



PLEASE READ CAREFULLY:

By opening the attached document you agree to the following terms of use:

- You may not use this document or the information contained herein to a regulatory authority in connection with an application for a marketing authorization or any other regulatory submission without the express written consent of Roche. Please contact global.data_sharing@roche.com
- You may not use this document or the information contained herein to identify clinical trial patients.
- You may not copy, reproduce, or make this document available in any manner that would permit a third-party to review or use the document without first agreeing to these terms of use.



Indirect Treatment Comparisons statistical analysis plan

Indirect treatment comparisons of treatments in SMA
Type 1

Roche compound of interest: RG7916/Risdiplam

Date: 20 – 12 -2019

Author(s): [REDACTED], Visible Analytics

[REDACTED], Visible Analytics

[REDACTED], Visible Analytics

Adapted by:

[REDACTED]

[REDACTED]

Table of contents

1	Introduction	3
2	Data	3
2.1	Clinical studies	3
2.2	Outcomes	8
2.3	Adjustment of prognostic factors	10
3	Statistical methods	11
3.1	MAIC methodology	11
3.2	STC methodology	13
3.2.1	Target population	13
3.3	Statistical models.....	13
3.3.1	Types of endpoints.....	13
3.3.2	Covariates used for the MAIC/STC.....	13
4	List of planned analyses	16
5	References	18

List of tables

Table 1 Inclusion criteria and study design	6
Table 2 Baseline characteristics	7
Table 3 Dosing and delivery	7
Table 4 List of efficacy outcomes amongst the published studies (from SR report)	8
Table 5 Motor Milestones Outcomes for the analysis*	9
Table 6: Motor Function Outcomes for the analysis*	9
Table 7: Survival Endpoints for the analysis*	10
Table 8 Covariates to be tested in MAIC/STC analyses	13
Table 9 Covariates excluded from the matching analysis	14
Table 10 Planned analysis for Risdiplam vs. Nusinersen	16
Table 11 Planned analysis for Risdiplam vs. AVXS-101.....	17

1 Introduction

Risdiplam (RG7916) is a product under investigation for spinal muscular atrophy (SMA). Three trials (JEWELFISH, FIREFISH and SUNFISH) of Risdiplam are currently ongoing amongst SMA Type 1 and Type 2/3 patients.

A systematic review of clinical trials in SMA was performed to identify available evidence for comparators of interest and assess the feasibility of Indirect Treatment Comparison (ITC) / Network meta-analyses (NMA) for SMA Type 1 and SMA Type 2/3 using data from FIREFISH/SUNFISH trials and results of a systematic review. The ITC/NMA will compare Risdiplam to relevant comparators. The output of the analyses will support the interpretation of the results for Risdiplam and inform market access strategy.

This SAP describes the analysis to be undertaken in Type 1 patients using individual patient-level data from FIREFISH.

2 Data

2.1 Clinical studies

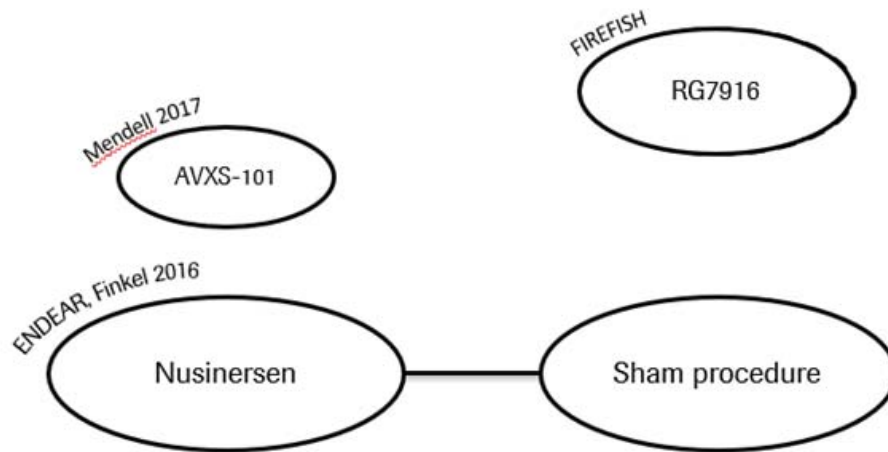
FIREFISH is a two-part seamless, open-label, multi-center study to investigate the safety, tolerability, pharmacokinetics, pharmacodynamics and efficacy of Risdiplam in infants with type 1 spinal muscular atrophy. The first part of the study has enrolled 21 infants to evaluate the safety, tolerability, pharmacokinetics, and pharmacodynamics of Risdiplam and to select the dose for Part 2. Part 2 was designed to assess the efficacy of Risdiplam measured as the proportion of infants sitting without support after 12 months of treatment, as assessed in the Gross Motor Scale of the Bayley Scales of Infant and Toddler development – Third Edition (BSID-III) (defined as sitting without support for 5 seconds). The same efficacy parameters are assessed in Part 1 and Part 2.

The data from the pooled population that received the final dose will be included in the indirect treatment comparison. This includes all part 2 patients as well as the subset of patients who received final dose in the part 1.

A systematic literature review and feasibility assessment of all published and ongoing trials in SMA types 1, 2 and 3 was performed to identify available evidence for comparators of interest (searches run on 31st January 2018).

Based on the results of the literature review, the following disconnected network was formed (Figure 1). The final analyses will be conducted following availability of the FIREFISH part 2 data (January 2020).

Figure 1: SMA Type 1 evidence network using FIREFISH



The network above is disconnected, hence a network meta-analysis based on a comparison of relative treatment effects is not possible. A naïve or unanchored indirect comparison of Risdiplam with Nusinersen (Finkel RS 2017) or AVXS-101 (Mendell JR 2017) may be confounded by systematic variation in prognostic factors. Therefore, a matching adjusted indirect comparison (MAIC) will be attempted. Sample size is a limitation in MAIC analyses, limited numbers of patients may limit the ability to adjust for potential confounding factors due to reductions in the effective sample size. This is expected to be the case for the comparison against AVXS-101 and in this case Simulated Treatment Comparison (STC) will be attempted.

MAIC is an indirect comparison method which adjusts for between-trial population imbalances. Unlike network meta-analysis which are based only on aggregate data, MAIC incorporates individual patient data (IPD) from one trial (e.g. FIREFISH) and reweights outcomes based on published aggregate data from other trials (e.g. ENDEAR and Mendell 2017). The broadly defined steps to perform MAIC are as follows:

1. Assess cross-trial similarities and differences
2. Match average baseline characteristics from ENDEAR and Mendell 2017 separately by applying “propensity score” based weights to individual patients in FIREFISH
3. Compare outcomes across balanced trial populations

Step 1 is discussed below, steps 2 and 3 are discussed in Section 3.1.

STC is also a population average method and as opposed to using weighting methods is an outcome regression method, which estimates the effect of covariates on outcomes and makes use of these to estimate the effect of an intervention in a population with a different profile. STC is discussed in section 3.2.

A challenge of population matching is that these techniques can only be used to compare with one trial at time. Hence, separate matching exercises and analyses will have to be conducted for comparisons to nusinersen and AVXS-101. Matching versus AVXS-101 will be performed using the Mendell 2017 study and matching versus nusinersen will be conducted using the nusinersen arm from the ENDEAR trial. Mendell 2017 was chosen for matching as the only study reporting AVXS-101 whilst ENDEAR was preferred as the largest study reporting Nusinersen. Finkel 2016 reported only 20 patients compared to 80 in ENDEAR.

Published methodology recommends comparing the inclusion criteria, study design, baseline characteristics and study procedures across the trials included in the MAIC (Signorovitch JE 2010, D. Phillippo, et al. 2016, D. M. Phillippo, et al. 2018).

Table 1 compares the inclusion criteria and study design between the included studies. All studies are in infants with SMA type 1, defined by the homozygous deletion or compound heterozygosity deletion or mutation of the SMN1 gene and two copies of the SMN2 gene. The inclusion criteria for age of SMA onset and age of enrollment differ between the studies. FIREFISH and Mendell 2017 are both open label studies with two cohorts of patients, treated with different doses whereas ENDEAR is a randomized, double-blind, sham-controlled study. Study designs cannot be adjusted for in the analysis. The implications of these differing study designs can be highlighted in any write-up of the analysis.

An important difference in study inclusion criteria is that the CL-101 study (Mendell 2017) did include only children who received their treatment before 6 months of age. This is an important difference that needs to be taken into consideration in the indirect treatment comparison against AVXS-101.

Table 1 Inclusion criteria and study design

	Inclusion	Study design
Risdiplam (FIREFISH)	<ul style="list-style-type: none"> • Diagnosis of 5q- SMA <ul style="list-style-type: none"> ○ Genetic confirmation of homozygous deletion or compound heterozygosity predictive of loss on function of the SMN1 gene ○ Clinical history, signs or symptoms attributable to Type 1 SMA (i.e. hypotonia) ○ Two SMN2 gene copies • Gestational age of 37-42 weeks • Onset of SMA after 28 days but prior to 3 months • Aged between 28 days and 210 days at enrollment • Receiving adequate nutrition and hydration support • Measuring to at least the third percentile in body weight 	<ul style="list-style-type: none"> • Open label • Single arm, two cohorts: Part 1) dose finding, Part 2) Dose confirming
Nusinersen (ENDEAR)	<ul style="list-style-type: none"> • Genetic diagnosis of 5q-SMA <ul style="list-style-type: none"> ○ Homozygous gene deletion or compound heterozygote deletion/mutation of SMN1 ○ Two SMN2 gene copies • Gestational age of 37-42 weeks • Younger than 180 days at SMA onset • Younger than 210 days at screening • Receiving adequate nutrition and hydration support • Measuring to at least the third percentile in body weight 	<ul style="list-style-type: none"> • Randomised • Double-blind • Sham-controlled
AVXS-101 (Mendell 2017),	<ul style="list-style-type: none"> • Genetically confirmed diagnosis of SMA 1 <ul style="list-style-type: none"> ○ Homozygous SMN1 exon 7 deletions ○ Two copies of SMN2 • Children up to the age of 6 months • Hypotonia and muscle weakness • Homozygous SMN1 exon 7 deletions 	<ul style="list-style-type: none"> • Open label • Single arm, two doses: High dose and Low dose

Table 2 compares the baseline characteristics between the included studies. FIREFISH and ENDEAR baseline characteristics are relatively similar. There are clear imbalances between FIREFISH and Mendell 2017 which may limit the feasibility of the matching, such as age at initiation of treatment (D. Phillippo, et al. 2016).

Table 2 Baseline characteristics

Baseline characteristic	Risdiplam (FIRE-FISH), (part 1 cohort, n=21)	Risdiplam (FIRE-FISH), (all, n=62)	Nusinersen (ENDEAR), (nusinersen arm, n=80)	AVXS-101 (Mendell 2017), (high dose arm only, n=12)
Female gender	71%	60%	54%	58%
White race	81%	63%	NR	92%
Mean weight (sd, [range])	6.7kg (1, [5.2-8.9])	6.8kg (1.1, [4.1-10.6])	NR	5.7kg (-, [3.6-8.4])
Mean age at first dose (sd, [range])	178 days (42, [102-213])	166 days (44, [68; 213])	163 days (-, [52-242])	~103 days (3.4 mts) (-, [27-240])
Mean age at symptom onset (sd, [range])	8.1 weeks (3.2, [4 – 13.1])	7.5 weeks (3.1, [4-13.1])	7.9 weeks (-, [2-18])	~6.1 weeks (1.4 mts) (-, [0-13])
Duration of symptoms at screening (sd, [range])	104 days (38, [37-163])	93 days (41, [7-163])	~92 days (13.2 wks) (-, [0-181])	NR
Patients with clinical support: Ventilatory	24%	29%	26%	17%
Patients with nutritional support: Gastrointestinal tube feeding	0%	3%	9%	NR
Mean total CHOP INTEND score (sd, [range])	24 (6, [10 – 34])	22 (7, [8-37])	27 (8, [-])	28 (-, [12-50])
HINE-2 score (sd, [range])	1.00 (0.7, [0-3])	0.95 (0.97, [0-5])	1.29 (1.07, [-])	NR
NR=Not Reported				

Table 3 compares the dosing and delivery for the three treatments. All three treatments are delivered differently and at different frequencies: Risdiplam is an oral treatment delivered daily, Nusinersen is delivered through injection into the spinal canal with six treatments over the course of around 10 months, and AVXS-101 is a one-off administration through an intravenous catheter inserted into the peripheral vein.

Table 3 Dosing and delivery

Treatment (Study)	Dosing and delivery
-------------------	---------------------

Risdiplam (FIREFISH)	<ul style="list-style-type: none"> • Delivery: oral • Dose not yet reported • Orally once daily for four weeks followed by an open-label extension phase
Nusinersen (ENDEAR)	<ul style="list-style-type: none"> • Delivery: intrathecal injection • Dose adjusted according to estimated volume of cerebrospinal fluid, equivalent to 12mg-dose if age \leq 24 months • Delivered on days 1, 15, 29, 64 and maintenance doses on days 183 and 302
AVXS-101 (Mendell 2017)	<ul style="list-style-type: none"> • Delivery: intravenous catheter inserted into the peripheral vein • Cohort 1: 6.7×10^{13}vg/kg • Cohort 2: 2.0×10^{14}vg/kg • One time administration: 10-12 ml/kg slowly infused over 60 minutes*
* Protocol specifies 15-20 minutes	

2.2 Outcomes

The efficacy outcomes reported in each of the studies found in the SLR are shown in Table 4.

Table 4 List of efficacy outcomes amongst the published studies (from SR report)

Study identifier	Study details	CHOP-INTEND	HINE-2	Motor milestones (BSID-III)	Motor milestones (WHO)	Survival rate	Event-free survival ^o	Time to death/use of permanent ventilation
ENDEAR	RCT; Nusinersen	Yes*	Yes*	-	-	Yes*	Yes*	Yes*
Finkel 2016	DC/DE; Nusinersen	Yes*	Yes*	-	-	Yes*	Yes*	Yes*
Mendell 2017	DC/DE; AVXS-101	Yes*	-	Yes*	Yes*	Yes*	Yes*	Yes*

^o Event is defined as permanent ventilation

Colour coding details:

Yes*: Outcome assessed; reported in publication

CHOP-INTEND= Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders; DC/DE= Dose comparison/ escalation trial; ; HINE-2= Hammersmith Infant Neurological Examination Module 2; RCT= randomised control trial; SA= Single arm trial; WHO= World Health Organisation; BSID-III= Bayley Scale of Infant and Toddler Development

Table 5 shows possible outcomes to consider in the analysis, and which are reported in each of the studies. For each outcome of interest, the publication must report either mean/median

and standard deviation for continuous covariates, or the proportion of individuals with a binary/categorical trait (D. Phillippo, et al. 2016). Differences in assessment schedule have been noted and for the analysis of the binary endpoint in order to account for differences in the length of the follow-up of studies FIREFISH and ENDEAR (terminated at the interim analysis) a sensitivity analysis will be conducted. For this sensitivity analysis any assessment that occurred in the three months preceding the date of clinical cut-off will be excluded.

Table 5 Motor Milestones Outcomes for the analysis*

Outcome	Type of outcome	Risdiplam FIREFISH*	Nusinersen ENDEAR	AVXS-101 Mendell 2017
Motor milestone –Percentage of infants sitting without support for >=5 seconds, as classified by item 22 of BSID-III	Binary	Yes	No	Yes
Motor milestone –Percentage of infants sitting without support for >=30 seconds, as classified by item 26 of BSID-III	Binary	Yes	No	Yes
Motor milestone- Percentage of infants with head control (classified by BSID-III)	Binary	Yes	No	Yes
Motor milestone- Percentage of infants that can roll over (classified by BSID-III)	Binary	Yes	No	Yes
Full head control (classified by HINE-2)	Binary	Yes	Yes	No
Sitting without support (classified by HINE-2)	Binary	Yes	Yes	No
Sitting with or without support (classified by HINE-2)	Binary	Yes	Yes**	No
Rolling (Classified by HINE-2)	Binary	Yes	Yes	No
Standing (Classified by HINE-2)	Binary	Yes	Yes	No
Motor-milestone response according to HINE-2, defined as meeting the following criteria: improvement in at least one HINE-2 category and more HINE-2 categories with improvement than categories with worsening	Binary	Yes	Yes	No
<p>* At 12 months for Mendell et al, at 12 months for Risdiplam (patients who died within 12 months will be classified as non-responders) and latest available of 6-13 months for Finkel et al</p> <p>** Only available interim efficacy dataset (n=51). Analysis of this outcome assumes that the baseline characteristics for the interim population are the same as for the ITT population (n=80).</p> <p>BSID-III: Bayley Scales of Infant and Toddler Development, version 3</p> <p>HINE-2: Hammersmith Infant Neurological Examination Module</p>				

Table 6: Motor Function Outcomes for the analysis*

Outcome*	Type of outcome	Risdiplam FIREFISH	Nusinersen ENDEAR	AVXS-101 Mendell 2017
Percentage of infants who achieve a CHOP-INTEND score of 40 or higher	Binary	Yes	Yes	Yes

Percentage of infants with \geq 4-point improvement in CHOP-INTEND score from baseline	Binary	Yes	Yes	Yes
*12 months of Follow-up for Risdiplam (patients who died within 12 months will be classified as non-responders), latest available of 6-13 months for Nusinersen; at 8 months for AVXS-101 CHOP-INTEND: Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders				

Table 7: Survival Endpoints for the analysis*

Outcome	Type of outcome	Risdiplam FIREFISH	Nusinersen ENDEAR	AVXS-101 Mendell 2017
Alive without permanent ventilation at 12 months	Binary	Yes	Yes	Yes
Alive at 12 months	Binary	Yes	Yes	Yes

2.3 Adjustment of prognostic factors

To inform the choice of prognostic and predictive factors for inclusion in the matching exercise, we conducted a systematic review of prognostic studies and subgroup analysis conducted in trials identified by the feasibility assessment for Nusinersen and AVXS-101. This complies with TSD 18 recommendations which state that evidence must be presented that there are grounds for considering variables as effect modifiers or prognostic factors (D. Phillipppo, et al. 2016).

EMBASE and Medline were searched using SMA disease terms combined with terms for observational studies in SMA. The systematic review comprises RCTs identified in the feasibility assessment and observational studies in SMA. Results of this review were discussed with the Roche internal medical team and with external medical experts and the following factors were selected for matching to be considered in all analyses (Survival, motor milestones and CHOP intend analyses):

- Age at first dose
- Duration of symptoms at screening/prior to initiation of treatment
- Baseline Motor function (Baseline total CHOP Intend Score)

Review of prognostic factors and expert medical opinion confirmed the age at onset of disease/symptoms to be a very important prognostic factor for survival in SMA type I (Farrar 2013, Ge 2012, Oskoui 2007). The younger the child at age of onset, the worst will be the prognosis. By a later onset the prognosis will be less severe because the muscular system will have had the chance to further develop. Age at onset of disease is not known exactly, but estimated from the recall of duration of symptoms by caregivers. Age at first dose is recorded in clinical trials and known very well and will be used for matching instead of age of symptoms in combination with the duration of the symptoms. Age at first dose was also found to be an important predictor of motor functional outcome for children treated with nusinersen in the review of prognostic factors (Pechmann 2018).

Duration of symptoms was indicated by medical experts to be a very strong predictor of treatment efficacy: the earlier the patients are treated in the course of their disease the better the results. The combination of both age at disease onset and disease duration is important. Symptoms/disease duration relies on the recall of the first symptoms by the caregivers and it is anticipated that caregivers' assessment will be highly variable. However, no better alternative exist, therefore this factor is included in the matching algorithm.

Total CHOP intend score was further identified by medical expert as a useful criterion to classify the severity of the disease, this score has a greater granularity than the HINE-2 total score and is therefore preferable to characterize patients with SMA type I.

Nutritional status and ventilatory function have been identified as important predictors of survival (Gregoret 2013, Oskoui 2007); nevertheless medical experts emphasized that differences in protocols across centers and over the course of the years make this variable not useful for the purpose of matching across the identified studies. The children weight percentile could be a preferred and more stable measure of nutritional status; yet this variable is not available for matching. Similarly, a better indicator of respiratory function would be the frequency of permanent ventilation and the frequency of lower respiratory tract infection; however, none of the studies recruited patients necessitating permanent ventilation, nor included patients with lower respiratory tract infections.

3 Statistical methods

3.1 MAIC methodology

MAIC is a form of propensity score weighting, in which individuals in the FIREFISH trial will be weighted by the inverse of their propensity to be in the FIREFISH trial compared to the comparator trial, to balance the covariate distribution with that of the ENDEAR and Mendell 2017 trials separately. This process will be implemented using published methodology, and is summarized below (Signorovitch JE 2010, D. Phillippo, et al. 2016, D. M. Phillippo, et al. 2018).

Age at first dose has been shown to be an important predictor of treatment efficacy. In the study CL-101 (Mendell 2017) only children below the age of 6 months were considered for inclusion. For this reason any child above the age of 180 days at first dose in the FIREFISH study will be excluded from the matching in the comparison against AVXS-101. As it is anticipated that only few patients included in the FIREFISH received their first dose of Risdiplam prior to the 6 months an alternative approach will be used to indirectly compare to AVXS-101, this method is described in section 3.2.

As the evidence for Risdiplam is from a single arm trial, both MAICs will be unanchored. This means that the weighting model must include every effect modifier and prognostic variable.

The first step is to create a propensity score model which will predict the probability of patients being included in the FIREFISH trial compared to the target trial (either ENDEAR or Mendell 2017). This is equivalent to a model on the log of the individual weights (w):

$$\log(w_i) = \alpha_0 + \alpha_1^T X_i$$

Where X_i is the covariate vector for the i^{th} individual. The regression parameters ($\hat{\alpha}_1$) are estimated using the method of moments, to match effect modifier distributions between trials. This is equivalent to minimizing:

$$\sum_{i=1}^{N_{R(R)}} \exp(\alpha_1^T X_i)$$

when $\bar{X}_{(Nu)}^{EM} = 0$, where R is Risdiplam and Nu is Nusinersen. Matching analysis using AVXS-101 will be performed separately; for this section we will continue using Nusinersen as the example.

Outcomes (Y) will be predicted for Risdiplam in the Nusinersen and AVXS-101 populations separately by reweighting the outcomes of the individuals in the FIREFISH trial according to the weights estimated by the propensity score model above. For example, to predict the outcomes on R in the Nu trial, the outcomes of the R individuals are reweighted:

$$\hat{Y}_{R(Nu)} = \frac{\sum_{i=1}^{N_{R(R)}} Y_{i(R)} \hat{w}_i}{\sum_{i=1}^{N_{R(R)}} \hat{w}_i}$$

Using the natural outcome scale (g), the unanchored indirect comparison between, e.g. N and R will therefore be:

$$\hat{\Delta}_{RN(Nu)} = g(\bar{Y}_{Nu(Nu)}) - g(\hat{Y}_{R(Nu)})$$

The matching will reduce the effective sample size (ESS) for the FIREFISH trial. The degree of reduction will depend on the degree of “overlap” between the FIREFISH and the target populations. Given the limited patient numbers in some of the target trials this may lead to marked uncertainty in estimated treatment comparisons. TSD 18 reported an average reduction in ESS of 80% which could lead to a high level of uncertainty (D. Phillippo, et al. 2016). The ESS will be approximated by:

$$ESS = \frac{(\sum_{t=R} \sum_{i=1}^{N_{t(R)}} \hat{w}_{it})^2}{\sum_{t=R} \sum_{i=1}^{N_{t(R)}} \hat{w}_{it}^2}$$

A small ESS indicates highly variable weights due to lack of population overlap, meaning that the estimate may be unstable. However, TSD 18 cautions that this ESS approximation is likely to underestimate the true ESS as the weights are not fixed, known or uncorrelated with outcome. The distribution of estimated weights will also be reported.

It is important to emphasize that the small numbers of patients in some of the studies will limit our ability to take account of confounders and increase uncertainty. Care will be taken not to ‘overfit’ and reduce sample size. Due to the numbers of patients in FIREFISH we anticipate only being able to match on mean baseline characteristics and not variability. Baseline characteristics to be used in the matching analysis are detailed in Section 3.3.

Bootstrapping will be used to obtain confidence intervals around our estimates. We will bootstrap the whole MAIC process to account for uncertainty in both the sampling error and uncertainty in the weights. We have considered methods for estimating the systematic error in the estimates as recommended in Appendix 3 of the TSD 18, however this is unfeasible given the sample size of the trials.

3.2 STC methodology

Outcome regression methods underlie STC. In this method a model for the conditional mean outcome is regressed on treatment and selected model covariates (D. Phillippo, et al. 2016,).

This method, similarly to the MAIC approach, assumes that absolute outcomes can be predicted from the covariates and hence that all effect modifiers and prognostic factors are accounted for. The outcome model is fitted using the IPD from the Firefish study, and covariates will be centered around the average observed in the comparator's trial (CL-101, Mendell 2017), this way the model's intercept will represent the outcome to be expected in the comparator's study (log hazard for time to event endpoint and log-odds for binary endpoints).

This method will be used in order to estimate the effect of covariates on outcome (Age at first dose and baseline functional score), and to attempt to estimate what might have been the result to be expected the population included in the CL-101 study (Mendell 2017).

As for the MAIC approach described above, bootstrapping will be used in order to obtain 95% confidence intervals around the estimates.

3.2.1 Target population

TSD 18 stipulates that the target population must reflect the real-world population. In the MAIC, the IPD from FIREFISH will be matched to that of the competitor trials, ENDEAR and Mendell 2017, respectively. In the event these populations differ from a real world SMA population, calibration of the results may be necessary.

3.3 Statistical models

3.3.1 Types of endpoints

Table 5 shows that the endpoints available are either binary outcomes or hazard ratios (HR). There is no continuous data available that reports the standard deviation around the mean. MAIC and STCs typically assume additivity on the natural outcome scale; for binary outcomes we will report endpoints on the logit scale and for HR we will report endpoints on the log HR scale.

3.3.2 Covariates used for the MAIC/STC

Considering the baseline characteristics available, the prognostic factors identified in the review and feedback from the internal medical team at Roche and external medical experts we have listed the covariates to be included in each of the matching analysis (Table 8).

Table 8 Covariates to be tested in MAIC/STC analyses

Characteristic	Justification
Mean age at first dose	<ul style="list-style-type: none"> • Age characteristics are most commonly reported as prognostic • Age of symptom onset and age of treatment onset specifically are highlighted by the prognostic review, age of at first dose is the most reliable measure in clinical trials.
Duration of symptoms/disease	<ul style="list-style-type: none"> • Flagged by internal and external medical experts • This factor was also reported to be associated with efficacy of Nuninersen in subgroup analyses of the ENDEAR trial (Finkel 2017)
Mean baseline total CHOP INTEND score	<ul style="list-style-type: none"> • Motor milestones (e.g. head and/or trunk control) flagged in the prognostic review • CHOP INTEND is seen as a stronger indicator than HINE-2 as was developed based on infants with SMA type 1 (Glanzman 2010) and is more granular than HINE-2 (Pechman 2018)

Baseline characteristics that have been excluded from the matching are outlined in Table 9 along with reasons for exclusion.

Table 9 Covariates excluded from the matching analysis

Characteristic	Reason for exclusion from matching
Mean age at symptom onset	Age at first dose already included
Mean age at diagnosis	<p>Age at first dose and duration of symptoms already included</p> <p>Age at diagnosis was not specifically flagged by the review and is subject to local accessibility of</p>

	genetical analyses, hence highly variable across centers
Mean weight	Percentile of growth curve would be a good factor, but this was not reported
Gender	Prognostic review did not found it to be a statistically significant predictor of outcomes
Race	Not flagged by prognostic review or medical experts
Mean HINE-2 score	CHOP INTEND already included as seen as a stronger indicator of motor function compared to HINE-2
Ulnar CMAP amplitude	CHOP INTEND already included as seen as a strong indicator of motor function Prognostic review showed uncertainty about whether CMAP is a prognostic factor
Peroneal CMAP amplitude	Not reported in FIREFISH

4 List of planned analyses

To act as a summary of the SAP, Table 10 and Table 11 show the complete list of planned analyses, for Risdiplam vs. Nusinersen and Risdiplam vs. AVXS-101, respectively.

Table 10 Planned analysis for Risdiplam vs. Nusinersen

Outcomes	Scale of outcome	Characteristics to match
Motor milestone response according to HINE-2	Logit (log OR)	<ul style="list-style-type: none"> • Age at first dose in days • Duration of symptoms in weeks • Total CHOP INTEND score at baseline
Full head control according to HINE-2	Logit (log OR)	
Sitting without support according to HINE-2	Logit (log OR)	
Sitting with or without support according to HINE-2	Logit (log OR)	
Rolling, according to HINE-2	Logit (log OR)	
Standing according to HINE-2	Logit (log OR)	
>=4 point improvement in CHOP-INTEND score from baseline	Logit (log OR)	
CHOP intend of 40 or higher	Logit (log OR)	
Alive	Log HR	
Alive without permanent ventilation	Log HR	
Adverse Events	Logit (log OR)	
Adverse Events leading to discontinuation	Logit (log OR)	
Serious Adverse Events	Logit(log OR)	

Table 11 Planned analysis for Risdiplam vs. AVXS-101

Outcomes	Scale of outcome	Characteristics to match
Sitting without support for ≥ 5 seconds (Item 22 on BSID-III)	Logit (log OR)	<ul style="list-style-type: none"> • Age at first dose in days • Age at onset of symptoms in weeks (duration not available from Mendell) • Total CHOP INTEND score at baseline
Sitting without support for ≥ 30 seconds, as classified by item 26 of BSID-III	Logit (log OR)	
Head control (classified by item 4 of BSID-III)	Logit (log OR)	
Roll-over (classified by item 20 of BSID-III)	Logit (log OR)	
CHOP-INTEND score of 40 or higher	Logit (log OR)	
≥ 4 point improvement in CHOP-INTEND score from baseline	Logit (log OR)	
Alive	Log HR	
Alive without permanent ventilation	Log HR	
Adverse Events	Logit (log OR)	
Serious Adverse Events	Logit (log OR)	

5 References

- Farrar MA, Teoh HL, Carey KA, Cairns A, Forbes R, Herbert K, et al. Nusinersen for SMA: Expanded access programme. *Journal of Neurology, Neurosurgery and Psychiatry*. 2018;89(9):937–942.
- Finkel RS, Mercuri E, Darras BT, et al. 2017. "Nusinersen versus Sham Control in Infantile-Onset Spinal Muscular Atrophy." *The New England Journal of Medicine* 1723-1732.
- Finkel, RS., Kuntz, N., Mercuri, E., et al. 2017. "Primary Efficacy and Safety Results from the Phase 3 ENDEAR Study of Nusinersen in Infants Diagnosed with Spinal Muscular Atrophy (SMA)." *43rd Annual Congress of the British Paediatric Neurology Association, Cambridge, UK*.
- Glanzman, AM., Mazzone, E., Main, M., et al. 2010. "The Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP INTEND): test development and reliability." *Neuromuscul Disord* 155-61.
- Mendell JR, Al-Zaidy S, Shell R, et al. 2017. "Single-Dose Gene-Replacement Therapy for Spinal Muscular Atrophy." *The New England Journal of Medicine* 377 (18): 1713-1722.
- Pechman, A., Langer, T., Schorling, D.,. 2018. "Evaluation of Children with SMA Type 1 Under Treatment with Nusinersen within the Expanded Access Program in Germany." *Journal of Neuromuscular Disease* 135-143.
- Phillippo, D M, A E Ades, S Dias, S Palmer, K Abrams, and N Welton. 2018. "Methods for Population-Adjusted Indirect Comparisons in Health Technology Appraisal." *Medical Decision Making* 38 (2): 200-211.
- Phillippo, DM, AE Ades , S Dias, S Palmer, and K Abrams. 2016. *NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submissions to NICE*. NICE DSU.
- Signorovitch JE, Wu EQ, Yu AP, et al. 2010. "Comparative effectiveness without head-to-head trials a method for matching-adjusted indirect comparison applied to psoriasis treatment with adalimumab or etanercept." *Pharmacoeconomics* 28: 935-45.



Indirect Treatment Comparison - statistical analysis plan

Network meta-analysis of treatments in SMA type 2/3

Roche compound of interest: RG7916/Risdiplam

Version:	1.0
Date:	9-Oct-2019
Authors:	[REDACTED], Visible Analytics Limited
	[REDACTED], Visible Analytics Limited
	[REDACTED], Visible Analytics Limited
Adapted by:	[REDACTED]
	[REDACTED]

Table of contents

1	Introduction.....	3
2	Data	4
2.1	Clinical studies.....	4
2.2	Systematic review to inform the ITC/NMA.....	4
2.3	Systematic review to identify prognostic and predictive factors	11
3	Statistical methods	11
3.1	MAIC methodology (base case)	11
3.2	Restricted NMA (Sensitivity 1).....	13
3.2.1	Priors	13
3.3	Statistical models.....	14
3.3.1	Types of endpoints and sampling models (overview)	14
3.3.2	Specific models	15
3.4	Consistency assessment	15
3.5	Reporting	15
4	List of planned analyses	16
5	References	17
6	Appendix	19
6.1	NMA models: GLM framework	19

List of tables

Table 1 Patient baseline characteristics amongst the studies included in the feasibility assessment.....	6
Table 2 Reporting of common endpoints and timepoints across studies in the network.....	10
Table 3 Sampling models (overview)	14
Table 4 Planned analyses	16
Table 5 Alternative likelihoods and treatment effect scales	20

1 Introduction

Roche currently have an asset (Risdiplam, RG7916) under investigation for spinal muscular atrophy (SMA). Four trials (JEWELFISH, FIREFISH, SUNFISH and RAINBOWFISH) of Risdiplam are currently ongoing amongst SMA type 1 and type 2/3 patients.

FIREFISH is single arm clinical trial to evaluate Risdiplam in type 1 SMA, SUNFISH is a randomized controlled clinical study to evaluate Risdiplam in type 2/3 patients, JEWELFISH is a single arm study to evaluate Risdiplam in pre-treated patients and RAINBOWFISH is a single arm trial to evaluate Risdiplam in pre-symptomatic SMA patients.

Roche are now seeking to conduct a network meta-analysis (NMA) for SMA type 2/3 using data from the SUNFISH trial and results of a systematic review. The NMA will compare Risdiplam to relevant comparators. The output of the NMA will support the interpretation of the results for Risdiplam and inform market access strategy.

This SAP describes the analysis to be undertaken in type 2/3 patients using data from SUNFISH, a RCT to compare Risdiplam vs placebo. (A separate SAP describes the analysis to be undertaken in type 1 patients using data from FIREFISH.)

2 Data

2.1 Clinical studies

SUNFISH is a two-part seamless, multi-center randomized, placebo-controlled, double-blind study to investigate the safety, tolerability, pharmacokinetics, pharmacodynamics and efficacy of risdiplam in type 2 and 3 spinal muscular atrophy patients.

The SUNFISH clinical trial was divided in two parts: SUNFISH part 1 was a dose finding study, that included an active and a control arm; patients in the control arm were switched to the active comparator after 12 weeks. For this reason, part 1 data will not be included in the analyses, which will focus on SUNFISH part 2 data.

In SUNFISH part 2, patients randomized to the active arm with a body weight above 20 kg received 5mg Risdiplam daily, whereas patients with a lesser weight received 0.25/kg daily. Patients on the control arm received placebo. All patients received Best Supportive Care in addition to Risdiplam/placebo.

2.2 Systematic review to inform the ITC/NMA

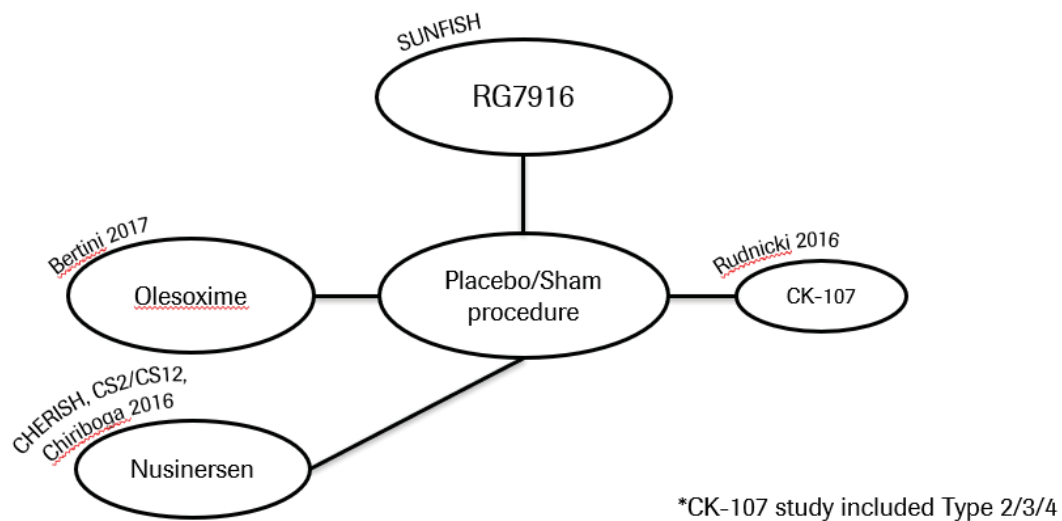
An update of a previously conducted Systematic Literature Review will be performed in 2020. New evidence will be included in a feasibility assessment. This SAP will then be updated accordingly.

A systematic literature review and feasibility assessment of all published and ongoing trials in SMA types 1, 2 and 3 for nusinersen and other key interventions was previously conducted on behalf of Roche. The final deliverables for the project were shared in June 2018. This analysis plan is based on the results of that review for SMA type 2/3.

Based on the results of the literature review and availability of SUNFISH interim analysis, the following connected network was formed including Risdiplam, Olesoxime, CK-107, and Nusinersen (Figure 1). The connectedness of the network, specifically to nusinersen, requires that placebo and sham procedures be treated as equivalent in terms of expected impact on outcomes.

The feasibility assessment concluded an NMA was feasible but that patient characteristics were sparsely reported and that heterogeneity between studies included age (Figure 2). A literature review to identify prognostic and predictive factors was recommended (see Section 2.3). SMN copy number, age and baseline ambulatory status/motor function were indicated to be prognostic/predictive factor of outcome in SMA type 2/3 therefore they will be taken into consideration in the analyses (Bertini 2017, Mercuri 2016, Montes 2018 Petit 2011, Qu 2015, Swoboda 2005).

Figure 1: All evidence network in SMA type 2/3 including Risdiplam



Note that comparisons for selected endpoints may not include all comparators (see Table 2).

SUNFISH included two parts: part 1 was a dose-finding study, part 2 was a RCT with the final dose.

Of the nusinersen studies, only CHERISH is placebo controlled. Chiriboga (2016) was a phase Ib study that investigated four dose levels of nusinersen (1, 3, 6, 9 mg) in sequential groups. CS2 was a non-randomized phase 1b/2a open-label study and included nine patients on a 12 mg dose on days 1, 29, 85. CS12 was their open-label extension and included 47 patients (all doses) who were re-dosed on days 1, 169, 351 and 533.

Rudnicki 2016 was a phase I study in healthy volunteers. Following the results of this study a phase II study in SMA patients type 2/3 and 4 aged 12 and above was initiated to assess the effect of 8 weeks of dosing of CK-2127107 on measures of muscle function in both ambulatory and non-ambulatory patients with SMA as compared to placebo (ClinicalTrials.gov Identifier: NCT02644668).

Bertini 2017 was a study of olesoxime that read out negatively.

Patient characteristics of the studies included in the feasibility assessment are reported in **Table 1**.

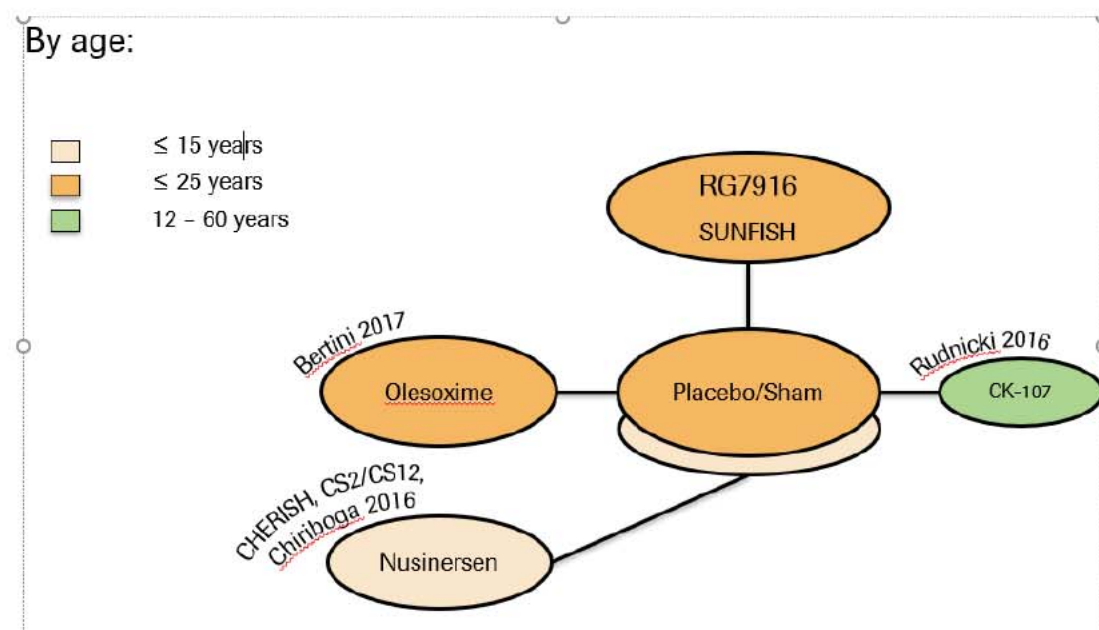
Table 1 Patient baseline characteristics amongst the studies included in the feasibility assessment

Study identifier	Details	Sample size	Age at trial inclusion		Disease duration [range]	Gender (%)		Ethnicity		SMN2 copies		Ambulatory status	RULM score	WHO motor milestone achieved	HFMS score
			Mean	(SD)[range]		Male	Female								
CHERISH (Mercuri 2015)	RCT; Nusinersen	126	Nusinersen: 4 years [2-9] Control: 3 [2-7] Eligible: 2-12 years		Nusinersen: 39.3 months [range 8- 94] Control: 30.2 [10-80]	NR	NR	NR		2/3/4 copies Nusinersen: 2 copies 75, 3 copies 88%, 4 copies 2%, unknown 2%. Control: 2 copies 10%, 3 copies 88%, 4 copies 2%	NR	Nusinersen 19.4 (6.2) Control 18.4 (5.7)	Nusinersen 1.4 (1.0) Control 1.5 (1.0)	Nusinersen 22.4 (8.3) Control 19.9 (7.2)	
Chiriboga 2016	DC/DE; Nusinersen	28	6.1 years [2-14]		NR	39%	61%	3/4/5 copies: 3 copies 89%, 4 copies 75, 5 copies 4%		36%	NR	NR	NR	NR	
CS2 and CS12 (Darras 2016)	DC/DE; Nusinersen	28	7.1 years (4.7) Eligible: 2-15 years		NR	54%	46%	2/3/4 copies: 2 copies 4%; 3 copies 75%, 4 copies 21%		46%	NR	NR	NR	NR	
SUNFISH part 2	RCT; RG7916	180	10 (10.01) [2-25]		NR	49%	51%	White: 71% Asian: 20% Black or African American: 1% Multiple: 1% Unknown: 7%		1 copy: 1% 2 copies: 10% 3 copies: 55% 4 copies: 6% unknown: 28%	Ambulatory: 1% Non-ambulatory: 99%	19.89 (7.04) [3-38]	NR	16.27 (12.32) [0-48]	
Bertini 2017	RCT; Olesoxime	165	9.9 years (5.7) [3-27] 34% <6 years 66% ≥6 years		NR	53%	47%	NR		NR	NR	NR	NR	NR	
Rudnicki 2016	RCT; CK-107	NR	≥12 years		NR	NR	NR	NR		NR	NR	NR	NR	NR	

Rudnicki 2016 included type 2/3/4 patients; DC/DE= Dose comparison/escalation trial; NR= not reported; RCT= Randomised controlled trial.

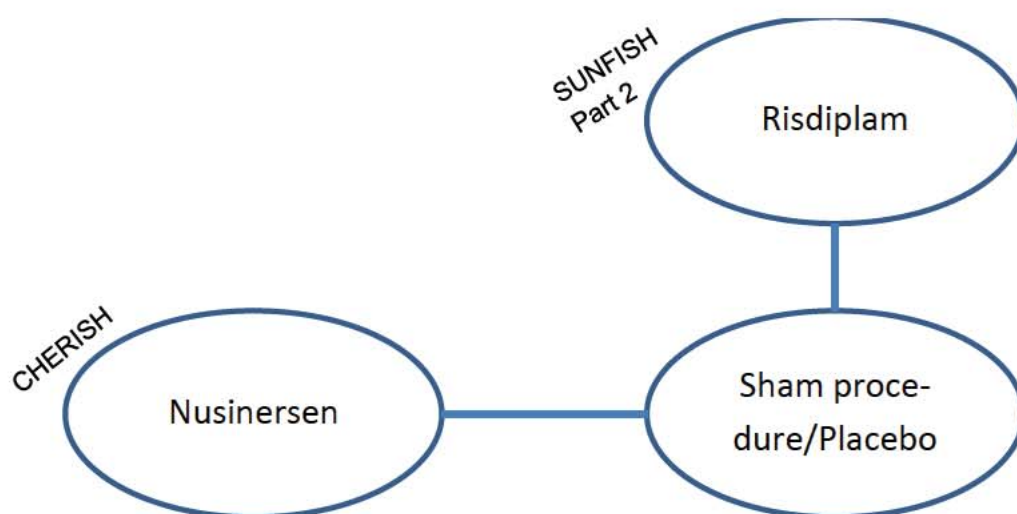
Age was identified as one of the factors predictive of outcome and in order to meet the similarity assumption required for an ITC/NMA, differences across study populations (Figure 2) should be adjusted for. Heterogeneity can be adjusted for in multiple ways: through matching-adjustment, by conducting subgroup analysis, or meta-regression on differences in patient characteristics between studies.

Figure 2 SMA 2/3 network by age



When the focus is on SMA type 2/3 patients only and on studies with licensed treatments and final doses, the evidence network results in the following (Figure 3):

Figure 3: Evidence network including Risdiplam and licensed treatments in SMA type 2/3



The connected network exhibits some heterogeneity in terms of factors that were found / are considered to be predictive of outcome, such as SMN2 copy number, age, ambulatory status and motor function at baseline.

Based on the network depicted in Figure 3, which includes only studies in patients type 2/3 and excludes dose finding studies/cohorts, SUNFISH Part 2 data will be compared to CHERISH data. Due to large differences in factors that may affect treatment efficacy (age, SMN2 copy number, motor function) between SUNFISH and CHERISH studies a standard NMA cannot be performed. Instead, we will be comparing risdiplam to nusinersen using the anchored Matching-Adjusted Indirect comparison (MAIC) methodology. This methodology allows pairwise indirect comparisons while taking into account the heterogeneity of the network. MAIC will be used as Base case. Through the MAIC process the SUNFISH population is matched to the CHERISH population. As a result of this matching, estimates of relative efficacy and safety are valid only for the population that was included in the CHERISH study. This methodology cannot provide estimates of relative efficacy in population subgroups that were not evaluated in the CHERISH study (eg adults). Details on the choice of matching factors are provided in section 2.3

In addition to the MAIC analyses, a restricted NMA will also be conducted (Sensitivity 1), while this analysis does not allow to match all of the factors that are deemed to be predictive of outcome (eg SMN2 copy number) it will allow to control some by excluding from the analyses patients who were not seen in the CHERISH study.

The criteria for selection of the SUNFISH analyses sets for the restricted NMA (Sensitivity 1) are:

- Age < 10 yrs (the oldest patient enrolled in the CHERISH study was 9 years old)
- No severe contracture or severe scoliosis at screening (CHERISH exclusion criteria)
- Motor function Score (Hammersmith Functional Motor Scale – Expanded) ≥10 at screening (CHERISH inclusion criteria)

For the restricted NMA, a Bayesian approach will use an informative prior for the random effects standard deviation and thus compare the fit of fixed and random effects ITC.

Table 2 shows the different efficacy outcomes reported across studies/timepoints that were. Outcomes are reported in binary (proportion of patients responding) or continuous (change from baseline) format.

The following efficacy outcomes will be included in the indirect treatment comparison against Nusinersen/CHERISH:

- Hammersmith Functional Motor Scale Expanded (HF MSE), mean change from baseline at 12 months

- Hammersmith Functional Motor Scale Expanded (HFMSE), proportion of patients showing an improvement of at least 3 points at 12 months¹
- Revised Upper Limb Module (RULM), mean change from baseline at 12 months

The following safety outcomes will be included in the ITC analysis to compare against Nusinersen/CHERISH:

- Adverse events occurring within 15 months from first dose for patients who have received at least one dose of treatment
- Severe adverse events occurring within 15 months from first dose for patients who have received at least one dose of treatment

¹ A change in the HFMSE score of at least 3 points is considered to be clinically meaningful (Swoboda 2010)

Table 2 Reporting of common endpoints and timepoints across studies in the network.

Outcomes of interest	Presentation	Nusinersen			Olesoxime	CK-107	Risdiplam
		CHERISH Mercuri 2015	Chiriboga 2016	CS2/CS12 Durras 2016			
Hammersmith Functional Motor Scale Expanded (HFMSE)	Continuous endpoint, change from baseline	✓ at 15 months	✓ at 9-14 months	✓ at 8.5 months	✓ at 21 months	8 weeks	12 months
Hammersmith Functional Motor Scale Expanded (HFMSE)	Proportion of patients showing maintenance or improvement	✓ at 15 months	✓ At 9-14 months	✓ at 8.5 months	✓ at 21 months	8 weeks	12 months
Revised Upper Limb Module (RULM)	Change from baseline in the Revised Upper Limb Module (RULM)	✓ at 15 months	-	-	-	8 weeks	12 months

2.3 Systematic review to identify prognostic and predictive factors

Following the recommendations of the feasibility assessment, a review was undertaken to identify prognostic and predictive factors. EMBASE and Medline were searched using SMA disease terms combined with terms for RCTs and observational studies in SMA. The systematic review comprises RCTs and 32 observational studies in SMA. A report is currently being prepared.

The review suggested that age could be predictive of treatment efficacy (Bertini 2017) and suggested that age, SMN2 copy number, and motor function (specifically ambulatory status) are prognostic factors. The role of gender as a prognostic factor was unclear. It is unclear whether factors which are prognostic of outcome could also be predictive of treatment efficacy.

Also, there is no clear evidence that baseline functional score is predictive of treatment efficacy once age has been taken into consideration. Previous RCTs in type 2/3 SMA have adjusted for both age and baseline functional score (Mercuri 2018).

In the lack of conclusive evidence on the role of those factors in predicting treatment efficacy the following criteria will be considered for the matching:

- Age
- SMN2 copy
- Baseline HFMSE Functional Score
- Severe scoliosis/contractures at baseline (as they impact on HFMSE assessment)

3 Statistical methods

3.1 MAIC methodology (base case)

MAIC is a form of propensity score weighting, in which individuals in the SUNFISH trial will be weighted by the inverse of their propensity to be in the SUNFISH trial compared to the comparator trial CHERISH, to balance the covariate distribution with that of the CHERISH trial. This process will be implemented using published methodology, and is summarized below (Signorovitch 2010, Filippo 2016, Filippo 2018).

As the evidence for Risdiplam is from an RCT and based on the assumption that Sham in CHERISH is comparable with Placebo in SUNFISH, an anchored MAIC will be performed. This means that the weighting model must include every effect modifier.

The first step is to create a propensity score model which will predict the probability of patients being included in the SUNFISH trial arms compared to the target trial (CHERISH). This is equivalent to a model on the log of the individual weights (w_{it}):

$$\log(w_{it}) = \alpha_0 + \alpha_1^T X_{it}$$

Where X_{it} is the covariate vector for the i^{th} individual on arm t . The regression parameters $\hat{\alpha}_1^T$ are estimated using the method of moments, to match effect modifier distributions between trials. This is equivalent to minimizing:

$$\sum_{i=1}^{N_t(SUNFISH)} \exp(\alpha_1^T X_{it})$$

when $\bar{X}_{(CH)} = 0$, where t is the treatment arm in SUNFISH and CH is CHERISH.

Outcomes (Y) will be predicted for Risdiplam (R) and Placebo (P) in the Nusinersen populations by reweighting the outcomes of the individuals in the SUNFISH trial according to the weights estimated by the propensity score model above. For example, to predict the outcomes on *Risdiplam or placebo* in the *CHERISH* trial, the outcomes of the *Risdiplam (R) and Placebo (P)* individuals are reweighted:

$$\hat{Y}_{t(CH)} = \frac{\sum_{i=1}^{N_t(SUNFISH)} Y_{it(SUNFISH)} \hat{w}_{it}}{\sum_{i=1}^{N_R(SUNFISH)} \hat{w}_{it}}$$

Where t is either Risdiplam or Placebo

Using the natural outcome scale (g), the anchored indirect comparison between, Risdiplam (R) and Nusinersen (N) will therefore be:

$$\hat{\Delta}_{RvsN(CH)} = (g(\bar{Y}_{N(CH)}) - g(\bar{Y}_{S(CH)})) - (g(\hat{Y}_{R(CH)}) - g(\hat{Y}_{P(CH)}))$$

Where S stands for sham, P for placebo and CH for CHERISH

The matching will reduce the effective sample size (ESS) for the SUNFISH trial. The degree of reduction will depend on the degree of “overlap” between the SUNFISH and the target populations. Given the limited patient numbers in some of the target trials this may lead to marked uncertainty in estimated treatment comparisons. TSD 18 reported an average reduction in ESS of 80% which could lead to a high level of uncertainty given the limited sample size of the Part 1 population in SUNFISH (Philippo et al 2016). The ESS will be approximated by:

$$ESS = \frac{(\sum_{t=R,P} \sum_{i=1}^{N_t(SUNFISH)} \hat{w}_{it})^2}{\sum_{t=R,P} \sum_{i=1}^{N_t(SUNFISH)} \hat{w}_{it}^2}$$

A small ESS indicates highly variable weights due to lack of population overlap, meaning that the estimate may be unstable. However, TSD 18 cautions that this ESS approximation is likely to underestimate the true ESS as the weights are not fixed, known or uncorrelated with outcome. The distribution of estimated weights will also be reported.

It is important to emphasize that the small numbers of patients in some of the studies will limit our ability to take account of confounders and increase uncertainty. Care will be taken

not to 'overfit' and reduce sample size. Due to the numbers of patients in SUNFISH we anticipate only being able to match on mean baseline characteristics and not variability. Baseline characteristics to be used in the matching analysis are detailed in section 2.2.

Bootstrapping will be used to obtain confidence intervals around our estimates. We will bootstrap the whole MAIC process to account for uncertainty in both the sampling error and uncertainty in the weights.

Target population

TSD 18 stipulates that the target population must reflect the real-world population. In the MAIC, the IPD from SUNFISH will be matched to that of the competitor trials, CHERISH. However, CHERISH investigated a very selective population and inferences generated from this anchored MAIC approach are valid only in CHERISH-like populations.

3.2 Restricted NMA (Sensitivity 1)

To attempt to adjust for the differences in age groups, baseline motor function (HFMSE score, scoliosis and contractures) between CHERISH and SUNFISH, a restricted NMA will be conducted using the subset of SUNFISH patients aged 9 and under, with a minimum HFMSE score of 10 and without severe scoliosis or contractures at baseline.

The restricted ITC described in section 2.2 will be performed using Bayesian Markov chain Monte Carlo (MCMC) techniques as implemented in JAGS using published methodology (Hawkins et al, 2016). Both fixed and random effects models will be conducted. Given the paucity of data and the star-shaped network, random effects models will be conducted using published informative priors (Turner et al, 2015).

Fixed and random effects models will be compared using the Deviance Information Criterion (DIC). The model with the smallest DIC is indicative of best fit.

Models will be coded in JAGS and run via R using R2jags. Convergence will be assessed using Brooks Gelman Rubin (BGR) plots and by examining trace plots (Brooks & Gelman, 1998). Three chains will be run starting from different initial values. The same priors and three sets of initial values will be used across models.

3.2.1 Priors

The following vague prior distributions are specified for study and treatment-specific terms, respectively:

$$\mu_s \sim \text{dnorm}(0, 10000)$$

$$d_k \sim \text{dnorm}(0, 10000)$$

where $s=1...S$ represent studies and $k=1...K$ treatments respectively.

The vague priors on the studies ensures that the estimates of treatment effect from the network meta-analysis are only influenced by the within trial treatment effect estimates and that between trial differences in response are conditioned out.

In addition, the following informative prior will be used for the random effects model (Source: Turner Table IV - General physical health indicators).

$$\tau^2 = \text{Lognormal}(-2.53, 1.58^2)$$

where τ^2 is the between study variance.

3.3 Statistical models

3.3.1 Types of endpoints and sampling models (overview)

Table 3 lists the efficacy outcomes that will be considered for the ITC.

We will select the most appropriate data type for each endpoint (i.e. binary, continuous, or categorical), depending upon how data are reported, and use an appropriate scale for each endpoint.

Table 3 Sampling models (overview)

Endpoint	Likelihood	Comments
Hammersmith Functional Motor Scale Expanded (HFMSE)	N (θ , se^2) with a linear link	Continuous endpoint, change from baseline
Hammersmith Functional Motor Scale Expanded (HFMSE)	Binomial with a logistic link	Proportion of patients showing maintenance or improvement, odds ratio
RULM	N (θ , se^2) with a linear link	Continuous endpoint, change from baseline
Safety endpoints (AEs, SAEs)	Binomial with a logistic link	Proportion of patients who experienced AEs/SAEs, odds ratio

3.3.2 Specific models

Not applicable.

3.4 Consistency assessment

The network is star-shaped, there were no loops hence an evaluation of network internal consistency is not required.

3.5 Reporting

Results will be presented in PowerPoint and will include tabulated results, evidence networks, Forest plots, and ranking plots. A report will also be produced.

4 List of planned analyses

Table 4 Planned analyses

ID	Network	Approach	Effect size	Timepoint	Comments
Hammersmith Functional Motor Scale Expanded (HFMSE), change from baseline	Base case	MAIC	Mean Difference	12 m	Least-squares mean change from baseline adjusting for each patient's age and their score at baseline
Hammersmith Functional Motor Scale Expanded (HFMSE), proportion of patients showing improvement	Base case	MAIC	OR	15 m for CHERISH 12 m for SUNFISH	Adjusting for patient's age and score at baseline
Hammersmith Functional Motor Scale Expanded (HFMSE), change from baseline	Sensitivity 1	Restricted NMA	Mean Difference	12 m	≤ 9 years age, HFMSE ≥=10, no severe scoliosis or contraindications subgroup using IPD from SUNFISH
Hammersmith Functional Motor Scale Expanded (HFMSE), proportion of patients showing improvement	Sensitivity 1	Restricted NMA	OR	15 m for CHERISH 12 m for SUNFISH	≤ 9 years age, HFMSE ≥=10, no severe scoliosis or contraindications subgroup using IPD from SUNFISH
RULM, change from baseline	Base Case	MAIC	Mean Difference	12 m	Least-squares mean change from baseline adjusting for each patient's age and their score at baseline ≤ 9 years age, HFMSE ≥=10, no severe scoliosis or contraindications subgroup using IPD from SUNFISH
RULM, change from baseline	Sensitivity 1	Restricted NMA	Mean Difference	12 m	Least-squares mean change from baseline adjusting for each patient's age and their score at baseline ≤ 9 years age, HFMSE ≥=10, no severe scoliosis or contraindications subgroup using IPD from SUNFISH
AEs/SAEs	Base Case	MAIC	OR	up to 15 m	
AEs/SAEs	Sensitivity 1	Restricted NMA	OR	up to 15 m	

OR=Odds Ratio

5 References

- Bertini, E., E. Dessaud, E. Mercuri, et al (2017). Safety and efficacy of olesoxime in patients with type 2 or non-ambulatory type 3 spinal muscular atrophy: a randomised, double-blind, placebo-controlled phase 2 trial. *The Lancet Neurology* 16(7): 513-522
- Brooks S, Gelman A. (1998) General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*.7 (4). p434-55
- Chiriboga, C. A., K. J. Swoboda, B. T. Darras, S. T. Iannaccone, J. Montes, V. D. C. De, D. A. Norris, C. F. Bennett and K. M. Bishop (2016). "Results from a phase 1 study of nusinersen (ISIS-SMN Rx) in children with spinal muscular atrophy." *Neurology* 86(10): 890-897
- Darras BT, Chiriboga CA, Montes J, et al. Nusinersen in treatment-naïve patients with later-onset spinal muscular atrophy (SMA): efficacy results from a phase 1b/2a multicenter study (CS2) and its open-label extension (CS12). Presented at the 21st International Congress of the World Muscle Society, 4-8 October, 2016, Granada, Spain.
- Hawkins N, Scott DA, Woods B. An alternative "arm based" parameterisation for net-work meta-analysis. *Research Synthesis Methods* 2016; 7:306-13
- Mercuri E, Finkel R, Montes J, Mazzone ES, Sormani MP, Main M, et al. Patterns of disease progression in type 2 and 3 SMA: Implications for clinical trials. *Neuromuscular Disorders*. 2016;26(2):126–131.
- Mercuri, E., B. T. Darras, C. A. Chiriboga, J. W. Day, C. Campbell, A. M. Connolly, S. T. Iannaccone, J. Kirschner, N. L. Kuntz and K. Saito (2018). "Nusinersen versus Sham Control in Later-Onset Spinal Muscular Atrophy." *New England Journal of Medicine* 378(7): 625-635.[CHERISH]
- Montes J, McDermott MP, Mirek E, Mazzone ES, Main M, Glanzman AM, et al. Ambulatory function in spinal muscular atrophy: Age-related patterns of progression. *PLoS ONE*. 2018;13(6).
- Petit F, Cuisset JM, Rouaix-Emery N, Cancés C, Sablonnière B, Bieth E, et al. Insights into genotype-phenotype correlations in spinal muscular atrophy: A retrospective study of 103 patients. *Muscle and Nerve*. 2011;43(1):26–30.
- Phillippo, D M, A E Ades, S Dias, S Palmer, K Abrams, and N Welton. 2018. "Methods for Population-Adjusted Indirect Comparisons in Health Technology Appraisal." *Medical Decision Making* 38 (2): 200-211.
- Phillippo, DM, AE Ades , S Dias, S Palmer, and K Abrams. 2016. *NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submissions to NICE*. NICE DSU.

Qu YJ, Ge XS, Bai JL, Wang LW, Cao YY, Lu YY, et al. Association of copy numbers of survival motor neuron gene 2 and neuronal apoptosis inhibitory protein gene with the natural history in a Chinese spinal muscular atrophy cohort. *Journal of Child Neurology*. 2015;30(4):429–436.

Rudnicki, S., J. Andrews, F. Malik, A. Wolff and J. Day (2016). "CK-2127107 a selective activator of the fast skeletal muscle troponin complex, for the potential treatment of spinal muscular atrophy." *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration* 17: 241

Signorovitch JE, Wu EQ, Yu AP, et al. 2010. "Comparative effectiveness without head-to-head trials a method for matching-adjusted indirect comparison applied to psoriasis treatment with adalimumab or etanercept." *Pharmacoeconomics* 28: 935-45.

Swoboda KJ, Prior TW, Scott CB, McNaught TP, Wride MC, Reyna SP, et al. Natural history of denervation in SMA: Relation to age, SMN2 copy number, and function. *Annals of Neurology*. 2005;57(5):704–712.

Swoboda KJ, Scott CB, Crawford TO, et al. SMA CARNI-VAL trial part I: doubleblind, randomized, placebo-controlled trial of L-carnitine and valproic acid in spinal muscular atrophy. *PLoS One* 2010; 5(8): e12140.

Turner RM, Jackson D, Wei Y, Thompson SG, Higgins JPT. Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. *Statistics in Medicine* 2015; 34:984-98

6 Appendix

6.1 NMA models: GLM framework

Treatment effects, reported as predicted mean change from baseline $\hat{y}_{s,k}$ for a treatment arm k from study s will be synthesized using a Bayesian multilevel framework following Hawkins et al, 2016.

Treatments included in the model will be indexed as positive integers k with b being the lowest index treatment in a study. For study arms receiving the baseline treatment (i.e. $k = b$),

$$\hat{y}_{s,k} = \alpha$$

where α is a study specific baseline term. The study level baseline is included to ensure that the treatment effect estimates are only informed by the within trial differences between arms. For study arms receiving other treatments (i.e. $k \neq b$),

$$\hat{y}_{s,k} = \alpha + (\beta_k - \beta_b)$$

Where $\beta_k - \beta_b$ is the difference between the treatment effect coefficients. A vague prior for the study specific baseline

$$\alpha_s \sim N(0, 10000)$$

is used to ensure estimates of treatment effect difference are informed by within trial differences between treatment arms and not by absolute response between trials. A vague prior is also used for the treatment effect coefficients

$$\beta_k \sim N(0, 10000)$$

The coefficient for the reference treatment is set to zero, i.e. $\beta_1 = 0$, so that the effect estimates for active treatments are relative to the reference. The observed trial data $\bar{y}_{s,k}$ are included in the model using a normal likelihood function with standard error se for the mean change from baseline:

$$\bar{y}_{s,k} \sim N(\hat{y}_{s,k}, se^2)$$

The corresponding random effects model is:

$$\hat{y}_{s,k} = \alpha + \beta_k + \varepsilon_{s,k} = \alpha + (\beta_k - \beta_b) + \varepsilon_{s,k}$$

where $\varepsilon_{s,k}$ is the random effect deviation for arm k of study s and is assumed to be normally distributed with zero mean and variance $\frac{\sigma^2}{2}$ where σ^2 is the random effect variance:

$$\varepsilon_{s,k} = N\left(0, \frac{\sigma^2}{2}\right)$$

As σ^2 is assumed to be constant across contrasts, each individual treatment response is associated with a random effect variance of $\frac{\sigma^2}{2}$. The study arm-specific random effects $\varepsilon_{s,k}$ are assumed to be uncorrelated within and across studies. This reflects the variation in response observed in individual treatment arms. The correlation in the random variation in treatment effects across multi-arm trials arises as the treatment effect estimates within a trial are jointly dependent on the variation in response in the base treatment arm of the study.

This model produces the correct co-variance for the common variance random effects model. For example, for a single study s comparing treatments 1, 2 and 3, the random effect variance for the treatment effect contrast comparing treatment 2 to treatment 1 is

$$\text{var}(\varepsilon_{s,2} - \varepsilon_{s,1}) = \frac{\sigma^2}{2} + \frac{\sigma^2}{2} = \sigma^2$$

The covariance of the random effects for the contrasts comparing treatment 2 to treatment 1 and comparing treatment 3 to treatment 1 is

$$\text{cov}(\varepsilon_{s,2} - \varepsilon_{s,1}, \varepsilon_{s,3} - \varepsilon_{s,1}) = \text{cov}(\varepsilon_{s,1}, \varepsilon_{s,1}) = \frac{\sigma^2}{2}$$

The random effects variance provides a measure of the variation in observed treatment effects (beyond the variation attributable to random variation in treatment responses within studies) from the predictions of the consistency model. The random effect variance captures both heterogeneity in treatment effects within pair-wise comparisons and inconsistency between direct and indirect estimates of treatment effects.

Alternative likelihoods are described in Table 5.

Table 5 Alternative likelihoods and treatment effect scales

Likelihood	Scale	Link function
Binomial $r_i \sim \text{binomial}(p_i, n_i)$	Arcsine	$p_i = \sin(\eta_i)^2$
	Hazard ratio event	$p_i = \exp(-\exp(\eta_i))$
	Hazard ratio no event	$p_i = 1 - \exp(-\exp(\eta_i))$
	Probit	$p_i = \Phi(\eta_i)$
	Relative risk event	$p_i = \exp(\eta_i)$
	Relative risk no event	$p_i = 1 - \exp(\eta_i)$
	Risk difference	$p_i = \eta_i$
	Odds ratio	$p_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}$
Poisson	Rate ratio	$\lambda_i = \exp(\eta_i)$

$r_i \sim \text{poisson}(\lambda_i E_i)$		
Normal $y_i \sim N(\bar{y}_i, se_i^2)$	Absolute difference	$y_i = \eta_i$
	Ratio of means	$y_i = \exp(\eta_i)$

For a binary variable r_i , n_i , and p_i are the number of events, number of patients and probability of an event for arm i . For a count variable r_i , λ_i , and E_i are the number of events, rate at which events occur and exposure time across all individuals for arm i . For a continuous outcome \bar{y}_i, se_i^2 and y_i are the mean outcome observed, observed variance and predicted outcome for arm i .