IQWiG Reports – Commission No. A21-107

Routine practice data collection for onasemnogene abeparvovec according to the Law for More Safety in the Supply of Medicines (GSAV): review of study protocol and statistical analysis plan

Addendum to Commission A20-61[1]

Addendum

Commission: A21-107
Version: 1.0
Status: 14 September 2021

[1] Translation of the addendum A21-107 *Anwendungsbegleitende Datenerhebung zu Onasemnogen-Abeparvovec: Prüfung des Studienprotokolls und des statistischen Analyseplans – Addendum zum Auftrag A20-61* (Version 1.0; Status: 14 September 2021). Please note: This translation is provided as a service by IQWiG to English-language readers. However, solely the German original text is absolutely authoritative and legally binding.

Publishing details

Publisher

Institute for Quality and Efficiency in Health Care

Topic

Routine practice data collection for onasemnogene abeparvovec according to the Law for More Safety in the Supply of Medicines (GSAV): review of study protocol and statistical analysis plan – Addendum to Commission A20-61

Commissioning agency

Federal Joint Committee

Commission awarded on

16 August 2021

Internal Commission No.

A21-107

Address of publisher

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
Im Mediapark 8
50670 Köln
Germany

Phone: +49 221 35685-0
Fax: +49 221 35685-1
E-mail: berichte@iqwig.de
Internet: www.iqwig.de

Statistical advice

- Til Stürmer, University of North Carolina, Chapel Hill, USA

IQWiG thanks the statistical advisor for his contribution to the present assessment. The advisor was not involved in the preparation of the assessment. IQWiG is solely responsible for the content of the assessment.

IQWiG employees involved in the addendum

- Beate Wieseler
- Ralf Bender
- Thomas Kaiser
- Guido Skipka
- Siw Waffenschmidt

Keywords

Onasemnogene Abeparvovec, Muscular Atrophy – Spinal, Registries, Benefit Assessment, Research Design, Peer Review – Research

Table of contents

List of tables

Page

List of abbreviations

Abbreviation	Meaning
AE	adverse event
AMNOG	Arzneimittelneuordnungsgesetz (Act on the Reform of the Market for Medicinal Products)
CHOP-INTEND	Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders
G-BA	Gemeinsamer Bundesausschuss (Federal Joint Committee)
GLMM	Generalized linear mixed model
GSAV	Gesetz für mehr Sicherheit in der Arzneimittelversorgung (Law for More Safety in the Supply of Medicine)
HFMSE	Hammersmith Functional Motor Scale Expanded
HINE	Hammersmith Infant Neurological Examination
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (Institute for Quality and Efficiency in Health Care)
MedDRA	Medical Dictionary for Regulatory Affairs
PICO	patient, intervention, comparator, outcome
RPDC	routine practice data collection
RULM	Revised Upper Limb Module
SAE	serious adverse event
SAP	statistical analysis plan
SGB	Sozialgesetzbuch (Social Code Book)
SMA	spinal muscular atrophy
SMArtCARE	SMA registry
SMN	survival motor neuron

1 Background

Preliminary note

The Federal Joint Committee (G-BA[2]) is the main decision-maker in the German statutory health insurance system. The Institute for Quality and Efficiency in Health Care (IQWiG[3]), founded in 2004, is the German health technology assessment agency. IQWiG's tasks are specified in Social Code Book (SGB[4]) V, which regulates the statutory health care services. Among other things, IQWiG is commissioned by the G-BA to assess drug and non-drug interventions. IQWiG also assesses new drugs at market entry following the introduction of early benefit assessments in 2011 according to the Act on the Reform of the Market for Medicinal Products (AMNOG[5]). As specified in §35a SGB V[6], "This includes, in particular, the assessment of the added benefit versus the appropriate comparator therapy,[7] the extent of the added benefit and its therapeutic relevance. The benefit assessment is conducted on the basis of evidence provided by the pharmaceutical company, including all clinical trials conducted or commissioned by the company...." IQWiG's tasks were further expanded in 2020 with the Law for More Safety in the Supply of Medicines (GSAV[8]). This law includes several measures to improve drug safety. Among other things, it stipulates that, for certain drugs, routine practice data can be used in early benefit assessments. As specified in §35a (3b) SGB V, the G-BA can require the collection of these data from the pharmaceutical company.

Commission

On 16 August 2021, the G-BA commissioned IQWiG to review the study protocol and statistical analysis plan (SAP) for a routine practice data collection according to the GSAV (RPDC-GSAV) on the gene therapy drug onasemnogene abeparvovec.

In its meeting on 4 February 2021, the G-BA decided to require an RPDC-GSAV and analyses according to §35a (3b) Sentence 1 SGB V for onasemnogene abeparvovec in the treatment of spinal muscular atrophy (SMA) [1,2]. The decision is based, among other things, on the concept for an RPDC-GSAV for onasemnogene abeparvovec developed by IQWiG (rapid report A20-61 of 1 October 2020 [3]).

In order to review whether the requirements of the G-BA for the RPDC-GSAV and for analyses have been implemented by the documents on the study protocol and SAP prepared by the pharmaceutical company, the G-BA forwarded these documents to IQWiG [4,5] and

[2] *Gemeinsamer Bundesausschuss* English website: g-ba.de/english/

[3] *Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen* English website: www.iqwig.de/en/

[4] *Sozialgesetzbuch* German website SGB V: sozialgesetzbuch-sgb.de/sgbv/1.html

[5] *Arzneimittelneuordnungsgesetz* An overview of how the AMNOG procedure is implemented at IQWiG is available in English on iqwig.de/en/presse/media-centre/figures-and-graphs/what-are-dossier-assessments/

[6] German website: sozialgesetzbuch-sgb.de/sgbv/35a.html (quotation translated by IQWiG)

[7] Appropriate comparator therapy = standard care specified by the G-BA

[8] *Gesetz für mehr Sicherheit in der Arzneimittelversorgung* An overview of the law is available in German on the website of the Federal Ministry of Health: bundesgesundheitsministerium.de/gsav.html

commissioned IQWiG to review these documents. In addition to the G-BA's decision on onasemnogene abeparvovec, the contents of the related consultations of the company on the study design of the RPDC-GSAV (2021-B-190 [6], 2021-B-122 [7]) are to be considered.

2 Review of the documents for the planning of the RPDC-GSAV for onasemnogene abeparvovec

2.1 General comments on the documents submitted by the company

2.1.1 Major deviations from the G-BA's decision

The planning of the company for the RPDC-GSAV deviates in essential points from the underlying decision of the G-BA [1]. The company changes the research question by defining the relevant patient populations differently from the G-BA. In particular, it does not divide the patients according to symptom status (presymptomatic vs. SMA type 1 vs. SMA type 2, see Section 2.2.1). In addition, it does not consider the intended hypothesis shift (shifted hypothesis boundaries, see Section 2.3.4), which accounts for the increased uncertainty of the non-randomized study design planned for the assessment. It also limits the use of existing data on treatment courses with nusinersen and onasemnogene abeparvovec in a manner that deviates from the G-BA (see Sections 2.1.2 and 2.2.3). These deviations are described in detail in the following sections.

2.1.2 Limited use of available data on treatment courses with nusinersen and onasemnogene abeparvovec

In the present case of an RPDC-GSAV in a rare disease, the collection of data from a sufficient number of patients presents a challenge. The G-BA considers this by providing for data collected concurrently as well as data not collected concurrently within one data source for the RPDC-GSAV if these meet the defined data quality requirements. Furthermore, the G-BA explicitly also provides for the inclusion of data from other (international) registries in addition to data from the primarily relevant disease-specific SMA registry (SMArtCARE) if these meet the requirements of the RPDC-GSAV. The aim of these specifications is to include as large a sample size as possible in the RPDC-GSAV within a reasonable period and thus enable implementation of the aim of the RPDC-GSAV, namely, the benefit assessment according to §35a SGB V.

The company limits the inclusion of patients in the RPDC-GSAV by a number of decisions:

- Only the SMArtCARE registry is intended for data collection. In contrast, the G-BA's decision potentially provides for the pooling of comparative data from different data sources (registries). The supporting rationale of the G-BA explicitly explains how data from different registries can be pooled [2].
- Only German centres are included from the SMArtCARE registry. According to the study protocol, this step excludes from the RPDC-GSAV 12 foreign hospitals (mostly Austrian) that report data in SMArtCARE. The company justifies this decision with the upcoming newborn screening in Germany and the quality requirements of the G-BA for the use of onasemnogene abeparvovec, which only apply in German centres.

- Of the German centres in SMArtCARE, only those that meet the G-BA's quality requirements for the use of onasemnogene abeparvovec are included. Thus, according to the study protocol, 16 of 34 hospitals in Germany that collect data in the SMArtCARE registry are excluded from data collection. In total, only treatment courses from 18 German hospitals will be used for the RPDC-GSAV (for restriction of centres for data collection, see also Section 2.2.3).
- The company only wants to optionally use data collected retrospectively on treatment courses with nusinersen. The company does not want at all to consider retrospectively collected data on treatment courses with onasemnogene abeparvovec. The company justifies the exclusion of retrospective data on onasemnogene abeparvovec with the G-BA's decision. It is unclear which part of the decision it refers to, as such a restriction is not provided for in the decision (see also Section 2.2.2).

Through these decisions, the company massively limits the sample size for the RPDC-GSAV. As a result, the generation of robust data for a benefit assessment of onasemnogene abeparvovec versus nusinersen will probably be severely delayed and may not be available in the extent required.

2.1.3 Timing of the preparation of the study protocol and the SAP

In a decision dated 4 February 2021, the G-BA requested the company to submit by 15 August 2021 final drafts of the study protocol and the SAP for the RPDC-GSAV on onasemnogene abeparvovec. The company submitted a consultation request on the protocol and SAP to the G-BA on 23 April 2021 and 15 June 2021, respectively. The corresponding consultation meetings took place on 29 June 2021 and 11 August 2021.

The drafts of the study protocol and SAP submitted by the company have a version date of 5 August 2021. The company thus explicitly does not take into account the feedback from the G-BA on its second consultation request. As a result, the study protocol and SAP submitted contain plans that do not coincide with the G-BA's decision (e.g. on the restriction of the centres to be included in the data collection or the definition of the patient populations). The G-BA had already informed the company about this issue.

2.2 Comments on the study protocol

2.2.1 Research question according to PICO

The G-BA's research question for the RPDC-GSAV and the subsequent analysis of the data is specified in the decision using the patient, intervention, comparator, and outcome (PICO) scheme. The following sections assess the implementation of the PICO scheme in the study protocol of the company.

Population

In its decision on the RPDC-GSAV, the G-BA stipulated that the company should collect and analyse comparative data on treatment with onasemnogene abeparvovec or nusinersen for 3 patient populations in the therapeutic indication:

- presymptomatic patients with 5q SMA with a bi-allelic mutation in the survival motor neuron (SMN)1 gene and up to 3 copies of the SMN2 gene
- symptomatic patients with 5q SMA with a bi-allelic mutation in the SMN1 gene and clinically diagnosed SMA type 1
- symptomatic patients with 5q SMA with a bi-allelic mutation in the SMN1 gene and clinically diagnosed SMA type II with up to 3 copies of the SMN2 gene

In this context, patients in the above-mentioned patient population who are older than 6 months or 6 weeks at the time of gene therapy with onasemnogene abeparvovec should also be included in the data collection.

The company deviates from this stipulation of the G-BA. In doing so, it makes inconsistent statements in various sections of the study protocol regarding the patient populations planned for the RPDC-GSAV and the analysis for assessment of the added benefit of onasemnogene abeparvovec versus nusinersen.

In the protocol, the company provides for a population of patients with a bi-allelic mutation in the SMN1 gene and up to 3 copies of the SMN2 gene in the sections on study design and study objectives, without considering symptom status. In addition, the company plans a patient population exclusively with SMA type I. The company does not justify this deviation from the G-BA's decision in these sections of the study protocol.

In contrast, in the inclusion criteria of the study the company depicts the G-BA's definitions of populations and describes how these inclusion criteria can be determined from the data set of the SMArtCARE registry. According to this information, the defined inclusion of the populations specified by the G-BA is possible.

In the data analysis section, the company describes 2 analysis populations, namely a population with a bi-allelic mutation in the SMN1 gene and up to 2 copies of the SMN2 gene and a second population with a bi-allelic mutation in the SMN1 gene and 3 copies of the SMN2 gene. In deviation from the G-BA's specification, the company again does not consider the symptom status. The company justifies this deviation by stating that the introduction of newborn screening from October 2021 would increase the relevance of the number of SMN2 copies compared to the clinical phenotype and that, due to immediate treatment after diagnosis, patients who were symptomatic in screening would play a subordinate role.

The following table provides an overview of the company's definitions of the populations in the study protocol (and SAP).

Table 1: Overview of definitions of patient populations in the study protocol of the company

Section on study design **Section on study objectives**	**Inclusion criteria**	**Section on analysis populations**
Treatment-naive patients with ▪ 5q SMA with a bi-allelic mutation in the SMN1 gene and up to 3 copies of the SMN2 gene and ▪ symptomatic patients with 5q SMA type I treated with onasemnogene abeparvovec or nusinersen	▪ Presymptomatic patients with 5q SMA with a bi-allelic mutation in the SMN1 gene and up to 3 copies of the SMN2 gene or ▪ Symptomatic patients with 5q SMA with a bi-allelic mutation in the SMN1 gene and clinically diagnosed SMA type 1 or ▪ Symptomatic patients with 5q SMA with a bi-allelic mutation in the SMN 1 gene and clinically diagnosed SMA type II and up to 3 copies of the SMN2 gene	▪ Population A: patients with 5q SMA with a bi-allelic mutation in the SMN 1 gene and up to 2 copies of the SMN2 gene ▪ Population B: patients with 5q SMA with a bi-allelic mutation in the SMN1 gene and 3 copies of the SMN2 gene

Overall, the description of the populations in the study protocol is inconsistent. Although the planned inclusion of patients is based on the requirements of the G-BA, the definition of the research question and the planning of the analysis deviate.

The definition of the analysis populations (and thus the populations for which conclusions on added benefit can be made in an assessment) is based on the assumption that no symptomatic patients will be treated in the short term after the introduction of newborn screening. This assumption is speculative. Furthermore, this planning does not take into account that a relevant number of treatment courses are also available for retrospective data collection. These have even been conducted concurrently with nusinersen since onasemnogene abeparvovec entered the market.

Overall, the company's deviation from the G-BA's definition of populations is inappropriate. Symptom status in combination with age contributes to the clinical diagnosis and has a relevant influence on the treatment outcome. The company does plan to consider symptom status at treatment initiation (symptomatic / presymptomatic) as a confounder and in a subgroup analysis. This seems inappropriate due to the relevance of this characteristic and the requirements of the G-BA. A possible effect modification by symptom status cannot be investigated by considering this factor as a confounder. Subgroup analyses, on the other hand, may be suitable for this purpose. However, the company plans subgroups analyses (for time-to-event outcomes) only if there is a statistically significant interaction between treatment and subgroup factor. However, with the expected small sample sizes, the power for a statistically significant interaction will be very low. According to the requirements of the G-BA, the definition of the populations and the data analysis should be performed separately for presymptomatic and symptomatic patients.

Intervention and control

In the data collection, the company includes patients treated with onasemnogene abeparvovec or nusinersen according to approval status. This approach is appropriate.

Outcomes

The company considers the outcomes defined by the G-BA as follows:

Mortality and respiratory function

In addition to mortality, the company plans to analyse a combined outcome of deaths and permanent ventilation. This combined outcome is appropriate in the present therapeutic indication and the operationalization is also appropriate. In addition, comparative effects for respiratory function alone are to be described. This is in accordance with the requirements of the G-BA. Most of the operationalizations planned are appropriate.

For the outcome "improvement in time of ventilator support from baseline", it remains unclear how the improvements to be measured are to be depicted in relation to the ventilatory status at baseline (the company assumes to include primarily presymptomatic patients) or in relation to potential initial deterioration.

Achievement of motor milestones and motor function

The company plans to record attainment of head control (measured with HINE[9]) and a selection of World Health Organization (WHO) motor milestones (sitting without support, crawling, standing without support, and walking without support). In addition, it plans to record motor function with a variety of instruments (HFMSE[10], RULM[11], CHOP-INTEND[12], HINE). This planning ensures a comprehensive description of the patients' motor development.

The following two tables summarize the outcomes planned for motor development. Regarding the motor milestone, it remains unclear why one of the milestones (crawling) is only recorded at the age when healthy children reach this milestone. Regarding the proportion of patients who can sit without support, the timing of the data collection at 18 months seems late. The recording of motor function seems appropriate in that it examines the achievement of the milestones as well as the maintenance of motor function.

[9] Hammersmith Infant Neurological Examination

[10] Hammersmith Functional Motor Scale Expanded

[11] Revised Upper Limb Module

[12] Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders

Table 2: Outcomes planned on motor milestones

Milestone	Proportion of patients reaching the milestone at the time stated				Time to first reaching of milestone	Time from first reaching up to loss of milestone
	Month 8	Month 18	Month 24	Age healthy children		
Head control	X		X			
Sitting without support		X	X	X (9.2 months)	X	X
Crawling				X (13.5 months)		
Standing without support			X	X (16.9 months)	X	X
Walking without support			X	X (17.6 months)	X	X

Table 3: Outcomes planned on instruments for recording motor function

Instrument	Time of data collection
HFMSE	Score at age of 36 months
RULM	Score at age of 36 months
CHOP-INTEND	Change in score from time of first treatment to month 6 and 12 after first treatment
HINE	Change in score from time of first treatment to month 12 and 24 after first treatment

The multiplicity of outcomes describing motor function is problematic for an assessment of added benefit. This multiplicity should be reduced by selecting the relevant outcomes and hierarchizing the outcomes overall. These decisions need to be prespecified in the study protocol.

The company does some hierarchization by defining time to event analyses of the motor function milestones as exploratory. The downgrading of these analyses is not appropriate as these operationalizations cover the complete observation period according to the G-BA decision (60 months), while the other outcomes suggested by the company are limited to a period of up to 24 or 36 months. Furthermore, for the time-to-event outcomes, all patients are included in the analysis, whereas for the outcomes on the proportions of patients reaching the milestone at a certain time point, only those who have reached this age are relevant.

Bulbar function (swallowing and speech ability, need for non-oral nutritional support)

The company depicts bulbar function by difficulties in swallowing and chewing and by the recording of non-oral nutritional support. Of the outcomes defined by the G-BA, speech ability is missing. Regarding the analysis, for the same reasons as described for motor function, time-

to-event outcomes appear to be more meaningful than analyses at fixed time points provided by the company.

Further complications of the disease

The company only plans to collect and analyse orthopaedic complications of the disease (scoliosis and orthopaedic surgery) as complications of the disease. It does not justify why further complications (e.g. pain) are not considered, at least for older patients.

Adverse effects (serious adverse events [SAEs], adverse events [AEs] leading to hospitalizations, specific SAEs [hepatotoxicity, thrombocytopenia, cardiac events, spinal ganglion cell inflammation, renal toxicity, hydrocephalus]).

The company's planning for the recording of AEs deviates from the specifications of the G-BA. The company only includes the outcome "AEs leading to hospitalizations" and does not consider SAEs and specific AEs.

The company justifies its decision not to collect data on specific AEs originating from the respective risk management plans of the European Medicines Agency (EMA) for nusinersen and onasemnogene abeparvovec by stating that no clinically relevant thresholds are currently defined for these AEs. Supplementation of data collection is planned after these thresholds are defined, the company stated. This delay in defining thresholds is not appropriate; definition of the data to be collected should be completed before the start of study. The lack of recording of SAEs is not justified.

In addition, it should be noted that analysis of related AEs can be omitted, as this information on the association with medication cannot usually be verified. Furthermore, it is unclear why MedDRA[13]-coded events should only be reported if they are already reported in the registry. MedDRA coding can also be performed subsequently on the basis of the reported free text on an AE and is absolutely necessary for a meaningful analysis.

Beyond the specifications of the G-BA, the company plans the following outcome:

Planned hospitalizations

The company adds an outcome on any planned hospitalizations to the outcomes defined by the G-BA. It remains unclear how these planned hospitalizations are related to the morbidity caused by SMA and how they can be distinguished from the outcomes already collected (e.g., on orthopaedic surgery). The handling of hospitalizations for the purpose of drug administration also remains unclear.

[13] Medical Dictionary for Regulatory Affairs

2.2.2 Study design

Prospective / retrospective data collection

The RPDC-GSAV is a comparative study without randomization. For the study design, the following 2 questions are relevant, among others:

1) Should the study be prospective, retrospective, or a combination of prospective and retrospective data collection?
2) In the case of retrospective data collection, should only data collected concurrently for the 2 interventions be considered in the study or should data collected non-concurrently also be used?

The section on study design in the study protocol of the company does not clarify these questions.

Regarding the question of prospective or retrospective data collection, the company describes elsewhere in the protocol, in connection with the study's inclusion/exclusion criteria, that it wants to use historical data (defined as data collected for the RPDC-GSAV before the start of the study, i.e. data to be collected retrospectively) only for nusinersen and only if the necessary sample size cannot be achieved by prospective data collection alone. The company justifies the exclusion of retrospective data on onasemnogene abeparvovec with the G-BA's decision.

This planning is not appropriate. The use of retrospectively collected data dependent on the future recruitment of patients is incomprehensible, especially in view of the company's comments on the recruitment expected. The company justifies use of retrospective data on nusinersen only as an option with potentially changed treatment standards over time. However, it does not describe whether and, if so, from which point in time such changes have occurred and which period could nevertheless potentially be considered retrospectively (e.g. at least the data from the time of the availability of onasemnogene abeparvovec [data collected concurrently]). The exclusion of the retrospective data collection for onasemnogene abeparvovec remains completely incomprehensible. It is unclear from which section of the G-BA's decision the company derives such a restriction. An exclusion of data to be collected retrospectively on onasemnogene abeparvovec means that the treatment courses recorded between market access (July 2020) and the start of the RPDC-GSAV (according to the study protocol: January 2022) are not considered for the RPDC-GSAV. These treatments even occurred concurrently with treatments with nusinersen. This approach is not appropriate in view of the limited number of patients with this rare disease.

No information is provided in the study protocol on the question of the use of data collected non-concurrently, i.e. data on nusinersen collected before onasemnogene abeparvovec was available. Thus, the company does not comment on the option offered by the G-BA to use such data.

Selection of confounders

In a comparative study without randomization, the relevant confounders must be prespecified in the study protocol. This step is necessary in order, after data collection, to approximate structural equality of the treatment groups in the analysis by (prespecified) adjustment for these confounders.

The procedure of the company for identifying confounders by means of a systematic literature search and the involvement of experts is basically appropriate. An assessment of the systematic search for guidelines and systematic reviews / meta-analyses is included in Appendix A.

However, the list of confounders reflects the company's decisions on patient populations for the analysis and thus for the benefit assessment, which differ from those of the G-BA. These should be corrected and the impact of this correction on the list of confounders should be considered.

For example, the company names "symptom status at baseline" as a confounder instead of dividing the patient population according to symptom status in the research question, as stipulated in the G-BA's decision. On the other hand, it does not consider the confounder "region" because, among other things, it excludes all centres outside of Germany from the data collection because they do not implement the quality requirements of the G-BA. As described in Sections 2.1.1 and 2.2.3, these limitations are questionable. If these decisions are corrected, "region" should be appropriately considered as a confounder.

Sufficient prespecification of the procedure for confounder adjustment in the analysis is missing; the planning is thus inappropriate (see Section 2.3.2).

2.2.3 Data source

The company chooses the SMArtCARE registry as a data source for the RPDC-GSAV. The registry is suitable for the RPDC-GSAV, as it fulfils the necessary quality criteria [3] and has been designated by the G-BA as the primary relevant registry [1]. The G-BA also refers to the inclusion of other registries, provided they meet the necessary requirements.

The company does not make use of the possibility to include further registries. It describes in the study protocol that the G-BA's decision provides for data collection within one data source. This seems to be a misinterpretation of the G-BA's decision. It is correct that the G-BA documents describe that the comparison of onasemnogene abeparvovec and nusinersen should be performed by parallel control (each) within one data source. However, these statements refer to the basic study design, not to the exclusive use of a single registry as a data source. The possibility of combining several sources by means of meta-analysis is explicitly referred to in the G-BA's decision [1,3].

Another registry that could potentially contribute data sets for the RPDC-GSAV is the RESTORE registry maintained by the company itself as a regulatory requirement. The registry

did not meet the requirements for an RPDC-GSAV at the time of concept development [3], but could be another suitable data source with appropriate adaptations. Appropriate adaptations of the registry would be possible for the company as the party responsible for the registry. The company itself stated in the expert discussion on the IQWiG concept for the RPDC-GSAV: "In principle, it is of course possible to make appropriate adaptations or planning for RESTORE based on a concept that considers the important issues discussed today" (translation from German) [8]. Therefore, the company should make the necessary adjustments (especially harmonization of data collection time points with SMArtCARE requirements, training of participating centres, equal efforts for inclusion of nusinersen patients).

Moreover, the company does not plan to use the entire SMArtCARE registry. It rather limits the centres whose data it wants to use in two steps 1) to German centres and 2) within Germany to the centres that meet the quality requirements of the G-BA for the use of onasemnogene abeparvovec. It justifies this planning with the newborn screening for SMA, which starts in Germany in October 2021, but is not yet performed in other countries. In addition, the company fears a bias due to different quality standards and in particular does not want to consider centres that do not use both interventions.

The exclusion of centres outside Germany is not appropriate. Wherever possible, studies involving patients with rare diseases should be conducted internationally in order to obtain timely and robust results, also with small sample sizes. Newborn screening does not prevent the inclusion of centres outside Germany. The G-BA explicitly provides for the investigation of symptomatic patients. Here in particular, centres outside Germany could also make relevant information available prospectively.

It is true that potential differences in quality standards or differences in care should be taken into account. However, the decision whether to include a centre or not should depend on the actual quality or care implemented in that centre. Therefore, (international) centres that are not bound to the quality requirements of the G-BA could potentially be included. The quality standard would have to be examined in each case. In the hearing on the assessment of nusinersen, it was discussed that, especially data from international study centres participating in SMA studies could probably be used [9].

The consideration to exclude centres that do not use both interventions is basically understandable for methodological reasons. However, in the present case of data collection for a rare disease, data from such centres should initially be considered in the analysis. The possible influence of these centres on the results should additionally be investigated in sensitivity analyses (see Section 2.3.3).

Overall, the limitations of the company regarding the data sources for the RPDC-GSAV are critical, as they reduce the number of patients included in the RPDC-GSAV in a relevant way and thus complicate timely and meaningful data collection.

2.2.4 Analysis of data collection

The information on the analysis of data collection in the study protocol is commented on in the context of the comments on the SAP.

2.3 Comments on the SAP

The company does not fulfil the requirements defined by the G-BA in the decision on the RPDC-GSAV regarding the planning of the data analysis. The planning is partly unclear and not described in sufficient detail or is unsuitable. Some of the requirements are not addressed.

2.3.1 Sample size calculation

In the study protocol and the SAP, the company describes a detailed sample size calculation for the 2 study populations defined by the company: A (all patients with 5q SMA with a bi-allelic mutation in the SMN1 gene and up to 2 copies of the SMN2 gene) and B (all patients with 5q SMA with a bi-allelic mutation in the SMN1 gene and 3 copies of the SMN2 gene). These populations differ from the G-BA's research question (see Section 2.2.1).

For the outcomes distinguished (overall survival [OS], event-free survival [EFS], sitting without support at month 18) and the varying assumptions (association between the factors "treatment" and "confounders", ratio of treatment group sizes), the required sample sizes for a power of 90% are presented. The sample sizes vary considerably for both study populations (for A between 48 and 820, for B between 155 and 10,820). On the other hand, under Section 8.3 (Expected Patient Numbers) the study protocol describes that 138 (for A) and 98 (for B) eligible children are expected for the SMArtCARE registry and, due to the study design, all eligible children will be included in the study. Against this background, the sample size calculation presented does not seem very helpful.

According to Section 6 of the SAP (Planned Analysis), a recalculation of the sample size calculation should take place 18 and 36 months after the start of the study. The analysis after 36 months is to decide whether the inclusion of initially only prospective cases should be extended to include retrospective cases or whether the study should even be terminated prematurely due to an insufficient sample size. The term "sample size" is always used here. However, it is not clear which sample size is precisely used, as this depends on various factors (see above). Overall, it remains unclear how the recalculation of the sample size calculation should proceed in detail. In particular, due to the importance of the 36-month analysis, a much more detailed description of these analyses in the SAP is required.

Under the assumptions for sample size calculation, it is outlined that the measure R^2 is used to describe the association between the factors "treatment" and "baseline confounders". It is unclear exactly how this measure R^2 is defined and to which analysis it refers. Presumably, it refers to the logistic regression used to calculate propensity scores and serves as a goodness-of-fit measure. The exact use of this measure and its precise definition should be added. However, in statistics, for binary data, the C statistic (area under the receiver operating characteristics

[ROC] curve [10]) and the Akaike Information Criterion (AIC) are more common and can be considered as alternatives.

For its population of "2 SMN-2 copies", the company refers to an unpublished comparison between onasemnogene abeparvovec and nusinersen, which it conducted for the sample size calculation of the RPDC-GSAV. According to the company, this is based on a comparison of individual study arms of the START and STR1VE-US studies on onasemnogene abeparvovec with study arms of the SHINE studies. A presentation of this unpublished indirect comparison, e.g. in the appendix of the study protocol or as an attachment to the study protocol, is missing. The resulting effects are therefore not verifiable. Furthermore, the dossier assessment A21-68 on onasemnogene abeparvovec [11] shows that

- further data on onasemnogene abeparvovec are available (study STR1VE-EU) and
- the studies on both drugs used by the company are not sufficiently similar. This is due to different exclusion criteria (exclusion of ventilated children from the onasemnogene studies) and marked differences in the duration of disease at study inclusion.

The effects determined by the company in its unpublished comparison are therefore potentially inappropriately justified by the data used by the company and may be considerably overestimated.

In the sample size calculation, the company does not take into account the shifted hypothesis boundaries for the assessment of effects (see Section 2.3.4) [1,3]. This is not appropriate.

2.3.2 Confounder adjustment

Appropriately prespecified confounder adjustment is of particular importance for the analysis of comparative studies without randomization. The information provided by the company on confounder adjustment in the protocol and SAP does not represent appropriate prespecification.

In the context of confounder adjustment, the company defines 3 treatment groups (page 42 of the SAP):

1) patients treated with onasemnogene abeparvovec only
2) patients treated with nusinersen only
3) patients treated with nusinersen and switched to onasemnogene abeparvovec

These groups do not represent an appropriate division of patients into analysable groups. An appropriate division of patients must be based on information available at the start of the study. No information may be used for this purpose that is only available during the course of the study and can therefore already be an effect of the treatment (such as the absence or occurrence of a switch in treatment).

The company states that it will initially perform confounder adjustment on the basis of a propensity score analysis. It remains unclear on which patient classification the propensity score analysis is ultimately based. The presentation on page 42 of the SAP suggests that Groups 1 and 2 above would be used for this purpose, and that patients from Group 3 would be assigned to both Group 1 and Group 2. This approach would be inappropriate in two respects. On the one hand, information is used that is only available during the course of the study while on the other hand, the patients from Group 3 are used twice. But the exact procedure remains unclear. However, the use of information only available in the course of the study means that a propensity score analysis based on this approach is not valid.

Furthermore, the SAP does not describe how the quality of the propensity score analysis is to be checked. According to rapid report A19-43 [12], a sufficient overlap and a sufficient balance of the groups to be compared must have been achieved. Although it is stated that graphical methods as well as permutation tests are to be applied (pages 42-43), further important details are missing. In particular, concrete criteria are missing with regard to what is understood by a sufficient overlap and a sufficient balance of the groups to be compared.

A serious deficiency is the plan described in the SAP to use a regression model as a substitute (a frailty model or a generalized linear mixed model [GLMM], depending on the measurement level of the outcome variable) if the groups to be compared are not sufficiently balanced after applying the propensity score. Insufficient overlap of the groups to be compared cannot be remedied by using a regression model. While it is possible to perform a regression in purely computational terms (unless the two groups are completely separated), this does not mean that the corresponding results can be interpreted in a meaningful way. If there is insufficient overlap between the groups to be compared, a regression model would use extrapolations that are not valid, since correlations are transferred to areas where no data are available at all [13].

Only one procedure is described for the application of the propensity score (fine stratification), although there are numerous other methods for this purpose [14,15]. The usual approach would be to choose a procedure for applying the propensity score so that sufficient overlap and balance of the groups to be compared are achieved. A description is missing in the SAP of a decision algorithm to adjust the propensity score analysis in case of lack of overlap and balance after applying the first method. Similarly, the correct consequence of this is missing if no propensity score method can be found by which sufficient overlap and sufficient balance of the groups to be compared can be achieved. In such a case, the attempt to estimate an effect by means of propensity scores or by means of regression models is not reasonable and the research question investigated must be reconsidered [14].

2.3.3 Analysis of outcomes

In the SAP, depending on the measurement level of the outcomes to be analysed (time-to-event, binary, continuous, count data), the methods “Cox model with time-dependent covariables” and “generalized linear models (GLMs) with different link functions” are mentioned, whereby the treatment is to be considered as a fixed, time-dependent effect in each case. If no confounder

adjustment is performed using the fine stratification method via the propensity score, frailty models and GLMMs are to be used instead, where in addition to the treatment effect, the centre is modelled as a random effect and all confounders are modelled as fixed effects.

The aforementioned overarching model classes in dependence of the measurement level of the outcomes to be analysed are appropriate. However, there are inappropriate sub-aspects in these model descriptions and ambiguities in the modelling details, so that the presentation of the models by which the treatment effect is to be ultimately estimated is insufficient overall. As shown in Section 2.3.2, it is inappropriate to use a regression model as a substitute if the propensity score analysis does not result in sufficient overlap and balance. Moreover, as shown in Section 2.3.7, modelling treatment as a time-dependent effect does not lead to a valid effect estimate.

Furthermore, the use of the centre as a random effect in the modelling must be questioned. Considering the limited sample size, it should rather be assumed that the centre has no relevant influence. Moreover, after considering centres that used only one intervention, the assumption of a random distribution across all centres is implausible. Therefore, the analysis should not include the centre as a random or fixed effect. Sensitivity analyses should then examine a possible centre effect, for example, by omitting centres that used only nusinersen as well as by performing descriptive analyses within centres.

Furthermore, it is not described in which form the confounders should enter the respective outcome model as fixed effects. Do the continuous confounders enter the model in their original unit or are they to be transformed beforehand? Is it assumed that all relationships between the confounders and the respective outcome are linear, or is the consideration of non-linear relationships planned? Is the investigation of interactions planned? These modelling aspects must be described in detail in an SAP, so that it is clear in what form the outcomes will be analysed for final effect estimation. Since these details are missing, the SAP is incomplete.

In its decision, the G-BA stipulates that, besides data collected concurrently, data not collected concurrently should possibly also be considered and requires that it is described how it is to be evaluated whether such data can be used for pooled analyses. The company does not address this issue in its protocol and SAP. Similarly, the protocol and SAP do not include any information on possible pooled analyses from different data sources, as the company intends to conduct data collection exclusively in a subset of the centres of the SMArtCARE registry (see Section 2.2.3).

2.3.4 Consideration of shifted hypothesis boundaries

Due to potentially unknown confounders, a conclusion on the benefit or harm of an intervention can only be derived from the effects observed in a non-randomized study once a certain effect size has been reached. A (positive or negative) conclusion on benefit or harm can be drawn when the confidence interval for the effect observed is above or below a threshold to be defined (test for shifted null hypothesis). The specific threshold results from the quality of the data in

the individual case, including knowledge about relevant confounders [3]. There is no information on this requirement [1] and its implementation in the study protocol or in SAP. This should be supplemented.

2.3.5 Subgroup analyses

The SAP does not contain any information on methods for subgroup analyses, except for the listing of the subgroup factors planned. Only the study protocol provides a rudimentary description of methods. Among other things, it is planned to perform subgroup analyses (for time-to-event outcomes) only if there is a statistically significant interaction between treatment and subgroup. In principle, this approach is methodologically correct. However, the very small sample size to be expected must be taken into account. With this sample size, the interaction test will have insufficient power. As a result, this requirement (significant interaction) probably means that no subgroup analyses can be expected. This is particularly relevant for the factor "symptom status" planned by the company, which it intends to use instead of the basic division of the patient population according to symptom status planned by the G-BA. It is to be feared that, according to these methods, this relevant analysis divided into presymptomatic and symptomatic patients will not be presented at all. It is suggested that because of the small sample sizes to be expected, all relevant subgroup analyses should be calculated without the requirement of a statistically significant interaction and the corresponding results presented.

In its decision, the G-BA requires the performance of subgroup analyses according to the number of copies of the SMN2 gene for the population of presymptomatic patients with up to 3 copies of the SMN2 gene in order to evaluate whether a joint analysis is appropriate. However, the company does not plan these subgroup analyses. In deviation from the G-BA, it plans to divide the population based on the number of SMN2 copies (up to 2 or 3 copies, see Section 2.2.1), irrespective of symptom status.

2.3.6 Handling of missing data

The SAP contains information on how to deal with missing data. It is described that individuals with missing data in the confounder variables should be excluded from all analyses that take these confounders into account. Given the small sample size expected, this approach does not seem appropriate. Every effort should be made to avoid missing information. Any remaining missing information should be replaced in a suitable manner to minimize the loss of results. It is suggested that these missing values be replaced by the multiple imputation approach [16].

There is no information in the study protocol and SAP on the extent to which and the reasons for which missing data are to be expected and how implausible data and outliers are to be dealt with. This information should be added to the SAP.

2.3.7 Handling of treatment switching

with 2.2.1patients treated with onasemnogene abeparvovec only

2) patients treated with nusinersen only

3) patients treated with nusinersen and switched to onasemnogene abeparvovec

The final analysis for effect estimation should then be performed on the basis of treatment episodes and not on the basis of treatment arms. For this purpose, the application is planned of the Cox model with time-dependent covariables, whereby treatment is to be considered as a time-dependent variable.

As shown in Section 2.3.2, the above division of patients is not valid. An appropriate division of patients must be based on information available at the start of the study. For this purpose, no information may be used that is only available during the course of the study and can therefore already be an effect of the treatment.

Furthermore, the Cox model with time-dependent covariables is not an appropriate method for dealing with treatment switching, since the time-dependent variables in this model must not be affected by the treatment itself. However, since the treatment itself is the time-dependent variable here, this assumption is trivially violated. Similarly, with treatment switching, the assumption that the effect of the time-dependent variable "treatment" is identical in all episodes is implausible. It also remains unclear how to deal in the final model with the fact that, after treatment with onasemnogene abeparvovec, there can be no more episodes without this intervention, given that this is a one-time treatment and assuming that the effects of gene therapy persist. As described in working paper GA14-04 [17], the naive application of the Cox model with time-dependent covariables is generally not an appropriate method for dealing with treatment switching.

The usual method in pharmacoepidemiological research for dealing with treatment switching is the new-user design, in which treatment-naive patients are assigned to the group of the respective initial treatment [18]. As with the intention-to-treat principle in randomized controlled trials, all subsequent intercurrent events (including treatment switching) are ignored in the analysis of the primary effect estimate. Initially, no censoring is performed for treatment switching, as this is informative censoring and can lead to biased effect estimates. Since the new-user design lacks randomization, appropriate adjustment must be made for confounding at the start of the study to reduce confounding bias as much as possible. Propensity scores are usually used for this purpose. Of course, the manner, extent, and corresponding time points of treatment switching must be presented here, since – as with the intention-to-treat principle – a great extent of treatment switching can have a relevant biasing influence on the effect estimate of onasemnogene abeparvovec versus nusinersen. Therefore, as a sensitivity analysis, supplemental analyses with censoring should be performed for treatment switching, varying the time of censoring to account for carry-over effects from prior treatment.

If it turns out that the extent of treatment switching is so great that a valid effect estimate of onasemnogene abeparvovec versus nusinersen no longer appears possible, it will not be possible to answer the initial research question (benefit of onasemnogene abeparvovec versus nusinersen) with the available data. A prevalent new-user design may be considered as an

alternative [19]. However, this investigates a different research question, namely, for example, whether patients taking nusinersen benefit from a switch in treatment to onasemnogene abeparvovec. Whether such a design is reasonable and feasible must be decided when information is available on the manner, extent, and time points of treatment switching.

In order to be able to appropriately take into account a greater extent of treatment switching, information on the number of patients switching treatment, including the respective periods under the different treatments, should be part of the information on the course of data collection that is regularly submitted to the G-BA. Depending on the extent of treatment switching, the study protocol can be adjusted via amendments, if necessary.

2.3.8 Analyses planned

The company describes 4 descriptive analyses, one interim analysis and one final analysis in the study protocol and SAP. The planned time points of the analyses differ from those of the G-BA. While in its decision the G-BA describes the analyses to be submitted in relation to the decision date (4 February 2021), the company plans analyses in relation to the start of the study (according to the protocol at the beginning of 2022). The time points for certain analyses also deviate; the G-BA specifies a review of the sample size after 18 months, the company only after 36. In addition, the time point of the final analysis is specified by the company as 60 months after the third interim analysis, deviating from the G-BA's decision. Reasons for these deviations remain unclear.

It remains unclear whether the character of the analyses planned is consistent with that of the G-BA's decision. While the G-BA requests interim analyses at different time points, the company describes descriptive analyses and additionally an interim analysis.

2.4 Overview of main deficiencies of the documents provided by the company

The review of the study protocol and SAP of the company for the RPDC-GSAV on onasemnogene abeparvovec revealed the following major deficiencies:

- the company's planning does not comply with the G-BA's decision, among other things, the company plans deviating research questions: exclusive division of the patient population according to SMN2 copies, no division between symptomatic / presymptomatic patients
- the company's planning potentially leads to results not being available on time or in sufficient volume
  - restriction to SMArtCARE in Germany
  - restriction to centres that fulfil the quality assurance directive of the G-BA
  - no planning for the adaptation and integration of the RESTORE registry, for which the company is responsible
  - no sufficient planning to integrate data already collected on nusinersen and

onasemnogene abeparvovec (from SMArtCARE or other registries).

- the planning of the analysis is insufficient
  - it is in part not detailed enough to ensure sufficient prespecification of the analyses (e.g. confounder adjustment, model selection and adjustment for the outcome analyses)
  - the consideration of shifted null hypotheses is missing in order to also be able to infer an effect in a non-randomized design with sufficient certainty
  - the methods proposed are partly inappropriate (e.g., formation of patient groups by using information that only becomes available during the course of the study, use of a regression model in the case of insufficient overlap of groups after application of propensity scores, consideration of treatment switching via a Cox model with treatment as a time-dependent covariable).

References

1. Gemeinsamer Bundesausschuss. Beschluss des Gemeinsamen Bundesausschusses über die Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V: Onasemnogen-Abeparvovec (spinale Muskelatrophie); Forderung einer anwendungsbegleitenden Datenerhebung und von Auswertungen [online]. 2021 [Accessed: 20.08.2021]. URL: https://www.g-ba.de/downloads/39-261-4702/2021-02-04_AM-RL-XII_awD_Onasemnogen-Abeparvovec_D-549_BAnz.pdf.

2. Gemeinsamer Bundesausschuss. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über die Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V: Onasemnogen-Abeparvovec (spinale Muskelatrophie); Forderung einer anwendungsbegleitenden Datenerhebung und von Auswertungen [online]. 2021 [Accessed: 20.08.2021]. URL: https://www.g-ba.de/downloads/40-268-7293/2021-02-04_AM-RL-XII_awD_Onasemnogen-Abeparvovec_D-549_TrG.pdf.

3. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Konzept für eine anwendungsbegleitende Datenerhebung – Onasemnogen-Abeparvovec; Rapid Report [online]. 2020 [Accessed: 09.02.2021]. URL: https://www.iqwig.de/download/a20-61_anwendungsbegleitende-datenerhebung-onasemnogen-abeparvovec_rapid-report_v1-0.pdf.

4. Novartis Gene Therapies. Routine data collection and evaluations of onasemnogene abeparvovec in Germany - Study Protocol. Protocol Number: COAV101A1DE01. Version: 1.01 (unpublished). 05.08.2021.

5. Novartis Gene Therapies. Routine data collection and evaluations of onasemnogene abeparvovec in Germany - Statistical Analysis Plan. Protocol Number: COAV101A1DE01. Version: 1.01 (unpublished). 05.08.2021.

6. Gemeinsamer Bundesausschuss. Niederschrift (vorläufige Fassung) zum Beratungsgespräch gemäß § 8 Abs. 1 AM-NutzenV vom 11.08.2021, Beratungsanforderung 2021-B-190 der Firma Novartis Gene Therapies zum Wirkstoff Onasemnogen-Abeparvovec (unpublished). 12.08.2021.

7. Gemeinsamer Bundesausschuss. Niederschrift (vorläufige Fassung) zum Beratungsgespräch gemäß § 8 Abs. 1 AM-NutzenV vom 29.06.2021, Beratungsanforderung 2021-B-122 der Firma Novartis Gene Therapies zum Wirkstoff Onasemnogen-Abeparvovec (unpublished). 12.07.2021.

8. Gemeinsamer Bundesausschuss. Zusammenfassende Dokumentation zum Beschluss des Gemeinsamen Bundesausschusses über die Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V: Onasemnogen-Abeparvovec (spinale Muskelatrophie); Forderung einer anwendungsbegleitenden Datenerhebung und von Auswertungen [online]. 2021 [Accessed: 13.09.2021]. URL: https://www.g-ba.de/downloads/40-268-7294/2021-02-04_AM-RL-XII_awD_Onasemnogen-Abeparvovec_D-549_ZD.pdf.

9. Gemeinsamer Bundesausschuss. Nusinersen: Mündliche Anhörung gemäß § 35 a Abs. 3 Satz 2 SGB V – Stenografisches Wortprotokoll [online]. 2021 [Accessed: 13.09.2021]. URL: https://www.g-ba.de/downloads/91-1031-625/2021_04_06_Wortprotokoll_Nusinersen_D-614.pdf.

10. Hosmer DW, Lemeshow S. Applied Logistic Regression. New York, NY: John Wiley & Sons; 2000.

11. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Onasemnogen-Abeparvovec (spinale Muskelatrophie) – Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung [online]. 2021 [Accessed: 17.08.2021]. URL: https://www.iqwig.de/download/a21-68_onasemnogen-abeparvovec_nutzenbewertung-35a-sgb-v_v1-0.pdf.

12. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Konzepte zur Generierung versorgungsnaher Daten und deren Auswertung zum Zwecke der Nutzenbewertung von Arzneimitteln nach § 35a SGB V: Rapid Report; Auftrag A19-43; Version 1.1 [online]. 13.05.2020 [Accessed: 02.06.2020]. URL: https://www.iqwig.de/download/A19-43_Versorgungsnahe-Daten-zum-Zwecke-der-Nutzenbewertung_Rapid-Report_V1-1.pdf.

13. Williamson E, Morley R, Lucas A et al. Propensity scores: from naive enthusiasm to intuitive understanding. Stat Methods Med Res 2012; 21(3): 273-293. https://dx.doi.org/10.1177/0962280210394483.

14. Desai RJ, Franklin JM. Alternative approaches for confounding adjustment in observational studies using weighting based on the propensity score: a primer for practitioners. BMJ 2019; 367: l5657. https://dx.doi.org/10.1136/bmj.l5657.

15. Webster-Clark M, Sturmer T, Wang T et al. Using propensity scores to estimate effects of treatment initiation decisions: State of the science. Stat Med 2021; 40(7): 1718-1735. https://dx.doi.org/10.1002/sim.8866.

16. Sterne JA, White IR, Carlin JB et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009; 338: b2393. https://dx.doi.org/10.1136/bmj.b2393.

17. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Treatment Switching in onkologischen Studien; Arbeitspapier [online]. 2018 [Accessed: 10.12.2018]. URL: https://www.iqwig.de/download/GA14-04_Treatment-Switching-in-onkologischen-Studien_Arbeitspapier_V1-0.pdf.

18. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. Am J Epidemiol 2003; 158(9): 915-920. https://dx.doi.org/10.1093/aje/kwg231.

19. Suissa S, Moodie EE, Dell'Aniello S. Prevalent new-user cohort designs for comparative drug effect studies by time-conditional propensity scores. Pharmacoepidemiol Drug Saf 2017; 26(4): 459-468. https://dx.doi.org/10.1002/pds.4107.

20. Scheffer H, Cobben JM, Matthijs G et al. Best practice guidelines for molecular analysis in spinal muscular atrophy. Eur J Hum Genet 2001; 9(7): 484-491. https://dx.doi.org/10.1038/sj.ejhg.5200667.

21. Wang CH, Finkel RS, Bertini ES et al. Consensus statement for standard of care in spinal muscular atrophy. J Child Neurol 2007; 22(8): 1027-1049. https://dx.doi.org/10.1177/0883073807305788.

22. Prior TW. Carrier screening for spinal muscular atrophy. Genet Med 2008; 10(11): 840-842. https://dx.doi.org/10.1097/GIM.0b013e318188d069.

23. Schroth MK. Special considerations in the respiratory management of spinal muscular atrophy. Pediatrics 2009; 123 Suppl 4: S245-249. https://dx.doi.org/10.1542/peds.2008-2952K.

24. Cuisset JM, Estournet B. Recommendations for the diagnosis and management of typical childhood spinal muscular atrophy. Rev Neurol (Paris) 2012; 168(12): 902-909. https://dx.doi.org/10.1016/j.neurol.2012.07.020.

Appendix A Review of information retrieval to identify confounders

Information retrieval

In order to identify relevant confounders, the company conducted a search for systematic reviews and guidelines on SMA in the following sources:

- bibliographic searches for guidelines and systematic reviews (last search on 23 March 2021
- hand search for guidelines on selected websites (last search on 23 March 2021)

Comment on information retrieval

To identify systematic reviews and guidelines, the company conducted a search in MEDLINE and the Cochrane Database of Systematic Reviews. In addition, the company states that it identified further guidelines via a hand search.

The company's search is not suitable to ensure the completeness of the search result. This is particularly due to the following reasons:

- The bibliographic searches in MEDLINE contain a time limit from 2015 without listing this in the inclusion and exclusion criteria. In the hand search, however, there is no time limit, which means that some documents published before 2015 are included [20-24].
- The search of the company in the Cochrane Database of Systematic Reviews is incomplete, as a time limit of the search is applied to the period June 2015 to June 2020. References published after June 2020 are therefore not considered.

In Table A25, the company lists HTA reports among the exclusion criteria. This exclusion is not appropriate.

In addition, several deficiencies were identified in the documentation of the information retrieval (e.g. a conclusive presentation of all included publications [e.g. study pool, etc.] is missing; the reporting of the results of the hand search in Section 5.3 is incomplete).

Summary

The company's information retrieval is not suitable to ensure the completeness of the search results. However, the basic approach of the company's information retrieval for confounders can be regarded as comprehensible and appropriate. It can therefore be assumed that a sufficiently complete list of potentially relevant confounders was identified.

Appendix B Disclosure of relationships (external experts)

This assessment was prepared with the involvement of an external expert (a statistician). According to § 139b (3) Sentence 2 SGB V, external experts who work on scientific research assignments for the Institute must disclose "all relationships with interest groups, contract institutes, in particular the pharmaceutical industry and the medical device industry, including the type and amount of remunerations." The Institute received a completed form "Form for disclosing relationships" from the expert. The disclosures were evaluated by the Institute's committee specifically established to assess conflicts of interest. No conflicts of interest were identified that would jeopardize professional independence with regard to the processing of the present commission.

The relationships of the external expert are summarized below. All information is based on self-disclosures by the individuals using the "Form for disclosing relationships" as of March 2020. The current form is available at www.iqwig.de. The questions listed in this form can be found following this summary.

Name	Q1	Q2	Q3	Q4	Q5	Q6	Q7
Stürmer, Til	yes	yes	no	yes	yes	yes	yes

The following 7 questions were asked in the "Form for disclosing relationships" (Version March 2020):

Question 1: Employment / self-employed activities / voluntary activities

Are you or have you been, within this year and the last three calendar years, employed by or, on a self-employed or voluntary basis, working for

- an organization in the health care system (e.g. a hospital, an organization in the self-government, a scientific society or a contract research organization),
- a pharmaceutical company
- a medical device manufacturer or
- an industrial interest group

or are you or have you been working on a self-employed or voluntary basis in an independent practice?

Question 2: Consulting relationships

Are you or have you been, within this year and the last three calendar years directly or indirectly advising

- an organization in the health care system (e.g. a hospital, an organization in the self-government, a scientific society or a contract research organization),

- a pharmaceutical company
- a medical device manufacturer or
- an industrial interest group

(e.g. as a reviewer, an expert, in connect with clinical trials as a member of an advisory board / Data Safety Monitoring Board [DSMB] or a steering committee)?

Question 3: Fees

Have you, within this year and the last three calendar years, directly or indirectly received fees (e.g. for talks, trainings, comments or articles) from

- an organization in the health care system (e.g. a hospital, an organization in the self-government, a scientific society or a contract research organization),
- a pharmaceutical company
- a medical device manufacturer or
- an industrial interest group

Question 4: Third-party funds

Have you or your employer or your practice or the institution for which you do voluntary work, within this year and the last three calendar years, received third-party funds (financial support, e.g. for research work, conduct of clinical studies, other scientific services or patent applications) from

- an organization in the health care system (e.g. a hospital, an organization in the self-government, a scientific society or a contract research organization),
- a pharmaceutical company
- a medical device manufacturer or
- an industrial interest group

(If you work in a large institution, it is sufficient to relate the required information to your working unit, e.g. hospital department, research group.)

Question 5: Other support

Have you or your employer or your practice or the institution for which you do voluntary work, within this year and the last three calendar years received, any other financial remuneration or payment in kind (e.g. equipment, staff, support for the organization of meetings, reimbursement of travel expenses or registration fees for trainings/meetings) from

- an organization in the health care system (e.g. a hospital, an organization in the self-government, a scientific society or a contract research organization),

- a pharmaceutical company
- a medical device manufacturer or
- an industrial interest group

(If you work in a large institution, it is sufficient to relate the required information to your working unit, e.g. hospital department, research group.)

Question 6: Stocks, shares, patents, utility models

Do you possess stocks, options or other shares from

- an organization in the health care system (e.g. a hospital, an organization in the self-government, a scientific society or a contract research organization),
- a pharmaceutical company
- a medical device manufacturer or
- an industrial interest group

Do you possess shares in a "sector-specific fund" that is targeted towards pharmaceutical companies or manufacturers of medical devices?

Do you possess patents for a pharmaceutical product or a medical device, or a medical method or a utility model for a pharmaceutical product or a medical device?

Question 7: Other

Are you or have you ever been involved in the development of a

- clinical practice guideline
- clinical study

with a topic similar to this project?

Are there any other circumstances that, from the point of view of an impartial observer, may be assessed as a conflict of interest (e.g. activities in health-related interest groups or self-help groups, political, academic, scientific or personal interests)?

The German report is published under https://www.iqwig.de/projekte/a21-107.html