGG (IRC)	7/1	18.43 (8.38, NE) 6 (85.7)	–	–
modified RANO-LGG (IRC)	–	–	9/7	33.6 (11.2, NE) 2 (22.2)
RANO-HGG (INV)	4/3	NE (22.18, NE) 1 (25.0)	19/NR	NE (20.0, NE) NR
RANO-HGG (IRC)	5/1	15.11 (7.66, NE) 4 (80.0)	NR	NR
RAPNO (IRC)	6/0	20.70 (8.38, NE) 6 (100)	–	–

Study reference/ID Time point Outcome Tumour response assessment criteria and method	Tovorafenib FIREFLY-1 N/N^{Cen}	Tovorafenib FIREFLY-1 Median time to event in months [95%-CI] patients with event n (%)	Dabrafenib + trametinib Bouffet 2023 N/N^{Cen}	Dabrafenib + trametinib Bouffet 2023 Median time to event in months [95%-CI] patients with event n (%)
TTR				
RANO-LGG (IRC)	12/NR	2.86 (range: 1.61, 11.30) 7 (50.0)	–	–
modified RANO-LGG (IRC)	–	–	NR	NR
RANO-HGG (INV)	10/NR	5.63 (range: 2.73, 16.36) NR	NR	NR
RANO-HGG (IRC)	10/NR	2.73 (range: 2.60, 16.36) 5 (50.0)	NR	NR
RAPNO (IRC)	12/NR	2.76 (range: 1.61, 3.02) 6 (50.0)	NR	NR

CI: confidence interval; CR: complete response; DOR: duration of response; HGG: high-grade glioma; HR: hazard ratio; INV: investigator; IRC: independent review committee; LGG: low-grade glioma; MR: minor response; N: number of patients in the analysis; N^{Cen}: number of censored patients; NE: not evaluable; NR: not reported; PD: progressed disease; PR: partial response; RANO: Response Assessment in Neuro-Oncology criteria; RAPNO: Response Assessment in Paediatric Neuro-Oncology; SD: stable disease; TTR: time to response

Table 33: Descriptive outcomes (DOR landmark rates) for Population 2 (BRAF V600E mutation in patients > 1 year) – PICO 5 – tovorafenib vs. dabrafenib + trametinib

Study reference/ID Outcome Tumour response assessment criteria and method	Tovorafenib FIREFLY-1 N/ N^{Cen}	Tovorafenib FIREFLY-1 Probability of continued response (95%-CI)	Dabrafenib + trametinib Bouffet 2023 N/ N^{Cen}	Dabrafenib + trametinib Bouffet 2023 Probability of continued response (95%-CI)
Time point		6 months		Study completion
DOR				
RANO-LGG (IRC)	NR/NR	100.00 (100.00, 100.00)	–	–
modified RANO-LGG (IRC)	–	–	NR	NR
RANO-HGG (INV)	NR/NR	100.0 (100.0, 100.0)	NR	NR
RANO-HGG (IRC)	NR/NR	100.00 (100.00, 100.00)	NR	NR
RAPNO (IRC)	NR/NR	100.00 (100.00, 100.00)	NR	NR
Time point		12 months		Study completion

Study reference/ID Outcome	Tovorafenib FIREFLY-1		Dabrafenib + trametinib Bouffet 2023	
Tumour response assessment criteria and method	**N/ N^{Cen}**	**Probability of continued response (95%-CI)**	**N/ N^{Cen}**	**Probability of continued response (95%-CI)**
DOR				
RANO-LGG (IRC)	NR/NR	57.14 (17.19, 83.71)	–	–
modified RANO-LGG (IRC)	–	–	NR	NR
RANO-HGG (INV)	NR/NR	100.0 (100.0, 100.0)	NR	NR
RANO-HGG (IRC)	NR/NR	60.0 (12.57, 88.18)	NR	NR
RAPNO (IRC)	NR/NR	66.67 (19.46, 90.44)	NR	NR
Time point		24 months		Study completion
DOR				
RANO-LGG (IRC)	7/1	0.00 (NE, NE)	–	–
modified RANO-LGG (IRC)	–	–	9/7	80.00 (30.00, 100.00)
RANO-HGG (INV)	4/3	66.67 (5.41, 94.52)	19/NR	80.0 (50.0, 90.0)
RANO-HGG (IRC)	5/1	40.0 (5.20, 75.28)	NR	NR
RAPNO (IRC)	6/0	16.67 (0.77, 51.68)	–	–

CI: confidence interval; CR: complete response; DOR: duration of response; HGG: high-grade glioma; INV: investigator; IRC: independent review committee; LGG: low-grade glioma; N: number of patients in the analysis; N^{Cen}: number of censored patients; NE: not evaluable; NR: not reported; RANO: Response Assessment in Neuro-Oncology criteria; RAPNO: Response Assessment in Paediatric Neuro-Oncology

C.2.4 Calculating relative risks for progression-free survival

Relative effect measures for the outcome of PFS measured as a dichotomous outcome at 6 and 12 months were requested in the assessment scope. Although these outcomes were reported by the HTD for each treatment group in the MAIC analysis (as Kaplan-Meier estimates of the probabilities of 6- and 12-month PFS, with associated confidence intervals), the corresponding relative effects were not reported. For completeness, the assessors have calculated these from the available data, measured as the relative risk (RR) p_1/p_2 where p_1 and p_2 denote the probabilities of PFS in the tovorafenib and dabrafenib + trametinib groups respectively, at the relevant timepoint. The point estimate of the RR was calculated as the ratio of the point estimates $\widehat{p_J}$ of p_J $(J = 1,2)$ reported by the HTD for the corresponding

treatment groups, while the standard error of this estimate was calculated on the logarithmic scale using the relation $Var\left(\log\left(\frac{p_1}{p_2}\right)\right) = Var(\log(p_1)) + Var(\log(p_2))$. The estimates of $Var(\log(p_J))$ were back-calculated from the confidence intervals for the estimates $\widehat{p_J}$ reported by the HTD, under the assumption that the latter have been calculated as $\exp(\widehat{p_J} \pm 1.96 * se(\widehat{p_J}))$ (truncated at 1 where necessary), where $se(\widehat{p_J})$ denotes the standard error. This formula is the default method used to calculate confidence intervals of Kaplan-Meier estimates in the "survfit" function of the "survival" package in R, which has been used in the code provided by the HTD [47]. Confidence intervals for the RR were then calculated as $\exp\left(\log\left(\frac{\widehat{p_1}}{\widehat{p_2}}\right) \pm 1.96 * se(\log(\frac{\widehat{p_1}}{\widehat{p_2}}))\right)$.

C.3 Safety

Table 34: Relative safety results for PICO 5

Intervention	**Tovorafenib**		**Dabrafenib + trametinib**		**Indirect comparison method**	**Tovorafenib vs. dabrafenib + trametinib**	
Study	**FIREFLY-1**		**Bouffet 2023**				
Data-cut	**2 year**		**NR**				
Outcome	**Events (sample size)**	**Rate [95%-CI]**	**Events (sample size)**	**Rate [95%-CI]**		**ESS (FIREFLY-1)**	**OR[a] [95%-CI] p-value[c]**
Safety outcomes							
Any AE	22 (22)	100% [NR]	36 (36)	100% [NR]	NA: Indirect comparison not feasible due to 100% event rate in both treatment groups		
Severe AE [CTCAE Grade ≥ 3]	NR	NR	22 (36)	61.1% [NR]	MAIC: base case[b]	14.26	12.65 [2.30, 69.58] p = 0.004
	20 (22)	90.9% [NR]			Naïve comparison	22	6.36 [1.28, 31.53] p = 0.023
Serious AE	NR	NR	15 (36)	41.7% [NR]	MAIC base case[b]	14.26	3.04 [0.77, 12.10] p = 0.144
	13 (22)	59.1% [NR]			Naïve comparison	22	2.02 [0.69, 5.94] p = 0.200
Treatment interruption due to AE	NR	NR	26 (36)	72.2% [NR]	MAIC: base case[b]	14.26	1.15 [0.28, 4.76] p = 0.849
	15 (22)	68.2% [NR]			Naïve comparison	22	0.82 [0.26, 2.62] p = 0.743
Treatment discontinuation due to AE	NR	NR	8 (36)	22.2% [NR]	MAIC: base case[b]	14.26	0.58 [0.11, 3.03] p = 0.520
	3 (22)	13.6% [NR]			Naïve comparison	22	0.55 [0.13, 2.35] p = 0.422
Death related to AE	2 (22)	9.1% [NR]	0 (36)	0% [NR]	NA: Indirect comparison not provided due to zero event rate in dabrafenib + trametinib group		

Intervention	**Tovorafenib**		**Dabrafenib + trametinib**		**Indirect comparison method**	**Tovorafenib vs. dabrafenib + trametinib**	
Study	**FIREFLY-1**		**Bouffet 2023**				
Data-cut	**2 year**		**NR**				
Outcome	**Events (sample size)**	**Rate [95%-CI]**	**Events (sample size)**	**Rate [95%-CI]**		**ESS (FIREFLY-1)**	**OR[a] [95%-CI] p-value[c]**
a: OR > 1 indicates higher event rates with tovorafenib b: adjusted for age, prior surgery, prior radiotherapy, Karnofsky/Lansky score, gender (male) c: No formal hypothesis testing was described in the dossier nor in the ITC SAP. All p-values are nominal, have not be obtained through prespecified analysis and are not controlled for multiplicity.							
AE: adverse events; CI: confidence interval; CTCAE: Common Terminology Criteria for Adverse Events; ESS: effective sample size; ITC: indirect treatment comparison; OR: odds ratio; NA: not applicable; NR: not reported; NA: not applicable; SAP: statistical analysis plan							

Appendix D Requests for further specifications, clarifications or additional information

D.1 First request for further specifications, clarifications or additional information

D.1.1 Assessors' first request for further specifications, clarifications or additional information to the HTD

JCA-MP-2024-06

INFORMATION, DATA, ANALYSES OR OTHER EVIDENCE TO BE REQUESTED FROM THE HTD

further to Article 11(2) of Regulation (EU) 2021/2282

The deadline for response, set by the assessor and co-assessor, is **7 days**

as per Article 12(5) of Commission Implementing Regulation (EU) 2024/1381

Request	**Rationale**
1. Analysis and results of relative effects on **ORR** based on **RANO-HGG (INV) criteria (investigator assessment),** to facilitate comparison with Bouffet 2023 using consistent RANO-HGG (INV) definitions. *The Bouffet data included in Table 57 of the Dossier V.2, but has been mislabelled as RANO-LGG (INV) when it should actually be RANO-HGG (INV). This is clear in the Bouffet 2023 publication, from the description of the methods (page 666), references (23 and 24) and results (Table 2); and clear from the Novartis Protocol for this study (specifically, Appendix 5 of the Protocol, which clearly outlines the use of RANO-HGG as the criteria used for investigator assessment of response, and Amendment 9, which clearly describes the addition of RANO-LGG only for independent assessment)* These analyses should be based on the following data: FIREFLY-1 Arm 1 (N=10, CR+PR, 0+4), Bouffet 2023 (N=36, CR+PR, 3+16).	There has been a mis-interpretation of the results reported in Bouffet 2023 (as outlined in the request herein).
2. Analysis and results of relative effects on **ORR** based on **RANO-HGG (INV) criteria (investigator assessment),** to facilitate comparison with Bouffet 2023 using consistent RANO-HGG (INV) definitions, (as requested above) using pooled data from Arm 1 and Arm 2 of FIREFLY-1.	Given the similarity of the patients in the full analysis sets of Arm 1 and Arm 2, their relevance to the population defined by PICO 5, and the availability of RANO-HGG (INV) based response assessments, pooling of results from these two arms is requested.
3. Analysis and results of relative effects on **PFS** based on **RANO-HGG (INV)** criteria (investigator assessment), to facilitate comparison with Bouffet 2023 using consistent RANO-HGG (INV) definitions, (as requested above) using pooled data from Arm 1 and Arm 2 of FIREFLY-1.	Extension of the evidence to Arm 2 in the FIREFLY-1 trial would facilitate additional PFS comparison with Bouffet 2023, based on RANO-HGG (INV) criteria

D.1.2 Response of the HTD to the first request for further specifications, clarifications or additional information

Dossier of the Joint Clinical Assessment (JCA) of a Medicinal Product (MP): Request for information, data, analyses or other evidence from the assessor and co-assessor - health technology developer (HTD) explanations

Tovorafenib (Ojemda®)

Ipsen Pharma

Code: JCA-MP-2024-06

Regulation (EU) 2021/2282 on Health Technology Assessment

Version 0.2

Date submitted: 21 November 2025

Pursuant to the template set out in Annex I of **Commission Implementing Regulation (EU) 2024/1381**

Ipsen Pharma (the health technology developer [HTD]) appreciates the opportunity to provide additional information related to the ongoing assessment of tovorafenib (Ojemda®), code JCA-MP-2024-06, as requested by the assessor and co-assessor in the "Request for information, data, analyses or other evidence from the assessor and co-assessor of the joint clinical assessment of the medicinal product Ojemda (Tovorafenib), case No JCA-MP-2024-06" document received on 14 November 2025. Please find below the necessary explanations for the requested information. There have been no updates to the dossier based on this request document; an updated dossier has therefore not been uploaded.

Pursuant to Article 11(2), third sentence, of Regulation (EU) 2021/2282, the HTD also takes this opportunity to inform the Coordination Group that Day One Biopharmaceuticals, Inc., sponsor of the FIREFLY-1 study and US marketing authorization holder of Ojemda™ (tovorafenib), has just recently shared with the HTD clinical data from an updated data cut-off for FIREFLY-1. This is a 3-year follow-up of patients in the study with a data cut-off date of 06 June 2025, and is not a pre-specified cut-off, nor is it part of any formal commitment to the Food and Drug Administration (FDA) for this study. Based on the HTD's assessment of the Tables, Figures and Listings (TFLs) shared by the sponsor, the clinical data is broadly consistent with the data presented in the European Medicines Agency (EMA) submission package, with consistent results for both efficacy and safety; no new safety signals were observed in the 3-year update. As a result, Day One Biopharmaceuticals, Inc., also confirmed to the HTD that there is no plan to update the US Product Information with this new clinical data cut-off. The HTD has also been informed that some of the results from this updated data cut-off are planned to be presented at an upcoming conference of the Society for Neuro-Oncology (SNO) Annual Meeting.

The HTD shared a similar communication around the 3-year data from the FIREFLY-1 study with EMA (via e-mail) on 14 November 2025. For the Coordination Group's awareness, EMA was informed in the same communication that health-related quality of life (HRQoL) data, which are not available in the clinical study reports (CSRs) due to the exploratory nature of these assessments, had been submitted as part of the Joint Clinical Assessment (JCA) dossier.

Table 1: Summary of HTD responses to the assessor's and co-assessors request for information, data, analyses or other evidence

Request	Rationale	HTD response to request
1. Analysis and results of relative effects on **overall response rate (ORR)** based on Response Assessment in Neuro-Oncology for high-grade gliomas (**RANO-HGG) (INV) criteria (investigator assessment)**, to facilitate comparison with Bouffet 2023 using consistent RANO-HGG (INV) definitions. *The Bouffet data included in Table 57 of the Dossier V.2, but has been mislabelled as Response Assessment in Neuro-Oncology criteria for low-grade gliomas (RANO-LGG) (INV) when it should actually be RANO-HGG (INV). This is clear in the Bouffet 2023 publication, from the description of the methods (page 666), references (23 and 24) and results (Table 2); and clear from the Novartis Protocol for this study (specifically, Appendix 5 of the Protocol, which clearly outlines the use of RANO-HGG as the criteria used for investigator assessment of response, and Amendment 9, which clearly describes the addition of RANO-LGG only for independent assessment).* These analyses should be based on the following data: FIREFLY-1 Arm 1 (N=10, CR+PR, 0+4), Bouffet 2023 (N=36, CR+PR, 3+16).	There has been a mis-interpretation of the results reported in Bouffet 2023 (as outlined in the request herein).	The HTD does not agree that there has been a mis-interpretation of the results in Bouffet 2023 (1). The HTD finds that the Novartis protocol for the study is unclear on several points; it is acknowledged that specific text sections in Appendix 5 and Amendment 9 of the protocol could indicate that RANO-LGG criteria were used only for independent assessment of response, in alignment with the comment from the assessor and co-assessor. However, the HTD's conclusion continues to be that in the Bouffet 2023 publication (1), both the independently- and investigator-assessed results presented are based on RANO-LGG criteria, in alignment with the information in Dossier V0.2. Please refer to Section 1 below for a comprehensive rationale.
2. Analysis and results of relative effects on **ORR** based on **RANO-HGG (INV) criteria (investigator assessment)**, to facilitate comparison with Bouffet 2023 using consistent RANO-HGG (INV) definitions, (as requested above) using pooled data from Arm 1 and Arm 2 of FIREFLY-1.	Given the similarity of the patients in the full analysis sets of Arm 1 and Arm 2, their relevance to the population defined by PICO 5, and the availability of RANO-HGG (INV) based response assessments, pooling of results from these two arms is requested.	The HTD previously provided descriptive RANO-HGG (INV) data for Arm 2 as part of the response to the request for providing "relative effectiveness results for PICO 5 and PICO 7 considering the efficacy results available for patients in Arm 2 of clinical trial FIREFLY-1" included in the Commission's Second Request. However, the HTD also clearly explained that the Arm 2 efficacy data are **not** suitable for inclusion on descriptive level nor for conducting comparative effectiveness analyses for PICO 5 and PICO 7 in the Dossier. Please refer to Section 2 below for a comprehensive rationale.

3. Analysis and results of relative effects on **progression-free survival (PFS)** based on **RANO-HGG (INV) criteria (investigator assessment)**, to facilitate comparison with Bouffet 2023 using consistent RANO-HGG (INV) definitions, (as requested above) using pooled data from Arm 1 and Arm 2 of FIREFLY-1.	Extension of the evidence to Arm 2 in the FIREFLY-1 trial would facilitate additional PFS comparison with Bouffet 2023, based on RANO-HGG (INV) criteria.	Please refer to response to request no. 2 above and Section 2 below.

1. Response to request for providing analysis and results of relative effects on ORR based on RANO-HGG (INV) criteria (investigator assessment), to facilitate comparison with Bouffet 2023 using consistent RANO-HGG (INV) definitions

The HTD understands that the rationale for above request is that, according to the assessor and co-assessor, RANO-HGG criteria were used for all results per investigator assessment provided in the Bouffet 2023 publication (1). The assessor and co-assessor state in their request letter that the use of RANO-HGG criteria for investigator assessment is "*clear in the Bouffet 2023 publication, from the description of the methods (page 666), references (23 and 24) and results (Table 2); and clear from the Novartis Protocol for this study (specifically, Appendix 5 of the Protocol, which clearly outlines the use of RANO-HGG as the criteria used for investigator assessment of response, and Amendment 9, which clearly describes the addition of RANO-LGG only for independent assessment)".* However, the HTD disagrees with this conclusion of the assessor and co-assessor and finds it to be clear in the Bouffet 2023 publication (1) that both the independently- and investigator-assessed results presented are based on RANO-LGG criteria, in alignment with the information in Dossier V0.2. The HTD has thoroughly described the rationale underlying this below:

- The methods section in the Bouffet 2023 publication (1) states the following: *"For the BRAF V600–mutant LGG cohort, radiographic disease assessments were performed at baseline, every 8 weeks X 3, and then every 12 weeks and evaluated by independent radiology review and investigators using Response Assessment in Neuro-Oncology (RANO) criteria [with references 23 and 24 provided]; RANO 2017 criteria use T2 FLAIR magnetic resonance imaging sequences, were considered more relevant for paediatric LGG, and are presented here."* From this statement, it is clear to the HTD that all results based on radiographic disease assessment (including both ORR and PFS) presented in the Bouffet 2023 publication (1), whether that be per independent radiology review or investigator assessment, are based on RANO-LGG criteria (Wen 2017 (2); reference 24 in the Bouffet 2023 publication (1)). As described in Wen 2017 (p. 2444) (2), the RANO-LGG criteria are overall similar to RANO-HGG criteria, except that RANO-LGG criteria include a minor response category and measures T2/FLAIR rather than contrast enhancement since these tumours rarely enhance. This is also reflected in Table 26 of Dossier V0.2, from which it is clear that T2/FLAIR magnetic resonance imaging (MRI) is the imaging basis for RANO-LGG criteria, whereas contrast-enhanced T1 MRI is the imaging basis for RANO-HGG criteria. The underlined parts of the methods section statement in Bouffet 2023 (1) above therefore makes it clear that only RANO-LGG criteria are presented in the publication; there is no differentiation between independent radiology review and investigator assessment in the statement, and there is no direct mentioning of RANO-HGG criteria in the publication.

- Table 2 in the Bouffet 2023 publication (1) does not specify the criteria used for investigator assessments, as opposed to the case for the independent radiology review for which RANO 2017 criteria are specified in the table. Considering the statement in the methods section referred to above, and that it would appear misleading to present results side-by-side in a table without specifying differences in response criteria, the HTD concludes that both the results per independent radiology review and investigator assessment provided are based on RANO-LGG criteria. This is further supported by the fact that footnote b for "RANO 2017" in Table 2 states that "*Minor response category was not used in this clinical trial; patients meeting criteria for minor response were considered to have SD";* this is a relevant comment only if "RANO 2017" in the Bouffet 2023 publication (1) refers to the RANO-LGG criteria described in Wen 2017 (2), as the minor response category does not exist in the RANO-HGG criteria. The use of RANO-LGG criteria for investigator assessment is also supported by Figure S3 in the Supplementary Materials to the Bouffet 2023 publication (1); in this, it is clearly stated that the Kaplan-Meier plots showing PFS per investigator assessment are based on RANO 2017 criteria. It would be irrational to provide median and 24-month PFS results in Table 2 that were based on different criteria than the only Kaplan-Meier curve provided for investigator assessment (Figure S3), and the 24-month landmark estimate for PFS appears aligned between Table 2 and Figure S3.
- The HTD finds that the Novartis protocol for the study is not clear on several points; for example, PFS analyses are not even mentioned in this, and the statistical analyses are not detailed in the protocol, but only available in a statistical analysis plan (SAP) that is not part of the Supplementary Materials to the Bouffet 2023 publication (1). This precludes the HTD from having full visibility of details and changes in the analyses planned for ORR and PFS over the course of the study. It is acknowledged that specific text sections in Appendix 5 and Amendment 9 of the protocol could indicate that RANO-LGG criteria were used only for independent assessment of response, in alignment with the comment from the assessor and co-assessor. However, the RANO-LGG criteria described in Section 15.5.1 of Appendix 5 include descriptions of assessments to be done by the investigator, e.g. *"The evaluation of overall lesion response at each assessment is a composite of the target lesion response, non-target lesion response, the presence of new lesions, corticosteroid use relative to baseline, and clinical status as assessed by investigator and supported by the ECOG/Karnofsky Performance Scale".* Thus, the HTD finds it difficult to follow that the Novartis protocol for the study clearly outlines the use of RANO-HGG as the criteria used for investigator assessment of response, as stated in the assessor's and co-assessor's request.

In conclusion, the HTD acknowledges that on a general level, some unclarity exists around the exact methods used in the Bouffet 2023 study. However, based on above, the HTD finds it to be clear that the investigator-assessed data presented in the Bouffet 2023 publication (1) are based on RANO-LGG criteria, not RANO-HGG criteria, as specified in Wen 2017 (2). Emphasis is placed

on the methods section, Table 2 and Supplementary Figure S3 in the Bouffet 2023 publication (1) in coming to this conclusion, as opposed to the information in the publicly available study protocol, since this is a peer-reviewed source of information. Thus, it is not possible to meet the assessor's and co-assessor's first request for additional analyses and results specified in the document received on 14 November 2025.

2. Response to requests for providing analysis and results of relative effects on ORR and PFS based on RANO-HGG (INV) criteria (investigator assessment), to facilitate comparison with Bouffet 2023 using consistent RANO-HGG (INV) definitions, (as requested above) using pooled data from Arm 1 and Arm 2 of FIREFLY-1

As described above, it is clear to the HTD that RANO-HGG (INV) data are not available in the Bouffet 2023 publication (1). The HTD previously provided RANO-HGG (INV) data for Arm 2 of FIREFLY-1 as part of the response to the request for providing "relative effectiveness results for PICO 5 and PICO 7 considering the efficacy results available for patients in Arm 2 of clinical trial FIREFLY-1" included in the Commission's Second Request. However, the HTD also clearly explained that the Arm 2 efficacy data are **not** suitable for inclusion on descriptive level nor for conducting comparative effectiveness analyses for PICO 5 and PICO 7 in the Dossier. In summary, the sponsor (Day One Biopharmaceuticals, Inc.) did not conduct formal efficacy analyses for Arm 2, and no Analysis Data Model (ADaM) datasets data are available. The Arm 2 efficacy data provided as part of the response to the Commission's Second Request was based on retrospective efficacy analyses using Study Data Tabulation Model (SDTM), and these data were generated from a snapshot without full source data verification. The data were provided by the HTD to EMA on 7 October 2025, addressing question 197 in the Day 120 list of questions, but no pooling of Arm 1 and 2 data have been done. It is important to highlight also that the inclusion criteria for Arm 2 are not aligned with the intended labelled indication as per the assessment scope. Thus, it is not possible to meet the assessor's and co-assessor's second and third requests for additional analyses and results specified in the document received on 14 November 2025.

Reference list

1. Bouffet E, Geoerger B, Moertel C, Whitlock JA, Aerts I, Hargrave D, et al. Efficacy and Safety of Trametinib Monotherapy or in Combination With Dabrafenib in Pediatric BRAF V600-Mutant Low-Grade Glioma. J Clin Oncol. 2023;41(3):664-74.

2. Wen PY, Chang SM, Van den Bent MJ, Vogelbaum MA, Macdonald DR, Lee EQ. Response Assessment in Neuro-Oncology Clinical Trials. J Clin Oncol. 2017;35(21):2439-49.

D.1.3 Brief assessment by the assessors of the HTD's response to the first request for additional information

In response to the assessors' first additional information request, the HTD did not provide an updated dossier. The HTD's responses and arguments for not submitting the requested analyses are insufficient in the following respects:

- The HTD initially disagreed with the assessors that the INV-assessed tumour responses in Bouffet 2023 were labelled incorrectly in the analyses provided by the HTD. Corrected analyses were provided after a further request by the assessors (see Section D.2.2).
- The HTD's arguments for not submitting relative effectiveness analyses based on data from Arms 1 + 2 of FIREFLY-1 are insufficient:
  - The HTD argues that the inclusion criteria for Arm 2 of the study do not correspond to the intended indication according to the assessment scope, and that Arms 1 + 2 should therefore not be pooled. The assessors acknowledge that the inclusion criteria differ slightly between Arm 1 and Arm 2 in a way that would potentially allow patients that are not aligned with the intended labelled indication as per the assessment scope to be included in Arm 2 of the study. However, the reported characteristics of the patients in Arm 2 (which are also included for the safety analyses) clearly show that the patients are within the intended assessment scope and could therefore be included in a pooled analysis.
  - For patients in Arm 2 of FIREFLY-1, analyses for the same outcomes as prespecified for Arm 1 were predefined in the study protocol. The HTD provides no explanation as to why the sponsor of the study (Day One Biopharmaceuticals) did not conduct the efficacy analyses for Arm 2. The HTD did not provide sufficient justification for not carrying out the requested analyses themselves.

D.2 Second request for further specifications, clarifications or additional information

D.2.1 Assessors' second request for further specifications, clarifications or additional information to the HTD

JCA-MP-2024-06

INFORMATION, DATA, ANALYSES OR OTHER EVIDENCE TO BE REQUESTED FROM THE HTD

further to Article 11(2) of Regulation (EU) 2021/2282

The deadline for response, set by the assessor and co-assessor, is **7 days**

as per Article 12(5) of Commission Implementing Regulation (EU) 2024/1381

Request	Rationale
1. Analysis and results of relative effects on **ORR** based on **RANO-HGG (INV) criteria (investigator assessment),** to facilitate comparison with Bouffet 2023 using consistent RANO-HGG (INV) definitions. *The Bouffet data included in Table 57 of the Dossier V.2, but has been mislabelled as RANO-LGG (INV) when it should actually be RANO-HGG (INV). This is clear in the Bouffet 2023 publication, from the description of the methods (page 666), references (23 and 24) and results (Table 2); and clear from the Novartis Protocol for this study (specifically, Appendix 5 of the Protocol, which clearly outlines the use of RANO-HGG as the criteria used for investigator assessment of response, and Amendment 9, which clearly describes the addition of RANO-LGG only for independent assessment)* These analyses should be based on the following data: FIREFLY-1 Arm 1 (N=10, CR+PR, 0+4), Bouffet 2023 (N=36, CR+PR, 3+16).	There has been a mis-interpretation of the results reported in Bouffet 2023 (as outlined in the request herein). In the response to the first additional request addressing this question, the HTD does not agree that there has been a mis-interpretation. This conclusion by the HTD is based on the information available in the publication Bouffet 2023, as opposed to the information in the publicly available study protocol. The HTD suggests that a peer-reviewed publication is a more reliable source of information than a study protocol. The Assessors do not agree that a journal publication is more reliable than a study protocol. In addition, further information on the study provided by the sponsor of Bouffet 2023 confirms the use of RANO-HGG criteria for the data presented in the publication (see Table 6, right column in the document): https://www.g-ba.de/downloads/92-975-7639/2024_04_25_Modul4A_Dabrafenib_Anhang_4_H.pdf This document was created by Novartis, the HTD for dabrafenib + trametinib, and sponsor of the Bouffet 2023 study. Thus, the analyses were conducted as described in the study protocol. Regardless of the ambiguous description given in the publication and the corresponding supplement, the assessors understand that results relating to RANO-HGG criteria are presented in Bouffet 2023. Therefore, the requested analyses are feasible and necessary for the assessment and the assessors again request the analyses already specified in the first additional request.
2. Analysis and results of relative effects on **ORR** based on **RANO-HGG (INV) criteria (investigator assessment),** to facilitate comparison with Bouffet 2023 using consistent RANO-HGG (INV)	Given the similarity of the patients in the full analysis sets of Arm 1 and Arm 2, their relevance to the population defined by PICO 5, and the availability of RANO-HGG (INV) based response assessments, pooling of results from these two arms is requested.

definitions, (as requested above) using pooled data from Arm 1 and Arm 2 of FIREFLY-1.	See point 1.
3. Analysis and results of relative effects on **PFS** based on **RANO-HGG (INV)** criteria (investigator assessment), to facilitate comparison with Bouffet 2023 using consistent RANO-HGG (INV) definitions, (as requested above) using pooled data from Arm 1 and Arm 2 of FIREFLY-1.	Extension of the evidence to Arm 2 in the FIREFLY-1 trial would facilitate additional PFS comparison with Bouffet 2023, based on RANO-HGG (INV) criteria See point 1.

Further request for additional information on safety outcomes:

Request	**Rationale**
4. Corrected analyses on all safety outcomes for PICO 5.	The analyses on safety outcomes submitted by the HTD appear to be incorrect. For example, for treatment interruption due to AE, a large effect estimate was reported while the proportion of patients with events did not differ substantially between the studies (see dossier Appendix A.2, Table 17). The assessors were not able to reproduce the effect estimates shown for the naïve comparisons for most of the safety outcomes (i.e., serious AE, treatment discontinuation due to AE and treatment interruption due to AE), whereas the results for severe AE [CTCAE grade ≥ 3] were reproducible. Therefore, corrected analyses are necessary for the safety outcomes, including a quality check of the base case analyses for all safety outcomes.

D.2.2 Response of the HTD to the second request for further specifications, clarifications or additional information

Regulation (EU) 2021/2282 on Health Technology Assessment

Dossier of the Joint Clinical Assessment (JCA) of a Medicinal Product (MP): Request for information, data, analyses or other evidence from the assessor and co-assessor - health technology developer (HTD) cover note

Tovorafenib (Ojemda®)

Ipsen Pharma

Code: JCA-MP-2024-06

Version 0.3

Date submitted: 5 December 2025

Pursuant to the template set out in Annex I of **Commission Implementing Regulation (EU) 2024/1381**

Ipsen Pharma (the health technology developer [HTD]) appreciates the opportunity to provide additional information related to the ongoing assessment of tovorafenib (Ojemda®), code JCA-MP-2024-06, as requested by the assessor and co-assessor in the "Second request for information, data, analyses or other evidence from the assessor and co-assessor of the joint clinical assessment of the medicinal product Ojemda (Tovorafenib), case No JCA-MP-2024-06" document received on 28 November 2025. Please find below a description of how each of the requests have been addressed and of the dossier revisions implemented in relation to these.

Table 1: Summary of HTD responses to the assessor's and co-assessors request for information, data, analyses or other evidence

Request	Rationale	HTD response to request
1. Analysis and results of relative effects on **overall response rate (ORR)** based on Response Assessment in Neuro-Oncology for high-grade gliomas (**RANO-HGG) (INV) criteria (investigator assessment)** to facilitate comparison with Bouffet 2023 using consistent RANO-HGG (INV) definitions. ***The Bouffet data*** *included in Table 57 of the Dossier V.2, but has been mislabelled as RANO-LGG (INV) when it should actually be RANO-HGG (INV). This is clear in the Bouffet 2023 publication, from the description of the methods (page 666), references (23 and 24) and results (Table 2); and clear from the Novartis Protocol for this study (specifically, Appendix 5 of the Protocol, which clearly outlines the use of RANO-HGG as the criteria used for investigator assessment of response, and Amendment 9, which clearly describes the addition of RANO-LGG only for independent assessment).* These analyses should be based on the following data: FIREFLY-1 Arm 1 (N=10, CR+PR, 0+4), Bouffet 2023 (N=36, CR+PR, 3+16).	There has been a mis-interpretation of the results reported in Bouffet 2023 (as outlined in the request herein). In the response to the first additional request addressing this question, the HTD does not agree that there has been a mis-interpretation. This conclusion by the HTD is based on the information available in the publication Bouffet 2023, as opposed to the information in the publicly available study protocol. The HTD suggests that a peer-reviewed publication is a more reliable source of information than a study protocol. The Assessors do not agree that a journal publication is more reliable than a study protocol. In addition, further information on the study provided by the sponsor of Bouffet 2023 confirms the use of RANO-HGG criteria for the data presented in the publication (see Table 6, right column in the document): https://www.g-ba.de/downloads/92-975-7639/2024_04_25_Modul4A_Dabrafenib_Anhang_4_H.pdf This document was created by Novartis, the HTD for dabrafenib + trametinib, and sponsor of the Bouffet 2023 study. Thus, the analyses were conducted as described in the study protocol. Regardless of the ambiguous description given in the publication and the corresponding supplement, the assessors understand that results relating to RANO-HGG criteria are presented in Bouffet 2023. Therefore, the requested analyses are feasible and necessary for the assessment and the assessors again request the analyses already specified in the first additional request.	The HTD appreciates that the assessors recognize the ambiguity in the publicly available study materials and the Bouffet 2023 publication (1), which led to the HTD's earlier conclusion that RANO-LGG criteria were used for both independently- and investigator-assessed results presented in Bouffet 2023 (1). The HTD previously provided multiple strong justifications for this interpretation in response to the first additional request addressing this question, demonstrating that the submitted evidence reflected the HTD's best understanding at the time. However, the HTD agrees with the assessors that the Bouffet 2023 study information in the Norvartis dossier submitted to The Federal Joint Committee (G-BA) (2) confirms the use of RANO-HGG criteria for the investigator-assessed data presented in Bouffet 2023 (1). Unlike the Bouffet 2023 publication (1), the Novartis dossier (2) explicitly states that RANO 2010 (i.e. RANO-HGG) criteria were used for investigator-based results, and the data are fully aligned between both sources (Bouffet 2023 publication (1) and Novartis G-BA submission (2). The HTD has therefore conducted new **ORR** analyses based on **RANO-HGG criteria per investigator (INV) assessment** vs. Bouffet 2023 (1) as requested. Furthermore, to ensure completeness and address the variability observed in ORR between RANO-HGG INV and independent review committee (IRC) assessments in the Bouffet 2023 study (1), the HTD has also conducted additional **ORR** analyses based on **RANO-HGG criteria per IRC assessment** vs. Bouffet 2023, using the RANO-HGG IRC results available in the Novartis dossier (2). It should be

		noted that RANO-HGG IRC results are not reported in the Bouffet 2023 publication (1); the publication includes IRC results only per RANO 2017 (i.e. RANO-LGG) criteria. As the RANO-HGG IRC results for the Bouffet 2023 study are available solely in a document type that falls outside the literature search scope requirements outlined in the "Guidance on filling in the joint clinical assessment (JCA) dossier template", the HTD understands that these results, as well as the indirect treatment comparison (ITC) analyses based on them, would not be appropriate for inclusion in the actual dossier. Therefore, these analyses are provided exclusively in this cover note; please refer to Section 1 below. It should be noted that the availability of **RANO-HGG IRC** results for the Bouffet 2023 study within the Novartis dossier (2) has influenced the relevance of one of the previously submitted ITC sensitivity analyses for **ORR** for PICO 5, i.e. the one based on **RANO-HGG IRC** results for **FIREFLY-1** vs. **RANO-LGG IRC** results for Bouffet 2023. As **ORR** analyses using **RANO-HGG IRC** results consistently for both studies were now feasible, this previously provided sensitivity analysis is no longer considered relevant and has been removed across the materials. In accordance with above, the following have been updated: - Sections 1.3, 5.2 and 5.3 of the main document - Appendix A.1 (for Bouffet 2023 only) - Appendix A.2 - Appendix B.2 - Appendix D.5 (both the PAIC_Evidence synthesis report and PAIC_SAP documents)

2. Analysis and results of relative effects on **ORR** based on **RANO-HGG (INV) criteria (investigator assessment)**, to facilitate comparison with Bouffet 2023 using consistent RANO-HGG (INV) definitions, (as requested above) using pooled data from Arm 1 and Arm 2 of FIREFLY-1.	Given the similarity of the patients in the full analysis sets of Arm 1 and Arm 2, their relevance to the population defined by PICO 5, and the availability of RANO-HGG (INV) based response assessments, pooling of results from these two arms is requested. See point 1.	Although the HTD acknowledges that RANO-HGG criteria were used for the investigator-assessed results presented in the Bouffet 2023 publication (1), it remains infeasible to fulfill the assessors' request to conduct efficacy analyses vs. Bouffet 2023 (1) using pooled data from Arm 1 and Arm 2 of FIREFLY-1. The key rationales for this position are summarised here; please refer to Section 2 below for further details. In accordance with the FIREFLY-1 study protocol, pooling of efficacy data from Arm 1 and Arm 2 was not planned; therefore, such evidence **does not exist** and **was not submitted to the European Medicines Agency (EMA).** As previously explained in responses to the Commission's Second Request on Arm 2 efficacy data from FIREFLY-1 and the first additional request on conducting indirect treatment comparison (ITC) analyses vs. Bouffet 2023 (1) using pooled efficacy data, the sponsor of FIREFLY-1 (Day One Biopharmaceuticals, Inc.) did not conduct formal efficacy analyses for Arm 2, and no Analysis Data Model (ADaM) datasets are available. The Arm 2 RANO-HGG (INV) data provided in earlier responses to the EMA Day 120 list of questions and the Commission's Second Request were based on retrospective analyses using Study Data Tabulation Model (SDTM). **These data were generated from a snapshot without full source data verification (SDV)**. Consequently, to meet the assessors' request, the HTD would need to conduct ITC analyses using mixed datasets (cleaned SDV'ed data for Arm 1 and uncleaned non-SDV'ed data for Arm 2). Such an approach would be highly inappropriate and inconsistent with International Council for Harmonisation of Technical Requirements for

		Pharmaceuticals for Human Use (ICH)-E6(R3) guidelines (3). In Section 2 below, the HTD has further expanded its explanation on why pooling of data from Arm 1 and 2 is not methodologically appropriate and why the requested analyses cannot be expected to meaningfully strengthen the evidence base for the assessment. It is also described that these analyses are considered outside the defined scope of the JCA as per the applicable regulation. No dossier revisions have been implemented in response to this request.
3. Analysis and results of relative effects on **progression-free survival (PFS)** based on **RANO-HGG (INV) criteria (investigator assessment)**, to facilitate comparison with Bouffet 2023 using consistent RANO-HGG (INV) definitions, (as requested above) using pooled data from Arm 1 and Arm 2 of FIREFLY-1.	Extension of the evidence to Arm 2 in the FIREFLY-1 trial would facilitate additional PFS comparison with Bouffet 2023, based on RANO-HGG (INV) criteria. See point 1.	Please refer to response to request no. 2 above and Section 2 below. No dossier revisions have been implemented in response to this request.
4. Corrected analyses on all safety outcomes for PICO 5.	The analyses on safety outcomes submitted by the HTD appear to be incorrect. For example, for treatment interruption due to AE, a large effect estimate was reported while the proportion of patients with events did not differ substantially between the studies (see dossier Appendix A.2, Table 17). The assessors were not able to reproduce the effect estimates shown for the naïve comparisons for most of the safety outcomes (i.e., serious AE, treatment discontinuation due to AE and treatment interruption due to AE), whereas the results for severe AE [CTCAE grade ≥ 3] were reproducible. Therefore, corrected analyses are necessary for the safety outcomes, including a quality check of the base case analyses for all safety outcomes	The HTD sincerely apologises for the data error identified in the previous submission based on this request from the assessors. The HTD fully acknowledges the importance of accuracy and transparency and regret any inconvenience this may have caused. Upon review, an error was identified in the input data related to the two analyses of patients experiencing adverse events (AEs) leading to treatment discontinuation and treatment interruption, respectively; the event rates for these two outcomes in the Bouffet 2023 study had been switched around in the analyses. The HTD has corrected the error for these two outcomes and conducted additional quality checks of the base case analyses for all safety outcomes; accuracy of previously submitted results for all other safety outcomes were confirmed. Please refer to Section 3 below for a description of the additional quality check steps taken.

		To address the assessors concerns on the reproducibility of the outcome for serious AEs, the HTD identified a need for clarifying the discrepancy between the FIREFLY-1 data presented for serious AEs in Appendix A.2, Table 17 and Appendix B.2, Table 32; this may explain why the assessors were not able to reproduce the effect estimates for this outcome. Please refer to Section 3 below for more details. The HTD apologises for any confusion this may have caused; additional language to clarify the approach taken has been added in Section 4.3.4.2 of the main document and in Appendixes A.2. and B.2.

1. Analysis and results of relative effects on ORR based on RANO-HGG criteria per IRC assessment for PICO 5

As noted above, to ensure completeness and address the variability observed in ORR between RANO-HGG INV and IRC assessments in the Bouffet 2023 study (1), the HTD has conducted additional ORR analyses based on RANO-HGG criteria per IRC assessment vs. Bouffet 2023 (1), using the RANO-HGG IRC results available in the Novartis dossier submitted to The Federal Joint Committee (G-BA) (2). The data inputs and results for these analyses are provided in Table 2 and described below.

Table 2: Relative effectiveness results (ORR) based on RANO-HGG criteria per IRC assessment for subpopulation 1 (BRAF V600E mutation in patients > 1 year) – PICO 5 – unanchored population-adjusted indirect comparison: tovorafenib vs. combination of dabrafenib and trametinib

Time point Outcome Comparison Scenario	Tovorafenib FIREFLY-1 (4)			Dabrafenib + trametinib Bouffet 2023 (2)			Group difference		
	Tumour response assessment criteria and method	N/N (with event)	% (95% CI)	Tumour response assessment criteria and method	N/N (with event)	% (95% CI)	ESS	OR[d] (95 %-CI) p-value	Shifted null hypothesis δ[e] (E-value)
Data-cut	**2-year**			**NR**					
ORR (CR+PR)	RANO-HGG (IRC)	10/5	50.0 (18.7, 81.3)	RANO-HGG (IRC)	36/7	19.4 (8.2, 36.0)	NA	NA	NA
Indirect comparison (PAIC) of tovorafenib and dabrafenib + trametinib:									
Naïve comparison[a]	RANO-HGG (IRC)	NA	NA	RANO-HGG (IRC)	NA	NA	10	4.14 (0.93, 18.37) p=0.061	-0.06 (7.75)
Base case[b]	RANO-HGG (IRC)	NA	NA	RANO-HGG (IRC)	NA	NA	6.64	7.26 (0.98, 53.68) p=0.052	-0.01 (14.01)
Scenario 1[c]	RANO-HGG (IRC)	NA	NA	RANO-HGG (IRC)	NA	NA	10	4.16 (0.83, 21.01) p=0.084	-0.19 (7.79)

a: Unweighted.
b: Adjustment for all mutually reported TEMs/PVs: age (median), prior surgery, Karnofsky/Lansky (combined; 80-100).
c: Adjustment for ranked variables by KOLs: age (median).
d: OR>1 presents an advantage for tovorafenib; 95% CIs excluding 1 and a p-value <0.05 present a statistically significant result.
e: Shifted null hypothesis δ of zero (or close to zero) and E-values >2 confirms robust model output (i.e., unaffected from unmeasured confounding).
CI: confidence interval; CR: complete response; ESS: effective sample size; IRC, independent review committee; KOL: key opinion leader; NR: not reported; OR: odds ratio; ORR: objective response rate; PR: partial response; PV, prognostic value; TEM, treatment effect modifier.

As of the 2 year data cutoff date (10 May 2024), the IRC-assessed ORR (CR or PR) for tovorafenib based on RANO-HGG criteria was 50.0% (95% confidence interval [CI]: 18.7, 81.3) (4). The IRC-assessed ORR (CR or PR) for patients treated with dabrafenib + trametinib in the Bouffet 2023 study based on RANO-HGG criteria was 19.4% (95% CI: 8.2, 36.0) (data-cut not reported) (2).

In the population-adjusted indirect comparisons (PAICs) for ORR of tovorafenib vs. dabrafenib + trametinib for patients > 1 year with the BRAF V600E mutation, the post-matching effective sample size (ESS) for tovorafenib was 6.64 for the base case (adjusted for all mutually reported TEM/PVs; age, prior surgery, Karnofsky/Lansky) and 10 for scenario analysis 1 (adjusted for ranked variables by KOLs; age), respectively, based on IRC-assessed RANO-HGG criteria (Table 2).

In the naïve comparison (using IRC-assessed RANO-HGG criteria), tovorafenib had a numerically higher ORR compared with dabrafenib + trametinib (OR [95% CI]: 4.14 [0.93, 18.37]; p=0.061; δ [E-value] = -0.06 [7.75]). In both the base case analysis and scenario analysis 1, tovorafenib also had a numerically higher ORR compared with dabrafenib + trametinib (base case OR [95% CI]: 7.26 [0.98, 53.68]; p=0.052; δ [E-value] = -0.01 [14.01] and scenario analysis 1 OR [95% CI]: 4.16 [0.83, 21.01]; p=0.084; δ [E-value] = -0.19 [7.79]). These findings suggest a clinical benefit of tovorafenib vs. dabrafenib + trametinib in the BRAF V600E mutation population; however, observed CIs reflect statistical uncertainty primarily due to the small sample size and the reduced ESS following matching in the PAIC. When adjusting for age only in scenario analysis 1, the ESS remained similar to the original sample size, highlighting statistical precision and direction of the findings. Overall, the analysis demonstrated that covariate balance was achieved successfully – without impacting other baselines, although the reduced ESS highlights that results should be interpreted with caution given the limited precision.

Robustness of the results was further explored through shifted null hypothesis testing. The estimated deltas of the shifted null hypotheses remained close to zero, suggesting that the findings are not sensitive to unmeasured confounding. Furthermore, the calculated E-values were greater than 2, indicating that substantial unmeasured confounding would be required to completely negate the observed treatment effect. These metrics confirm that the results are directionally robust, despite the statistical imprecision reflected in the wide CIs due to the small sample size and ESS.

It should be noted that the relative effectiveness results for ORR based on IRC-assessed RANO-HGG criteria for PICO 5 are directionally different from the relative effectiveness results for ORR based on INV-assessed RANO-HGG criteria, which were added to the dossier as requested; the relative effectiveness results for ORR based on INV-assessed RANO-HGG criteria for PICO 5 showed that tovorafenib had a numerically lower ORR compared with dabrafenib + trametinib (please refer to dossier Section 5.3, appendix B.2 and appendix D.5 for detailed results of this

analysis). The difference in direction of results for IRC-based and INV-based RANO-HGG analyses is caused by two key factors: 1) there is a one-event difference between IRC-assessed and INV-assessed RANO-HGG results in FIREFLY-1, but the small sample size means that this translates to an ORR decrease from 50.0% per IRC assessment to 40.0% per INV assessment (4) and 2) there is a major increase in the ORR per INV assessment vs. IRC assessment based on RANO-HGG criteria in the Bouffet 2023 study (52.8% vs. 19.4%) (2). This magnitude of difference in ORR between IRC and INV assessment translates into a change in direction of the PAIC results.

2. Response to requests for providing analysis and results of relative effects on ORR and PFS based on RANO-HGG (INV) criteria (investigator assessment), to facilitate comparison with Bouffet 2023 using consistent RANO-HGG (INV) definitions, (as requested above) using pooled data from Arm 1 and Arm 2 of FIREFLY-1

The HTD would first like to emphasize that, as specified in the FIREFLY-1 study protocol, **there were no predefined objectives to pool efficacy data from Arm 1 and Arm 2**. Therefore, the analyses requested by the assessors would rely on evidence that does not exist and was not submitted to the European Medicines Agency (EMA). In contrast to Arm 1, Arm 2 was designed as an expanded access cohort to provide continued treatment and collect supplementary safety data, not for efficacy evaluation. Detailed rationales for the omission of Arm 2 efficacy data from FIREFLY-1 in the European Medicines Agency (EMA) submission and the JCA dossier were previously provided in response to question 197 in the EMA Day 120 list of questions and the Commission's Second Request on Arm 2 efficacy data. EMA rapporteur acknowledged the limitations of any Arm 2 efficacy analyses in early assessment, and no follow-up questions on this are expected in the Day 180 list of questions. There have been no requests from EMA to submit pooled Arm 1 and 2 efficacy data at any point in the process.

While the assessors note that "*Given the similarity of the patients in the full analysis sets of Arm 1 and Arm 2, their relevance to the population defined by PICO 5, and the availability of RANO-HGG (INV) based response assessments, pooling of results from these two arms is requested",* the HTD would like to clarify that there were differences in the patient populations enrolled in Arm 1 and Arm 2. Arm 2 enrolled patients with low-grade glioma harboring activating RAF alterations, including BRAF or CRAF/RAF1 fusions and BRAF V600 mutations, and distinct inclusion and exclusion criteria between the two arms could introduce confounding variables and compromise the robustness of any efficacy analyses. The relevance of Arm 2 data in supporting efficacy claims for the intended labelled indication as per the assessment scope is therefore limited.

In addition, as previously explained in responses to the Commission's Second Request on Arm 2 efficacy data from FIREFLY-1 and the first additional request on conducting ITC analyses vs. Bouffet 2023 using pooled efficacy data, the sponsor of FIREFLY-1 (Day One Biopharmaceuticals, Inc.) did not conduct formal efficacy analyses for Arm 2, and no Analysis Data Model (ADaM)

datasets are available. The Arm 2 RANO-HGG (INV) data provided in earlier responses to the EMA Day 120 list of questions and the Commission's Second Request were based on retrospective analyses using SDTM. These data were generated from a snapshot without full SDV. Consequently, to meet the assessors' request, the HTD would need to conduct ITC analyses using mixed datasets (cleaned SDV'ed data for Arm 1 and uncleaned non-SDV'ed data for Arm 2). Such an approach would be highly inappropriate and inconsistent with ICH-E6(R3) guidelines (3).

The limited relevance of including Arm 2 efficacy data in comparative analyses in the dossier is further supported by the JCA guidance document "Methodological Guideline for Quantitative Evidence Synthesis: Direct and Indirect Comparisons", which states that *"In conducting treatment comparisons, bias must be minimised. Bias reflects a systematic error in the results and results in deviation of the estimated treatment effectiveness from the true treatment effectiveness. When performing evidence synthesis, some key potential sources of bias should be considered.* ***The first is bias in the results of the individual studies included in the review. If the individual study results are biased, then a synthesised summary of the individual results will also be biased and can yield misleading conclusions****".* In this specific case, including Arm 2 efficacy data could bias the results of FIREFLY-1 and, consequently, the evidence synthesis/comparative efficacy analyses.

Finally, the HTD's understanding is that the assessors' request to include such analyses in the dossier falls outside the scope of the JCA as defined in Regulation (EU) 2021/2282. As noted above, pooled Arm 1 and Arm 2 efficacy data do not exist and were not submitted to EMA. In accordance with Annex I, point (b), of Regulation (EU) 2021/2282, the dossier shall include all up-to-date published and unpublished information, data, analyses and other evidence as well as study reports and study protocols and analysis plans from studies with the medicinal product for which the health technology developer was a sponsor. The HTD has complied with these requirements and previously explained in response to the Commission's Second Request why Arm 2 efficacy data were not suitable for inclusion in the dossier. This rationale also applies to pooled Arm 1 and Arm 2 efficacy data. Pooled Arm 1 and Arm 2 efficacy data do not exist and any analyses using such data cannot be expected to meaningfully strengthen the evidence base for the assessment.

3. Response to request for conducting additional quality checks of safety outcomes for PICO 5

As mentioned earlier, the HTD has conducted additional quality checks of the base case analyses for all safety outcomes, which led to identification of an error in the input data related to the two analyses of patients experiencing AEs leading to treatment discontinuation and treatment interruption, respectively; the event rates for these two outcomes in the Bouffet 2023 study (8/36 and 26/36) had been switched around in the analyses. Data accuracy of previously submitted results for all other safety outcomes were confirmed. Please find below a description of the additional quality check procedure conducted:

- Two statisticians independently extracted and reviewed the data, then programmed and executed the analyses separately.
- A reconciliation meeting was held to compare outputs. In cases of discrepancies, a third statistician was engaged to review and reconcile the programming code.
- A fourth statistician conducted an independent analysis of raw rates to provide an extra layer of verification and confirm consistency with the other analyses.

This multi-tiered approach — combining blinded programming with independent verification — provides strong assurance that the results are accurate for all outcomes.

Open-envelope calculations for the unweighted results are provided in Table 3 for all safety analyses.

Table 3: Summary of open-envelope calculations for the unweighted results for safety analyses

Analyses	AE Grade ≥3	Serious AE	AE leading to treatment discontinuation	AE leading to treatment interruption
Dabrafenib + trametinib (Bouffet 2023)	22/36	15/36	8/36	26/36
Tovorafenib (FIREFLY-1)	20/22	13/22	3/22	15/22
ORR rate	6.36	2.02	0.55	0.82
SE*	0.82	0.55	0.74	0.59
95% CI	1.28, 31.53	0.69, 5.94	0.13, 2.35	0.26, 2.62

*SE calculated using normal approximation.

Definitions:

SE(logORRrate) = sqrt(1/a + 1/(n1-a) + 1/b + 1/(n2-b))

95% CI = EXP(log[ORRrate] +- NORM.S.INV(0.975) * SE(logORRrate)

To address the assessors concerns on the reproducibility of the outcome for serious AEs, the HTD identified a need for clarifying the discrepancy between the FIREFLY-1 data presented for serious AEs in Appendix A.2, Table 17 and Appendix B.2, Table 32. Bouffet 2023 does not specifically report treatment-emergent AEs (TEAEs) but any adverse events. To ensure consistency with the AE definition used in Bouffet 2023(1), as understood from the publication, the AEs in FIREFLY-1 were recalculated without restriction to TEAEs for the purpose of conducting the relative safety analyses vs. Bouffet 2023 (1). This translated into one more event for the serious adverse events in Appendix B.2, Table 32 compared to Appendix A.2, Table 17, which may explain why the assessors were not able to reproduce the effect estimates shown for this outcome. Only TEAE data have been included for FIREFLY-1 in Appendix A.2 to remain consistent with the clinical study report (CSR). The HTD apologises for any confusion this may have caused; additional language to clarify the approach taken have been added in Section 4.3.4.2 of the main document and in Appendixes A.2. and B.2.

Reference list

1. Bouffet E, Geoerger B, Moertel C, Whitlock JA, Aerts I, Hargrave D, et al. Efficacy and Safety of Trametinib Monotherapy or in Combination With Dabrafenib in Pediatric BRAF V600-Mutant Low-Grade Glioma. J Clin Oncol. 2023;41(3):664-74.
2. Gemeinsamer Bundesausschuss. Benefit assessment procedure for the active ingredient trametinib (malignant glioma, BRAF V600E mutation, ≥ 1 year, low-grade (LGG)/high-grade (HGG) after at least 1 prior therapy; combination with dabrafenib); Module 4A, Appendix 4-H 2024. Available from: https://www.g-ba.de/downloads/92-975-7639/2024_04_25_Modul4A_Dabrafenib_Anhang_4_H.pdf.
3. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). ICH HARMONISED GUIDELINE. GUIDELINE FOR GOOD CLINICAL PRACTICE: E6(R3) 2025. Available from: https://database.ich.org/sites/default/files/ICH_E6%28R3%29_Step4_FinalGuideline_2025_0106.pdf.
4. Ipsen Pharma. Clinical Study Report - FIREFLY-1 (DCO 10 May 2024). Data on file; 2025 06 Jan.

D.2.3 Brief assessment by the assessors of the HTD's response to the second request for additional information

In response to the assessors' second additional information request, the HTD provided an updated dossier. Several aspects of the response are commented on briefly:

- The HTD provided corrected safety analyses.
- The HTD provided analyses for ORR based on RANO-HGG criteria (using comparable definitions).
- The HTD further elaborates on why the requested pooled relative effectiveness analyses based on data from Arms 1 + 2 of FIREFLY-1 were not submitted:
  - The HTD mentions that there were no predefined objectives to pool efficacy data from Arm 1 and Arm 2. The assessors point out that even though a pooled analysis was not prespecified, it could still be informative, especially since the patient numbers for the relevant subpopulation are very low. Furthermore, according to the study protocol of FIREFLY-1, patients in Arm 2 underwent the same tumour assessment and were to be evaluated via the same statistical methods as described for Arm 1.
  - According to the HTD, unlike for Arm 1, no full source data verification was performed for Arm 2 and it would therefore not be methodologically appropriate to perform analyses with pooled data from both study arms. The assessors do not agree that the lack of full source data verification should prevent the analyses requested for the JCA.