Document Type:	Statistical Analysis Plan
Official Title:	EPI VITRAKVI: A comparison of clinical outcomes in Infantile Fibrosarcoma (IFS) patients treated with larotrectinib in the phase I/II SCOUT study versus (an) external historical cohort(s)
NCT Number:	NCT05236257
Document Date:	28 JUL 2022

Statistical analysis plan (sec. data collection)

Study Title EPI VITRAKVI: A comparison of clinical outcomes in Infantile Fibrosarcoma (IFS) patients treated with larotrectinib in the phase I/II SCOUT study *versus* (an) external historical cohort(s)			
Bayer Study Drug	BAY 2757556 / Larotrectinib / Vitrakvi®		
Study Purpose:	The *Haute Autorité de Santé* (HAS), French National Authority for Health, provided positive feedback on the reimbursement of larotrectinib only in a subpopulation of EMA approved indication, i.e. for the treatment of pediatric patients with IFS or another Soft Tissue Sarcoma (STS), harboring a Neurotrophic Tyrosine Receptor Kinase (*NTRK*) gene fusion, which is locally advanced or metastatic, and refractory or in relapse. HAS provided a conditional reimbursement and requested comparative data of larotrectinib treatment *versus* standard of care at least *versus* an external control arm, in particular in IFS for a future reassessment. This study is designed to address this request.		
IMPACT No.:	21767	**Protocol Version/Date:**	v 1.0, 27 JUL 2021
SAP Version, Date:	v 3.0, 28 JUL 2022	**SAP includes TLF specification:**	yes ☒ no ☐
Author:	PPD Bayer U.S. LLC100 Bayer Boulevard, Whippany, 07981 NJ, USA		

Throughout this document, symbols indicating proprietary names (®, TM) are not displayed. Hence, the appearance of product names without these symbols does not imply that these names are not protected.

Table of Contents

Abbreviations

AE	Adverse Event
BIS	*Bis In Die*
CNS	Central Nervous System
CR	Complete Response
eCRF	Electronic Case Report Form
CSP	Clinical Study Protocol
CWS	*Cooperative Weichteilsarkom Studiengruppe*
DCR	Disease Control Rate
DLT	Dose Limiting Toxicity
EFS	Event-Free Survival
EMA	European Medicines Agency
FAS	Full Analysis Set
GU	Genito-Urinary
HAS	*Haute Autorité de Santé* (French National Authority for Health)
HGG	High-Grade Glioma
HR	Hazard Ratio
IFS	Infantile Fibrosarcoma
IPTW	Inverse Probability Treatment Weighting
IRS	Intergroup Rhabdomyosarcoma Study
KDE	Kernel Density Estimate
LGG	Low-Grade Glioma
LKAD	Last Known Alive Date
MA	Marketing Authorization
MedDRA	Medical Dictionary for Regulatory Activities
MTD	Maximum Tolerated Dose
NTRK	Neurotrophic Tyrosine Receptor Kinase
ORR	Overall Response Rate
OS	Overall Survival
OVL	Overlapping Coefficient
PD	Progressive Disease
PFS	Progression-Free Survival
PH	Proportional Hazards
PIP	Pediatric Investigation Plan
PM	Parameningeal
PR	Partial Response
PS	Propensity Score
RECIST	Response Evaluation Criteria in Solid Tumors
RR	Risk Ratio
SAP	Statistical Analysis Plan
SD	Stable Disease
SLR	Systematic Literature Review
SOP	Standard Operating Procedure
STS	Soft Tissue Sarcoma
TLF	Tables, Listings and Figures
TRK	Tropomyosin Receptor Kinase
TRKi	TRK Inhibitor
USA	United States of America
WHO-DD	World Health Organization – Drug Dictionary

1. Introduction

The clinical development program of the first-in-class selective Tropomyosin Receptor Kinase inhibitor (TRKi) larotrectinib, including the pediatric Phase I/II SCOUT study[1], to date has shown that the tumor response observed under treatment with the drug, is pronounced and durable based on results from single arm trials.

In September 2019, larotrectinib was granted a conditional Marketing Authorization (MA) in Europe as monotherapy for the treatment of adult and pediatric patients with solid tumors that display a *NTRK* gene fusion, who have a disease that is locally advanced, metastatic or where surgical resection is likely to result in severe morbidity, and who have no satisfactory treatment options. In France, a dossier was submitted to the National Authority for Health (*Haute Autorité de Santé* [HAS]) in October 2019 to obtain reimbursement of larotrectinib in the country in line with the targeted population described in the MA. In July 2020, the HAS provided positive feedback on the reimbursement of larotrectinib only in pediatric patients with IFS or another STS harboring a *NTRK* gene fusion, which is locally advanced or metastatic, and refractory or in relapse. However, this appraisal is conditional upon the provision of comparative data of larotrectinib treatment *versus* standard of care, in particular in IFS. This study is designed to address this request but also to possibly use the data to provide comparison with historical data as agreed in the pediatric investigation plan (PIP) with European Medicines Agency (EMA).

Given the rarity of the condition and ethics considerations, the use of an external control arm allows for the timely generation of comparative data *versus* SCOUT study data. However, patients who received larotrectinib in the SCOUT trial will likely differ in various underlying characteristics from those who are recorded in the historical control cohorts. This presents a key statistical challenge of adjusting the analysis of primary and secondary endpoints for these patient characteristics in order to avoid biased results.

1.1 Background

Please refer to the protocol for a thorough account of the rationale and background of this study.

1.2 Protocol Version and Amendments

Study 21767/EPI VITRAKVI, v 1.0, 27 JULY 2021.

2. Study Objectives

The primary objective of this study is:

- to compare the time to medical treatment failure (defined as: next systemic treatment or mutilating surgery or radiation therapy or death due to any cause) between larotrectinib and

[1] SCOUT study, a Phase I/II study of the oral tropomyosin receptor kinase (TRK) inhibitor larotrectinib in pediatric patients with advanced solid or primary central nervous system tumors – international clinical study.

standard of care in IFS patients using externally-controlled comparison performed with phase I/II SCOUT study and eligible external historical cohort(s).

The secondary objectives of this study include the comparison of:

- Treatment outcomes (next systemic treatment, mutilating surgery, radiation therapy, death due to any cause),
- Treatment discontinuation rates due to toxicity.

Exploratory endpoints include:

- Progression-free survival (PFS): defined as the time (months) from the start of treatment to the date of first observed disease progression (radiological or clinical, whichever is earlier) or death due to any cause, if death occurs before progression is documented.
- Overall response rate (ORR): defined as the proportion of patients with a best overall response of complete response (CR) or partial response (PR) assessed by investigators.
- Disease control Rate (DCR): defined as the proportion of patients with a best overall response of CR, PR, or Stable Disease (SD).

3. Study Design

This study is a retrospective observational externally-controlled study. Data from patients with IFS within the following the two groups will be compared:

- From the international phase I/II clinical trial called 'SCOUT' conducted in a pediatric population (up to 21 years old) with advanced solid or primary central nervous system (CNS) tumors. Data are collected via electronic case report forms (eCRFs) completed by the investigators during the study conduct in each site.
- External historical control cohorts including:
  - *Institut Curie* database.
  - Database from the *Cooperative Weichteilsarkom Studiengruppe* (CWS, Germany).

3.1 SCOUT study

The pediatric subgroup of patients from both phase I and II of the SCOUT study, having IFS, or with other STS for complementary analyses described in sections 5.1.1 and 5.1.2, will form the larotrectinib population in the analysis.

3.1.1 Phase 1 Dose Escalation and Dose Expansion

Phase 1 is a sequential-cohort dose escalation study designed to identify the maximum tolerated dose (MTD) through incidence and characterization of DLT. The adverse events (AE) defining dose limiting toxicity (DLT) are defined in integrated clinical study protocol (CSP) amendment, version 12.0, section 3.1.1.4.

Larotrectinib is administered twice daily (BID) in either an oral liquid or capsule form. Escalation in Phase 1 was to proceed through 5 planned dose levels or until the MTD was reached. When the optimal dose was thus identified, a Phase 1 expansion cohort of up to 18 patients was recruited to further

define the safety profile. Further details on the design of Phase 1 can be found in the integrated CSP amendment.

Individual patients will continue daily larotrectinib dosing until progressive disease (PD), unacceptable toxicity, or other reason for treatment discontinuation. All treated patients will undergo a safety follow-up visit at 28 days (± 7 days) after the last dose. Patients will be followed for long-term survival.

3.1.2 Phase 2

Phase 2 is an expansion study in 3 selected cohorts of pediatric patients with tumors bearing *NTRK* fusions: IFS, other extracranial solid tumors, and primary CNS tumors. At least 30 patients less than 18 years of age with a primary CNS tumor harboring an *NTRK* gene fusion, including 15 patients with high-grade gliomas (HGG) and another 15 patients with low-grade gliomas (LGG) were planned for enrollment to satisfy the PIP commitment that would show larotrectinib as a safe and effective treatment with significant anti-tumor activity and durable clinical benefit in children with a primary CNS tumor.

The Phase 2 study structure consists of screening, the treatment period, safety follow-up, and long-term follow-up until the study is closed. On-going safety (referred to as the "active safety follow-up" and described in integrated CSP amendment, version 12.0, section 7.6.1), survival, and subsequent anticancer therapies will be assessed during the long-term follow-up.

Larotrectinib is administered as an oral capsule or a liquid formulation at the recommended dose determined in Phase 1 (100 mg/m^2; not to exceed 100 mg BID). Cycles are 28 days of continuous BID dosing. Patients undergo periodic radiologic evaluation and ongoing safety assessments as outlined in the protocol. Individual patients will continue daily larotrectinib dosing until PD, unacceptable toxicity, or other reason for treatment discontinuation, as outlined in integrated CSP amendment, version 12.0, section 6.4.

Patients with locally advanced disease and a response of at least a partial response (PR) for a period of one year after PR is initially confirmed may have their larotrectinib treatment withheld, a "wait and see" period, after discussion with the Sponsor. Patients with LGG and a response of stable disease (SD) for a period of at least 2 years after SD is initially confirmed, may similarly have their larotrectinib treatment withheld. Patients in the "wait and see" period will continue to undergo periodic radiologic evaluation and ongoing safety assessments as outlined in section 7.8 of the protocol. If a patient in this "wait and see" period is noted to have evidence of radiographic disease progression, the patient may restart larotrectinib. At the time of larotrectinib re-initiation, the patient will restart all per protocol assessments as delineated for patients on active treatment. Patients who reinitiate larotrectinib treatment are permitted to have their larotrectinib treatment stopped again.

Patients who undergo surgical resection resulting in negative margins may have their study treatment stopped and continue to undergo periodic radiologic evaluation and ongoing safety assessments as outlined in section 7.7 of the protocol. These patients are permitted to restart study treatment, if they experience PD, after discussion with the Sponsor. At the time of larotrectinib re-initiation, the patient will start all assessments per protocol as delineated for patients on active treatment.

3.2 External Cohorts

The choice of the control cohorts constituting the comparator arm of the study is not arbitrary. The non-arbitrary choice of the data sources to constitute the control cohort has been ensured by a comprehensive review of the existing relevant databases in France and internationally, notably based on a systematic literature review (SLR). After screening for full texts, 11 studies were included for data extraction. All these studies were critically reviewed against pre-defined criteria, validated by a scientific committee (Bayer study team, clinicians, methodologists, Institut Curie), to select cohorts of patients who present characteristics closest to those included in the SCOUT study and with sufficient methodological quality in order to be deemed appropriate for use as an external control arm to the SCOUT trial. All the studies identified within this SLR presented shortcomings on one or more parameters. However, two multicenter studies (Orbach et al. 2016 and Sparber-Sauer et al. 2020) had a lower risk of bias (ROBINS-I tool), especially as patient recruitment was prospective and had advanced IFS patients and minority of metastatic IFS patients. Both the studies reported longitudinal clinical efficacy data.

The Curie Institute database is the only database available in France describing French pediatric patients with IFS or STS harboring a *NTRK* gene fusion. Patient-level data contained in this cohort may provide useful information in the management of these patients. The Curie database will be augmented with data on IFS pediatric patients emanating from CWS. The two data sources will be pooled as a single comparator group for larotrectinib patients from the SCOUT trial, in order to address the request from HAS for further comparative analysis on larotrectinib treatment.

Despite those two studies had methodological limitations, among the 11 studies reviewed they are the most suitable to be considered as valid external control arm.

4. General Statistical Considerations

All issues concerning patient eligibility, data consistency checks, permissible data modifications, and coding of medical terms and concomitant medication will be described in detail in the Data Management Plan.

Sample Size and Power Considerations

The number of patients available is quite small, and no formal hypothesis and power calculations are considered for this study.

We can calculate the power for the comparison of two groups under a Cox Proportional Hazards Model, with the following applicable formula (**Error! Reference source not found.** 2006)[1]:

$$\text{Power} = \Phi\left(\frac{\sqrt{km}|\text{IRR-1}|}{k \times \text{IRR+1}} - z_{1-\alpha/2}\right)$$

where m = expected total number of events over both groups

$= n_1 p_E + n_2 p_C$

n_1, n_2 *= number of participants in experimental (E) and control (C) groups*

$k = n_1/n_2$ *(ratio of participants in E versus C groups)*

p_C *= probability of failure in group C over time period of study*

p_E *= probability of failure in group E over time period of study*

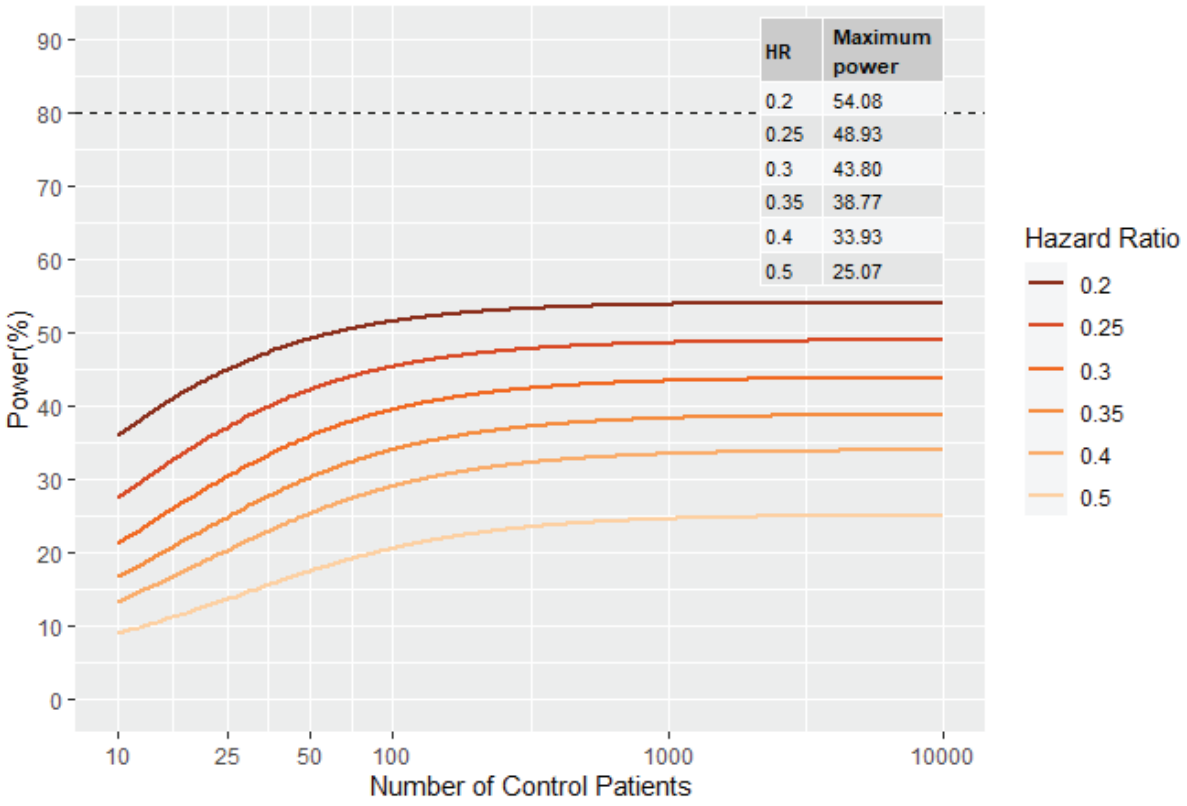
IRR = underlying hazard ratio

$z_{1-\alpha/2}$ *is the 100(1-α/2)-th percentile of the standard normal distribution*

Φ *is the CDF of N(0,1)*

The power calculations provided in the protocol assumed 35 IFS patients and probabilities of failure equal to 0.04 and 0.19 in SCOUT and the external historical control cohorts, respectively [Orbach (2020)]. Even with a strong treatment effect (hazard ratio (HR)=0.25), the power can only reach 49%, as further additions to the control cohort lead to smaller increases in power as the two groups become increasingly unbalanced. Therefore, having a powered study seems unfeasible.

Figure 1: Representation of the study power in function of the number of patients included in the comparator arm (control) for different hazard ratios



nE=35 (number of patients in the experimental group (treated with larotrectinib)
pE=0.04 (probability of failure in the experimental group during the study)
*pC=0.19 (probability of failure in the control group during the study) ([*2] Orbach *2020[2])*
RR=0.5 to 0.2 (risk ratio)
α=0.05 (type 1 error)

Patients from SCOUT study will be taken from 20 July 2021 cut-off, using EPAS6 and ISAS1 subsets. EPAS6 is defined as those patients who were either in the primary analysis set on which approval of larotrectinib was based (N=55), or who satisfied the following conditions:

- *NTRK* Fusion Analysis Set
- Not CNS Primary Tumor
- Response evaluation criteria in solid tumors (RECIST) measurable disease at baseline
- First Dose Date on or prior to 19 January 2021

The ISAS1 subset extends the first dose date window of EPAS6 to on or prior to 20 July 2021.

As of the 20 July 2021 cut-off, there are 49 SCOUT IFS patients available for analysis. The pooled external cohorts from Curie and CWS will comprise N=42. Therefore, the power calculations in the protocol can be updated with current information on sample size as follows:

- N=49 IFS patients for SCOUT
- N=42 for pooled external control cohorts
- pE=0.04 (probability of failure in SCOUT) and pC=0.19 (probability of failure in control) per protocol assumptions
- Optimistic scenario of a strong treatment effect (HR=0.25)
- Power=51%
- Power cannot exceed 62.9% for N=49 for SCOUT and an unlimited number of subjects in the external arm

4.1 General Principles

The statistical evaluation will be performed by using the software package SAS version 9.4 (SAS Institute Inc., Cary, NC, USA) for all outputs assigned for publication, or R (version 4.1.0 [2021-05-18]) or higher for exploratory analysis.

The creation of TLFs will follow the standard as described in RD-SOP-1215.

Descriptive analysis of the data will be performed using summary statistics for categorical and quantitative (continuous) data. Continuous data will be described by the number of non-missing values, median, mean, standard deviation, minimum, and maximum as well as lower and upper quartiles. Frequency tables will be generated for categorical data.

Time-to-event analyses, and other model-based analyses, will be described in detail in subsequent sections.

4.2 Handling of Lost to Follow Up and Premature Discontinuation

Lost to follow-up or withdrawal of consent for study observations will result in the patient being censored for all time-to-event endpoints. Censoring times and status can be accommodated using either Kaplan-Meier methodology or Cox Proportional Hazards models, as detailed in later sections. Censoring details for each endpoint will be given in section 6.

4.3 Handling of Missing Data

4.3.1 Baseline Data

No imputation of missing information will be applied except for partial dates with missing day only. These dates will be imputed to either first day of month, or first dose date if it falls in the same month.

Dates with month or year missing in addition to day will remain missing. This rule affects dates used in the analysis of patient and disease characteristics: dates of initial disease diagnosis and locally advanced/metastatic disease; dates of radiation therapy used to calculate duration of most recent radiotherapy.

4.3.2 Time to Event Data

No imputation will be performed for missing tumor response evaluations or dates.

The last known alive date (LKAD) is defined as the latest date indicating that the patient is still alive, among all available visit dates and other dates pertaining to a patient record (e.g. AEs) in the database. Partial documented dates (i.e. with day only missing) will be imputed using the first day of the month.

4.4 Interim Analyses and Data Monitoring

Interim analyses are not applicable to this retrospective observational externally-controlled study.

4.5 Data Rules

All documented therapies will be coded using the World Health Organization – Drug Dictionary (WHO-DD). Medical history and diagnoses will be coded using the latest Medical Dictionary for Regulatory Activities (MedDRA) version coding system.

Unless otherwise specified, baseline value is defined as last non-missing value on or prior to first dose of larotrectinib for the SCOUT study, or as last non-missing value on or prior to start date of chemotherapy (first line) for the external historical control cohorts.

Calculation of months is derived as: 1 month = 30.4375 days.

4.6 Definition of Derived Variables and Subgroups

A subgroup analysis will be performed on the primary endpoint for patients who take larotrectinib as second-line therapy or are treated with second-line or higher systemic therapy in the external historical cohorts. A similar subgroup analysis will be conducted for comparing first line Larotrectinib patients versus external controls taking at least 1 line of therapy. This is described fully in section 6.2.5.

5. Analysis Sets

All analyses will be performed on the Full Analysis Set (FAS), comprising patients in the SCOUT trial, or the pooled population of Institut Curie and CWS.

5.1 Assignment of Analysis Sets

FAS

A patient will be included in the FAS if the patient has taken at least one dose of larotrectinib for SCOUT study, or has fulfilled the eligibility criteria for inclusion into the pooled external cohorts and has initiated chemotherapy.

All analyses detailed below will be repeated separately for two populations of FAS:

5.1.1 IFS Population

- Age ≤ 21 years old
- Locally advanced or metastatic IFS
- Patients receiving
    - larotrectinib in the SCOUT trial, or
    - receiving at least one chemotherapy-based regimen in the external historical control cohorts.

5.1.2 IFS with documented NTRK gene fusion

.

- Age ≤ 21 years old
- Locally advanced or metastatic IFS, which is NTRK fusion-positive.
- Patients receiving
    - larotrectinib in the SCOUT trial, or
    - receiving at least one chemotherapy-based regimen in the external historical control cohorts.

6. Statistical Methodology

A comparison of patients in the larotrectinib group with patients from the external historical control cohorts who received systemic therapy (chemotherapy-based regimen) will be conducted. The external cohorts will comprise a pooled database from the *Institut Curie* and *CWS*.

Patients who received larotrectinib in the SCOUT trial will likely differ in various underlying characteristics from those who are recorded in the historical control cohorts. These characteristics are listed as potential confounders in section 6.1.2. A SLR was conducted to identify key covariates, as described in section 6.1.3. The analyses will need to be adjusted for these patient characteristics to avoid biased results.

The Inverse Probability of Treatment Weighting [IPTW] method will be employed to reduce the effects of measured confounding variables in the interpretation of the treatment effect (larotrectinib *versus* control group). The propensity score (PS) is defined as the probability of a patient receiving larotrectinib conditional on his observed baseline covariates, and is obtained from a logistic regression model. In order to balance the analysis, a weight is assigned to each patient as the inverse probability of being in a certain group conditioning on these baseline covariates. Therefore, for larotrectinib patients, this is the reciprocal of the PS, while for external comparator patients this is the reciprocal of one minus the PS. If a patient has a higher probability of being in a group, it is considered as over-

represented, and therefore is given a lower weight. On the other hand, if the patient has a smaller probability of being in the group, it is considered as under-represented and is given a higher weight. Therefore, the statistical design of this study intends to use this weighting adjustment to remove sampling bias, though diagnostics to check whether balance has been achieved in baseline covariates between larotrectinib and external cohort groups will be performed.

6.1 Population Characteristics

6.1.1 Disposition of Patients

The number and percent of patients included in the FAS will be summarized for both the larotrectinib and external cohorts.

6.1.2 Demographic and Baseline Characteristics

Descriptive summaries of demographics and baseline characteristics will be presented for each cohort and overall. The following baseline data will be summarized:

- Patients and disease characteristics:
  - Sex,
  - Age,
  - Disease history:
    - Locally Advanced or Metastatic Disease
    - Primary Tumor Location
    - Time from initial diagnosis to first dose of larotrectinib (SCOUT) or first line of chemotherapy (external cohorts),
    - Time from locally advanced/metastatic disease diagnosis to initiation of larotrectinib (SCOUT) or first line of chemotherapy (external cohorts),
    - The resectability (Intergroup Rhabdomyosarcoma Study) IRS classification stage
- Collection of treatment-related information:
  - Prior surgical treatment(s) for cancer:
    - Number of previous surgical resections
    - Amputated leg (yes/no)
    - Amputated arm (yes/no)
    - Other surgical intervention (categorization to be decided after medical review)
  - Prior radiation treatment(s) for cancer:
    - Prior radiotherapy (yes/no)
    - Duration of most recent radiotherapy
    - Anatomical sites
    - Start / stop dates (listing only).
  - Systemic treatment(s):
    - Number of prior systemic regimens, and categorization of treatment regimens
    - Reasons for discontinuation
    - Start / stop dates (listing only)
    - Best response to last systemic treatment (complete response (CR), partial response (PR), SD, PD, NE). Note: NE is defined as not evaluable / not done.
    - Number of cycles

6.1.3 Systematic Literature Review

A SLR was conducted in order to identify and quantify the impact of prognostic factors on the primary endpoint. The report found that there was a lack of evidence in the literature on prognostic factors that might have an influence on clinical outcomes for patients with IFS.

The SLR identified only 5 studies, and due to the rarity of the disease all studies had a small sample size leading to a high degree of imprecision in the results. The report has highlighted the following core subset of prognostic factors for inclusion into the PS model.

6.1.4 Core subset of prognostic variables

IRS initial resectability classification

IRS group was the most studied prognostic factor in the literature in terms of its impact on overall survival (OS). IRS Stage has 4 groups,but most of the studies (Orbach 2010, Cecchetto 2001, Sulkowski 2013) compared the impact of IRS I *versus* IRS II or IRS I *versus* IRS II/III on clinical outcomes. Those studies suggested that IRS I may be an indicator of better prognosis, either in terms of OS or event-free survival (EFS), than IRS II or III. No studies investigated the impact of IRS IV (which corresponds to metastatic disease stage) on clinical outcomes compared to other IRS groups.

The IRS initial resectability classification was not collected in SCOUT study, however given it is reported as a prognosis factor in the literature, in order to allow to feature it in the PS model, IRS for SCOUT IFS patients will be derived from their disease status. Locally advanced disease corresponds as per medical expert opinion (PI. Dr Orbach) to IRS III and metastatic disease to IRS IV (as per IRS definition).

Locally advanced *versus* metastatic disease stage

We have no knowledge of locally advanced *versus* metastatic disease stage being studied on clinical outcomes due to a lack of metastatic patients included in any IFS studies. However, the broad expert consensus is that metastatic disease stage is likely to have a significant effect on endpoints in comparison to locally advanced patients. Given IRS initial resectability classification already captures the stage of the disease, the impact of disease stage will be already accounted for and an additional variable locally advanced vs metastatic disease stage is not needed to feature in the PS model.

Primary Tumor Localization

Two studies explored the impact of primary tumor localization (axis *versus* extremity) on clinical outcomes (Orbach 2010[11], Cecchetto 2001). Only Orbach 2010 explored its impact on EFS, which did not reveal a statistical difference between tumor locations. Both studies found consistent results that tumor localization did not have an impact on OS. However, the Cecchetto study did include both infantile and adult fibrosarcoma patients, and results from this study may not be fully representative of the current study. Therefore, tumor location will not form part of the core prognostic factors for the PS model.

Tumor Size

Two studies explored the impact of tumor size on clinical outcomes (Orbach 2010, Sparber Sauer 2020). Both studies concluded that tumor size (≤5 cm, >5 cm) did not have an impact on EFS. Furthermore, the Orbach [2010] study did not find tumor size to be a prognostic factor of OS. As tumor size factors into IRS group stage (IRSG), tumor size will only contribute indirectly to the propensity score through IRSG.

Age

The SLR did not uncover any evidence of age having a statistically significant prognostic value on EFS and OS, probably due to small sample sizes. However, age will be retained as a prognostic factor due to the clinical importance of the covariate.

Time from locally advanced/metastatic disease diagnosis to initiation of larotrectinib (SCOUT) or first line of chemotherapy (external cohorts)

Based on the SLR, no study investigated how time from diagnosis to initiation of therapy affects clinical outcomes. Time from locally advanced/metastatic disease to initiation of treatment will be added to the core subset, acknowledging that there will likely be an initial imbalance on lines of therapy between treatment groups. Line of therapy will be substituted by elapsed time from locally advanced/metastatic disease in the PS model.

6.2 Analysis of Primary Variable(s)

6.2.1 Primary Endpoint

The primary endpoint of the study is the time to medical treatment failure: defined as the time (months) from the start of treatment to the date of the earliest event from: subsequent systemic treatment, radiation therapy, mutilating surgery or death due to any cause.

The **index** date of the analysis of the primary endpoint will be defined as either the start date of larotrectinib treatment (for patients in the SCOUT study, regardless of line of therapy) or start date of chemotherapy (first line) for the external historical control cohorts. The end date is defined as the earliest of the following endpoints:

- Start date of a post-treatment systemic anti-cancer therapy, if any, or,
- Start date of a post-treatment radiation therapy, if any, or,
- Date of a post-treatment mutilating surgery, if any, or,
- Date of death due to any cause, if applicable.

Patients who do not record an event will be censored at the last known alive date.

6.2.2 Inverse Probability of Treatment Weighting (IPTW)

Patients who received larotrectinib in the SCOUT trial will likely differ in various underlying characteristics from those who are recorded in the external historical control cohorts. These characteristics are listed above as potential confounders (See sections 6.1.2 and 6.1.4). The analyses will need to be adjusted for these patient characteristics to avoid biased results.

An adjustment methodology using a PS will be employed to reduce the effects of measured confounding variables in the interpretation of the treatment effect (larotrectinib *versus* external historical group). A PS can be seen as an overall "balancing score", that is calculated for each patient based on his or her measured underlying characteristics.

The PS will be obtained from a logistic regression model with dependent variable: larotrectinib group (yes/no). Covariates will be taken from the list of potential baseline confounders above. The predicted probabilities of receiving larotrectinib from this logistic regression model will form the PS for each patient. Example SAS code is as follows:

```
proc logistic data=temp;
class cat1 cat2 … / param=ref;
model tx (event = "Laro") = cat1 cat2 var1 var2 …;
output out=temp2 pred=propensity;
run;
```

where all categorical variables appear in the `class` statement in addition to the `model` statement. Predicted probabilities in the code example above appear in `temp2`.

Therefore, the weight assigned to each patient is the inverse probability of being in a certain group conditioning on these baseline covariates. For larotrectinib patients, this is the reciprocal of the PS, while for external comparator patients this is the reciprocal of one minus the PS. These weights are stabilized by multiplying by the marginal probability of treatment that was actually received, which avoids very large weights being assigned to larotrectinib patients with a very low PS or to comparator patients with a PS close to 1.

The selection of covariates for retention in the logistic regression model for the PS will be considered. Brookhart et al [3] showed in simulations that: (1) including a variable in the PS model related only to survival but not to treatment group decreased the variance of the treatment effect without affecting bias; (2) failure to include a true confounder (i.e. related to both treatment group and survival) led to a biased estimator; (3) the inclusion of a variable related only to treatment group increased the variance of the treatment effect without altering bias. Similar results were reported by Andrillon, Pirracchio and Chevret [4], in small samples, who confirmed that matching on a PS model including instrument variables (associated with treatment allocation) increased the bias of the estimated treatment effect compared to matching on a prognostic model or on the true confounder model.

Two PS models will be considered:

1. **SLR model**: The core subset of variables in section 6.1.4, comprising the prognostic factors identified by the SLR (, IRS group, age) and time from locally advanced/metastatic disease diagnosis to initiation of treatment.
2. **Full model**: All covariates in [1] and additional covariates listed in section 6.1.2:
    a. Additional demographic variables: sex;
    b. Additional covariates on disease history: time from initial diagnosis to first dose of treatment tumor location;
    c. Covariates on treatment-related information: prior surgical treatments for cancer, prior radiation treatment for cancer, prior systemic treatments.

The SLR only identified IRS group as a significant prognostic factor, and we have augmented the SLR model above with age (due to its clinical importance) and locally advanced versus metastatic disease stage (expert consensus on its likely significance despite lack of data). The failure of the SLR to identify other prognostic factors may in part be explained by the small sample sizes limiting the

power of those studies referred to in the report. Notwithstanding these data deficiencies, it does seem unlikely that the analysis could be biased by unmeasured confounders. However, this will be explored as detailed in section 6.5 to allay any concerns. Balance metrics presented in the following section will help to discriminate between these two approaches for the PS model.

6.2.3 IPTW Diagnostics

The distribution of the PS will be graphed for both larotrectinib and control groups. Ideally, there will be no extreme PS and good overlap of treatment and control. Sensitivity analyses described below will address any concerns about outliers in the distribution of PS.

For the two PS models detailed above (SLR and full model), standardized differences between larotrectinib and comparator groups will be computed. A graph will show the absolute standardized difference for each of the baseline covariates in section 6.1.2, comparing larotrectinib and historical control, for both the unweighted and weighted samples. A standardized difference of 10% or below is considered a reasonable guideline for demonstrating sufficient balance between the two groups.

For continuous covariates in the PS model, side-by-side boxplots comparing the distribution of the covariate between the two treatment groups will be presented for both unweighted and weighted samples. After weighting, the boxplots for larotrectinib and comparator should appear similar.

A measure of the overlap in the distribution of a covariate for both groups after applying IPTW weights can be calculated using the overlapping coefficient (OVL) for each baseline covariate, per Andrillon, Pirracchio and Chevret [5]. The OVL is defined as the common area under two probability density curves, and is used here as a measure of agreement between the two treatment groups. For each continuous variable, kernel density estimates (KDE) are found for each group on a common set of grid points, and the value of the minimum KDE is calculated. The area under the minimum of the two density curves is computed based on the trapezoidal rule of integration. This can be calculated using PROC KDE in SAS, and referring to the SAS blog on difference of density estimates[10]. When calculating the OVL for the weighted sample, a WEIGHT statement can be added to the KDE procedure that incorporates the IPTW weights. For categorical covariates within the unweighted sample, the OVL is simply the sum of the minimum proportions (for the 2 groups) across all categories. For the weighted OVL, the group proportions for each category are replaced by their weighted equivalent.

Therefore, per Belitser et al[6], a median OVL can be calculated across all baseline covariates (section 6.1.2) to get an overall measure of balance for SLR and full models. We also consider a weighted average, where weights are chosen to emphasize that balance on strong prognostic factors is more important than on factors that are only weakly related to the primary outcome. We use the following weights[6]

$$w_i = 1 + \left|\log\left(\widehat{HR}_{x_i y}\right)\right| - \frac{1}{I}\sum_{k=1}^{I}\left|\log\left(\widehat{HR}_{x_k y}\right)\right|$$

where $\widehat{HR}_{x_i y}$ is the hazard ratio for covariate x_i in relation to time-to-event primary outcome y in a univariate Cox proportional hazards (PH) model. This assumes there are also a total of I covariates considered in section 6.1.2. Therefore, covariate i is upweighted if the strength of association between the factor and outcome y is greater than average, in terms of the coefficient of the Cox PH model. Conversely, covariates whose strength of association with the outcome is below the average will be

downweighted. The median and weighted OVL value for SLR and full PS models, along with other diagnostic outputs detailed above, will be used to choose the appropriate PS model for use with the primary outcome model in section 6.2.4 and other secondary endpoints. We can decide whether the SLR model is sufficient, or whether there is an advantage in expanding the set of baseline covariates in the PS model.

6.2.4 Outcome Models for Primary Endpoint

The primary endpoint, time to medical treatment failure, will be investigated by plotting Kaplan-Meier survival curves for the original unweighted samples. Superimposed on this survival plot (e.g. with dashed lines) will be adjusted Kaplan-Meier curves for the two groups that incorporate the IPTW balancing weights. Therefore, the dashed lines will represent the survival curves after balancing the larotrectinib and external cohort databases. Median and first and third quartile estimates of survival will be reported using Kaplan-Meier method for both original unweighted and weighted samples. 95% confidence intervals will also be reported. In addition, duration of follow-up will be calculated using the Kaplan-Meier method after reversing *censoring* and *event* status.

```
proc lifetest data=temp outsurv=temp2;
    time osdur*oscnsr(1);
    strata trt;
    weight wgt;
RUN;
```

Example code for incoporating the IPTW weights is shown above. Here, the SAS routine `proc lifetest` calculates confidence intervals for quartiles by applying the formula given by Xie and Liu[7] for the variance of the adjusted Kaplan-Meier estimator.

Statistical tables will report the log-rank test, comparing larotrectinib and comparator groups. Per Xie and Liu[7], an analogous log-rank test for the weighted sample will also be computed. This adjusted log-rank test is available within `proc lifetest` by specifying that the `weight` statement should contain IPTW weights, e.g. `weight wgt`.

Additionally, a hazard ratio will be computed using a weighted Cox Proportional Hazards model, regressing survival time on an indicator variable denoting treatment status (larotrectinib/comparator) and incorporating the IPTW weights. To construct the confidence interval of the hazard ratio, the bootstrap-based variance estimator will provide approximately correct estimates of standard errors as it takes into account the fact that the IPTW weights are only estimates from the available data. For each of bootstrap samples (e.g. 200 samples) drawn from the data, weights will be drawn from the PS model and the log-hazard ratio will be calculated from the weighted Cox model. The standard deviation of the estimated log-hazard ratios (i.e., the estimated regression coefficient for the treatment status indicator) across the 200 bootstrap samples will be used as the bootstrap estimate of the standard error of the estimated regression coefficient obtained in the original data set. Ninety-five percent confidence intervals of the hazard ratio can be constructed as:

$$\exp\left[\widehat{\boldsymbol{\beta}} \mp \mathbf{1.96} \times \mathrm{SE}\left(\widehat{\boldsymbol{\beta}}\right)\right]$$

where $\widehat{\boldsymbol{\beta}}$ denotes the estimated treatment effect in the original weighted sample, and $\mathrm{SE}\left(\widehat{\boldsymbol{\beta}}\right)$ denotes the estimated standard error of the treatment effect using bootstrapping.

Hazard ratios will also be reported for the unweighted samples.

6.2.5 Sensitivity Analyses for Primary Endpoint

Sensitivity analyses for IPTW stabilized weights

A plot of the distribution of IPTW stabilized weights by group will be used to identify very large weights, and to assess the degree of overlap of the PS distributions of the treatment and control groups.

PS scores close to 0 (for larotrectinib) or 1 (for historical control) may be problematic for IPTW due to the large weights assigned to these observations. Extreme PS values may also signal a lack of comparability between the treatment groups, as these patients may lack a match in the other group. Therefore, trimming at 5% level will be considered, which amounts to dropping the individuals with the most extreme PS values in both groups. Essentially, trimming aims to exclude patients who fall well outside the region of overlap in the PS distributions of the two treatment groups. The minimum PS threshold is set as the value of the 2.5^{th} percentile observed among larotrectinib patients, and the maximum PS threshold is set as the value of the 97.5^{th} percentile observed among historical control patients. All patients who have a PS score outside the window defined by these 5% thresholds will be excluded from this sensitivity analysis. The objective of this sensitivity analysis is to remove 'outlier' patients.

A second sensitivity analysis will apply weight truncation that reduces any 'large' weight down to a maximum weight. Following Elze *et al* [8], we set the weight threshold to 10.

In summary, in addition to the main analyses, the following sensitivity analyses will also be considered:

a. IPTW (5% trimming)
b. IPTW (truncating large weights)

A consistency in the findings regarding the treatment effect between the main analysis and the sensitivity analyses should instill confidence in the primary results.

Sensitivity analyses – exact matching

The primary endpoint analysis is expected to provide a rather conservative estimate of the treatment effect. The index date for the patients in the external cohort(s) will correspond to first line of treatment, while for SCOUT patients, the index date will correspond to the second line of treatment for the majority of patients.

To balance the two groups, the following sensitivity analyses will be conducted:

Exact Matching sensitivity analysis

- Exact matching will be applied by the Larotrectinib LOT (line of therapy) by 1^{st} line and >= 2^{nd} line. In this matching 2^{nd} line or higher will be lumped together due to smaller number of patients in the control cohort had 3 or more therapies.
- Propensity scores will be generated after identifying matched external control patients to Larotrectinib patients
- The propensity scores will be obtained from a logistic regression model with the dependent variable: Larotrectinib group (yes, no). Key covariates will include: IRS

group, age, and time from locally advanced/metastatic disease diagnosis to initiation of treatment

- The predicted probabilities of receiving Larotrectinib from this logistic regression will form the propensity scores for each patient
- The weight assigned to each patient will be the inverse probability of being in a certain group conditional on the key covariates. For Larotrectinib patients, this is the reciprocal of the propensity score. For the external controls, this is the reciprocal of 1 minus the propensity score.
- These weights will be stabilized by multiplying by the marginal probability of treatment that actually received and will be referred to as the stabilized IPT weights.
- For both weighted and unweighted samples, the primary endpoint will be evaluated using unstratified and stratified log-rank test, where the stratification factor is the LOT (1L, >=2L)
- The hazard ratio (Larotrectinib/control) will be computed using a weighted Cox proportional hazards model, regressing survival time on an indicator variable denoting treatment group (Larotrectinib, external controls), stratification factor (1L, >=2L), and incorporating the stabilized IPTW weights.
- The index date for the Larotrectinib group is the date of first dose of Larotrectinib administration
- The index date for the external controls is the date of initiation of the matched line of therapy (index therapy)

Subgroup treated with second line of treatment in both groups.

Only those patients who are treated with larotrectinib as second-line treatment (for SCOUT) or take a second line of systemic therapy (for external historical cohorts) will be included in the analysis. The index date is defined as the second line of treatment in both the treatment and control groups. This analysis reflects the current reimbursement scope of larotrectinib, i.e. in patients who are refractory or in relapse.

Subgroup treated with first line of treatment

A similar subgroup analysis will be conducted for comparing first line Larotrectinib patients versus external controls taking at least 1 line of therapy.

6.3 Analysis of Secondary and Exploratory Variable(s)

The following time-to-event secondary endpoints, sections 6.3.1 to 6.3.8, will be analyzed using the steps outlined above, and incorporating the IPTW weights calculated for the primary endpoint. The sensitivity analyses detailed in section 6.2.5 will be limited to the primary endpoint.

6.3.1 Time to Subsequent Systemic Treatment

The endpoint, time to subsequent systemic treatment, will be defined as the time from start date of larotrectinib (for SCOUT) or start date of chemotherapy (for historical control cohorts) till the start

date of a post-treatment systemic anti-cancer therapy, if any. Patients who do not take a subsequent systemic treatment will be censored at the last known alive date.

6.3.2 Time to Mutilating Surgery including Limb Amputation

The endpoint, time to mutilating surgery including limb amputation, will be defined as the time from start date of larotrectinib (for SCOUT) or start date of chemotherapy (for historical control cohorts) till the start date of a mutilating surgery (including limb amputation), if any. Patients who do not record a mutilating surgery (including limb amputation) will be censored at the last known alive date.

6.3.3 Time to First Radiation Therapy

The endpoint, time to radiation therapy, will be defined as the time from start date of larotrectinib (for SCOUT) or start date of chemotherapy (for historical control cohorts) till the start date of a radiation therapy, if any. Patients who do not record a radiation therapy will be censored at the last known alive date.

6.3.4 Time to Complete Surgical Resection (excluding Amputation and Any Other Mutilating Surgeries)

The endpoint, time to complete surgical resection (excluding amputation and any other mutilating surgeries), will be defined as the time from start date of larotrectinib (for SCOUT) or start date of chemotherapy (for historical control cohorts) till the start date of a complete surgical resection (excluding amputation), if any. Patients who do not record a complete surgical resection (excluding amputation) will be censored at the last known alive date. Subject recording a R0 resection (no residual tumor) are considered to have achieved complete surgical resection. Additionally for this study, subjects recording a R1 surgery (macroscopic residual tumor) will also be included for complete surgical resection as there is no additional treatment in this case for IFS.

6.3.5 Overall Survival

OS is the time from start date of larotrectinib (for SCOUT) or start date of chemotherapy (for historical control cohorts) till time of death due to any cause. Patients alive at the time of analysis will be censored at the last known alive date.

6.3.6 Progression-Free Survival

PFS is the time from start date of larotrectinib (for SCOUT) or start date of chemotherapy (for historical control cohorts) till earlier of disease progression/relapse assessed by radiological or clinical examination or death due to any cause. Patients without an event at the time of analysis will be censored according to the rules contained within

Rules for Determining Date of Progression or Censoring for Progression-Free Survival

Rule	Situation	Date of Progression or Censoring	Outcome
1	No baseline tumor assessments	Date of First Larotrectinib Treatment or Chemotherapy	Censored
2	No post-baseline assessments and no death	Date of First Larotrectinib Treatment or Chemotherapy	Censored
3	No progression (clinical or radiological) and no death (with a post-baseline tumor assessment)	Date of last available tumor assessment	Censored
4	Patient lost to follow-up (or withdrew consent from study participation) prior to progression (clinical or radiological) or death	Date of last available tumor assessment	Censored
5	Progression (clinical or radiological)	Date of assessment where progression was confirmed	Progressed
6	Death without earlier progression	Date of death	Progressed
7	New anti-cancer therapy prior to progression (as assessed by medical expert)	Date of last tumor assessment before initiation of new therapy	Censored

6.3.7 Overall Response Rate

ORR is defined as the proportion of patients with a best overall response of CR or PR assessed by investigators.

Summary statistics for ORR with 95% 2-sided exact binomial confidence intervals will be reported. In addition, a weighted analysis will be conducted. The binary response outcome (ORR), is regressed on group indicator (larotrectinib, comparator) using a logistic regression model, incorporating the IPTW weights calculated for the primary endpoint. Standard errors are calculated using a robust sandwich estimator. The odds ratio comparing larotrectinib to comparator, with 95% confidence interval, will be provided. A corresponding odds ratio will also be reported for the unweighted samples.

6.3.8 Disease Control Rate

Disease control rate (DCR) is defined as the proportion of patients with a best overall response of CR or PR or SD as assessed by investigators.

Summary statistics for DCR with 95% 2-sided exact binomial confidence intervals will be reported. In addition, a weighted analysis will be conducted. The binary response outcome (DCR), is regressed on group indicator (larotrectinib, comparator) using a logistic regression model, incorporating the IPTW weights calculated for the primary endpoint. Standard errors are calculated using a robust sandwich estimator. The odds ratio comparing larotrectinib to comparator, with 95% confidence interval, will be provided. An odds ratio for the unweighted samples will also be reported.

6.4 Safety Analysis

Safety Analyses will be limited to calculating incidence of treatment discontinuation due to treatment-emergent adverse events (regardless of relatedness to treatment). This endpoint will be summarized descriptively by treatment group, with proportions and 95% 2-sided exact binomial confidence intervals, in the original unweighted sample only.

6.5 Confounder or Bias Adjusted Analyses

A bias analysis will investigate the effect of an unmeasured covariate for the primary endpoint, time to medical treatment failure. Per Austin (2014)[9], pairs of larotrectinib and control patients will be matched on the logit of the PS using a caliper that is defined as a proportion of the standard deviation of the logit of the PS. In this case, we use a standard quoted in Austin[9] of 0.2 × SD of logit PS. The PS model will be defined by the final model chosen for the primary endpoint using IPTW.

We define response as control time-to-event earlier than paired larotrectinib time-to-event, after removing pairs where the earlier time was censored (i.e. indeterminate response). A stratified log-rank test on the matched samples is equivalent to a 2-sided binomial test of the null proportion of responses, $p = 0.5$. If there exists an unmeasured confounding variable that increases the odds of larotrectinib exposure by Γ (e.g. 10%), then we can consider two extreme scenarios where a commensurate decrease or increase in responses occurs after adjusting for this variable in the PS model. Therefore the range of the true significance level can be found by repeating the binomial test using $p = \frac{\Gamma}{\Gamma+1}$ and $p = \frac{1}{\Gamma+1}$. In reality, the effect of the unmeasured covariate will lie somewhere between these 2 extremes.

The stratified log-rank test will be reported assuming there were no covariates unaccounted for in the PS model, i.e. p-values assume a null proportion of responses, $p = 0.5$ for a 2-sided binomial test. P-values will be also be reported for the scenarios where an unmeasured covariate increases odds of exposure by 5% and 10%. For 5%, p-values will be reported for a range of null proportions = [0.4878, 0.5122]. For 10%, p-values will refer to null proportions = [0.4762, 0.5238].

6.6 Analysis of Representativeness

Not applicable.

6.7 Additional Analyses Planned to be Reported Outside the Study Report

Not applicable.

7. Document History and Changes in the Planned Statistical Analysis

The complementary analysis on patients with IFS or other STS with *NTRK* gene fusion will no longer take place as there is only one STS patient in the Curie Database and none in the CWS. In its place, analyses will be performed on the population of patients with IFS having documented NTRK gene fusion.

The protocol described a sensitivity analysis on choice of index date of the primary endpoint, for subjects in either arm who are treated with a second line or more of systemic therapy or larotrectinib. The index date was specified as the start of second line in both groups. However, as shown below in 2, fewer control subjects received a second line than anticipated, and a larger number of subjects in SCOUT took larotrectinib as the secondline therapy. Therefore, the analysis would have had an unbalanced sample size at N=32 versus 8. In addition, 14 of the 32 larotrectinib subjects start at line 3 to 6, leading to a conservative analysis that unduly favors the control group.

In order to provide a more balanced analysis, we propose to replace this sensitivity analysis for the choice of index date with the sensitivity analysis based on exact matching on the line of therapies, approach as detailed in Section 6.2.5 A subgroup analysis was planned in those treated with second line of treatment in both groups. A similar subgroup analysis will be conducted when comparing first line Larotrectinib patients versus external controls taking at least 1 line of therapy, which was added to Section 6.2.5.

Table2 Line of Therapy

	Larotrectinib	**Control**
1st Line	19 (37%)	34 (81%)
2nd Line or Higher	32 (63%)	8 (19%)

8. References

[1] Rosner B. (2006). *Fundamentals of Biostatistics*. (6-th edition). Thomson Brooks/Cole.

[2] Orbach D, Sparber-Sauer M, Laetsch TW, Minard-Colin V, Bielack SS, Casanova M, et al. Spotlight on the treatment of infantile fibrosarcoma in the era of neurotrophic tropomyosin receptor kinase inhibitors: International consensus and remaining controversies. Eur J Cancer. 2020; 137:183-192.

[3] Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. Am J Epidemiol. 2006 Jun 15;163(12):1149-56.

[4] Lawrence Rasouliyan, Estel Plana, Jaume Aguado. Considerations in the Use of Propensity Scores in Observational Studies https://www.rtihs.org/publications/considerations-use-propensity-scores-observational-studies

[5] Anais Andrillon, Romain Pirracchio and Sylvie Chevret. Statistical Methods in Medical Research 2020, Vol. 29(3) 644–658.

[6] Svetlana V. Belitser, Edwin P. Martens, Wiebe R. Pestman, Rolf H.H. Groenwold, Anthonius de Boer and Olaf H. Klungel. Measuring balance and model selection in propensity score methods. Pharmacoepidemiology and drug safety 2011; 20: 1115–1129

[7] Xie J. and Liu C. Adjusted Kaplan–Meier estimator and log-rank test withinverse probability of treatment weighting for survival data. Statist. Med. 2005; 24:3089–3110

[8] Elze CM, Gregson J, Baber U, Williamson E, Sartori S, Mehran R, Nichols M, Stone GW, Pocock SJ. Comparison of Propensity Score Methods and Covariate Adjustment. J Am Coll Cardiol 2017;69:345–57.

[9] Austin PC. The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. *Statist. Med.* ***2014***, 33 1242–1258.

[10] Wicklin R. The difference of density estimates: When does it make sense? SAS BLOGS. https://blogs.sas.com/content/iml/2013/04/03/the-difference-of-density-estimates.html

[11] Orbach D, Rey A, Cecchetto G, Oberlin O, Casanova M, Thebaud E, et al. Infantile fibrosarcoma: management based on the European experience. J Clin Oncol. 2010;28(2):318-23.

9. Appendix

Not applicable.

Statistical Analysis Plan – Signature Page

OS statistical analysis plan approval form (sec. data collection)

Study number: BAY 2757556 / 21767

Statistical Analysis Plan (SAP)

Version and Date: v3/28JUL2022

I have read and approve the SAP Amendment referred above.

	Name	Signature and Date
Approved by:		
PPD	PPD By delegation: PPD By delegation: PPD By delegation: PPD	PPD

Electronic Signatures – Consent and Privacy statement

Herewith, signers agree that the electronic signatures are intended to be the legally binding equivalent of traditional handwritten signatures according to CFR – Code of Federal Regulations Title 21, Part 11 Electronic Records; Electronic Signatures, Subpart C – Electronic Signatures.

Bayer AG (hereinafter "us" or "we") wishes to inform you about **processing of your personal data in the context of Electronic Signatures**. Your personal data will be processed when you use the technical platform DocuSign Digital Transaction Management Platform ("DocuSign Website") and as part of electronically signing or approving a document. DocuSign Inc., United States of America, provides the technical platform. For respective data processing, please refer to the privacy statement from DocuSign Inc. available on the DocuSign Website. The Privacy Statement at hand informs about the personal data processed by Bayer AG as part of electronically signing or approving a document.

The **purpose for which we collect and process your personal data** is the handling of electronic signatures. This includes signing and approving of documents as well as legally required retention and archiving of signed documents. The **legal bases** for this processing are Art. 6(1)(b) and (c) General Data Protection Regulation.

The personal data, which we process, are **identifying information** like your name, e-mail address and contact details, and **document management and signature related information** like date of signature, history of document-related actions (e.g. review, sign and IP address). Furthermore, the electronically signed documents may contain **personal data of any type** that is not directly related to the electronic signature process.

Access to your personal data will be given to our departments, which are involved in processing of electronic signatures. This may involve departments from other Bayer affiliates within and beyond Europe. Your personal data may also be provided to service contractors supporting our electronic signature processes. Such service contractors are carefully selected and regularly monitored by us. They will only process personal data in accordance with our instructions and on basis of appropriate data processing agreements.

Your personal data may be transferred to and processed in a country outside the European Economic Area for which the European Commission has not decided that it ensures an adequate level of data protection. In such cases, we will set up appropriate protection measures as required by GDPR. For example, we generally will apply the so-called "standard

data protection clauses" adopted by the European Commission. You can obtain a copy from us on request. Alternative protection measures may be implemented as well.

Your personal data will be **retained and archived** according to legally required retention and archiving periods applicable to business related files. Your personal data will then be deleted if there are no other legal reasons to further processing.

You have the right to request information from us about your personal data, access to and rectification or erasure of personal data or restriction of processing concerning your personal data or to object to processing as well as the right to data portability. You also have the right to lodge a complaint with the data protection supervisory authority.

For any questions, you may have with respect to data privacy, please contact digitalsignature@bayer.com.

Last updated: November 12, 2020.