CADTH

CADTH COMMON DRUG REVIEW

Clinical Review Report

CERLIPONASE ALFA (BRINEURA)

(Biomarin Pharmaceutical (Canada) Inc.)

Indication: For the treatment of neuronal ceroid lipofuscinosis type 2 (CLN2) disease, also known as tripeptidyl peptidase 1 (TPP1) deficiency.

Service Line:	CADTH Common Drug Review
Version:	Final (with redactions)
Publication Date:	June 2019
Report Length:	70 Pages

CADTH

Funding: CADTH receives funding from Canada's federal, provincial, and territorial governments, with the exception of Quebec.

CADTH

Table of Contents

Tables

Figures

Abbreviations

AE	adverse event
CDR	CADTH Common Drug Review
CLN2	neuronal ceroid lipofuscinosis type 2
CLN2QL	Neuronal Ceroid Lipofuscinosis Type 2 Quality of Life Questionnaire
CTD	Common Technical Document
Denver II	Denver II Developmental Screening Test
EQ-5D-5L	EuroQol 5-Dimensions 5-Levels Questionnaire
HRQoL	health-related quality of life
ICV	intracerebroventricular
ITT	intention-to-treat population
LINCL	late infantile neuronal ceroid lipofuscinoses
MCID	minimal clinically important difference
MRI	magnetic resonance imaging
PedsQL	Pediatric Quality of Life
SAE	serious adverse event
SD	standard deviation
TPP1	tripeptidyl-peptidase-1
VAS	Visual Analogue Scale

Drug	Cerliponase alfa (Brineura)
Indication	Treatment of neuronal ceroid lipofuscinosis type 2 (CLN2) disease, also known as tripeptidyl peptidase 1 (TPP1) deficiency
Reimbursement Request	As per indication
Dosage Form(s)	150 mg/5 mL solution for administration by intracerebroventricular infusion
NOC Date	December 19, 2018
Manufacturer	Biomarin Pharmaceutical (Canada) Inc.

Executive Summary

Introduction

Neuronal ceroid lipofuscinosis type 2 (CLN2) disease is an autosomal recessive, neurodegenerative lysosomal storage disorder caused by deficient activity of tripeptidyl-peptidase-1 (TPP1). Deficiency in TPP1 activity results in the accumulation of lysosomal storage materials in the central nervous system, which leads to progressive decline in motor function.[1] It is a pediatric-onset disease that is characterized by motor deterioration, language delay, seizures, ataxia, dementia, blindness, and early death.[2] CLN2 disease estimates of incidence per 100,000 live births range from 0.15 (Portugal), 0.46 (West Germany), and 0.78 (UK) to as high as 9 (Newfoundland).[2,3] Estimates of prevalence of CLN2 disease in Canada range from 10 to 40 patients, according to published sources and a clinical expert consulted by CADTH.[4]

Diagnosis is confirmed through genetic testing and verification of TPP enzyme levels.

The most common initial symptoms are language delay and seizures, which occur in the late infantile period (two to four years of age). Rapid deterioration in motor function and cognitive ability often leads to a loss of voluntary movement and speech by age six. Loss of vision by the ages of seven to 10 is common in those with CLN2 disease, as is dependence on feeding via gastrostomy because of loss of the ability to swallow. Death often occurs in mid-adolescence.[2] CLN2 disease can be diagnosed by the measurement of white blood cell TTP1 activity and by CLN2 gene sequencing. Many mutations have been identified and there are two mutations that have been identified with greater frequency: c.509-1 G>C and c.622 C>T.[5] Screening for CLN2 disease at birth is not performed routinely in Canada.

Cerliponase alfa is a recombinant human TPP1 and the Health Canada–approved indication is for treatment of CLN2 disease. The recommended dosage in patients two years of age and older is 300 mg administered once every other week by intracerebroventricular (ICV) infusion via a surgically implanted reservoir and catheter. The manufacturer proposes that the treating hospitals will provide the devices and facilitate implantation in accordance with the product monograph, as is done similarly with other therapies that require infusion access devices.[1,6]

The objective of this report was to perform a systematic review of the beneficial and harmful effects of cerliponase alfa for the treatment of pediatric patients with CLN2 disease.

Results and Interpretation

Included Studies

Three studies met the inclusion criteria for this review. Study 201 and its extension, Study 202, were phase I/II, multi-centre, open-label, single-arm studies of cerliponase alfa in 24 patients with CLN2. The age range at study entry was between 3.0 and 8.0 years. Doses were escalated during Study 201 until a stable dosage of 300 mg every two weeks was reached. The primary objectives of studies 201, 202, and 203 were to evaluate the safety of cerliponase alfa and the impact on motor-language scores. Post hoc comparisons were made between the treated patients in studies 201 and 202 and historical controls. [redacted] Study 202 is ongoing (N = 23) and Study 203 is ongoing ([redacted]). The most recent efficacy data from these studies is from November 2016. [redacted]

There was no concurrent control group in studies 201 and 202 or 203. Rather, the manufacturer conducted statistical comparisons of the results from studies 201 and 202 with those from a natural history cohort (Study 901). There are several reports that use patients from Study 901 and other sources, and compare these cohorts with treated patients from studies 201 and 202. The main comparisons are found in Schulz et al., who performed an unmatched (control N = 42) and a matched (control N = 17) comparison. [redacted]

The primary efficacy outcome of studies 201 and 202 was the CLN2 motor-language rating scale, which was adapted from pre-existing scales. The resulting motor-language scale has a range of 0 to 6 points with lower scores representing worse symptoms (0 to 3 for motor and 0 to 3 for language). The primary efficacy end point was a responder analysis during the 300 mg dose period based on the motor-language scale in the intention-to-treat (ITT) population. Response was defined as the absence of an unreversed two-point decline or score of zero in the motor-language score by week 96 (Study 202).

Key limitations in both trials were the small sample sizes, lack of concurrent control groups, uncertain validity, and inconsistencies in the scales used to compare treated patients with historical controls.

Efficacy

There were no deaths in patients treated with cerliponase alfa in studies 201 and 202 or Study 203.

In studies 201 and 202, 20 out of 23 (87%) of patients taking cerliponase alfa had an absence of an unreversed two-point decline or score of 0 in the motor-language score at 96 weeks (response). Patients had a mean decline (worsening) of 0.7 (standard deviation [SD]: 0.8) points in the motor-language score after more than 97 weeks of treatment with the 300 mg dose, relative to baseline, while the motor-language-vision scores (range: 0 to 9) decreased by a mean of 1.0 point (SD: 1.2) and the total CLN2 scale (motor, language, vision, seizure, and range: 0 to 12) decreased by a mean of 0.2 points (SD: 2.4). The mean rate of decline of motor-language score per 48-week period was 0.27 (SD: 0.35) for the ITT population (N = 23) and was 0.42 for the safety population (N = 24).

In studies 201 and 202, 13 patients (56%) showed worsening of the motor-language score, nine patients (39%) showed no change, and one patient improved by one point, relative to baseline. Very few efficacy data were available from Study 203 because of the small sample size and the short duration of drug exposure at the time the interim report was written.

Health-related quality of life scores [redacted] The clinical significance of these changes is unclear, according to a clinical expert consulted by CADTH.

Post hoc comparisons were made between the treated population from studies 201 and 202, and the historical control populations. There were several historical control populations referenced in the study reports and publications with sample sizes of 42 (unmatched comparison), 17, and 21 (matched comparisons). The manufacturer stated that the matched analysis using a historical control of [redacted] was the most appropriate comparison, but did not explain why a matched comparison with N = 17 was used in the main publication for these studies.

The median time until a two-point decline in the motor-language scale score among the unmatched historical controls (N = 42) was 345 days (49.3 weeks), but the median was not reached among the treated patients in studies 201 and 202; 9% of the treated patients had a decline of two points at 345 days. The unadjusted mean rate of decline in the motor-language scale score per 48-week period was 0.27 ± 0.35 points among the 23 treated patients as compared with 2.12 ± 0.98 among the 42 unmatched historical controls, a difference of 1.85 ± 0.21 points (95% confidence interval, 1.51 to 2.18; $P < 0.001$). The hazard ratio comparing the treated population in studies 201 and 202 with the unmatched historical control (N = 42) was 0.08 (95% confidence interval, 0.02 to 0.23) favouring treated patients for the time to unreversed two-point or greater decline in the motor-language score. The hazard ratio comparing the treated population in studies 201 and 202 with the matched historical control ([redacted]) for the same outcome was [redacted] favouring treated patients.

Compared with the historical control data, patients receiving cerliponase alfa appeared to experience a two-point or greater decline in the CLN2 motor-language scale score at a slower rate. However, comparison with a historical control group cannot produce results that are as reliable as those within a randomized study. The investigators attempted to overcome this limitation by matching the treated patients with historical control patients using several known prognostic factors (e.g., age, genotype, baseline motor-language score). These methods would be expected to create populations that are more similar than unmatched populations, but the amount of residual confounding and the degree to which the treated and untreated patients differ remains unknown. After matching, there may be

other known and unknown prognostic factors that remain unbalanced. There are other limitations of using historical controls, including the difficulties in assessing motor, language, seizure, or vision symptoms retrospectively from patient records.

Another significant limitation is that while the scales used to assess symptom severity (motor, language, seizure, and vision) applied the same numerical gradations in the treated patients and the historical control groups, the definitions corresponding to the severity levels (0, 1, 2, and 3) were not the same. This creates uncertainty in the interpretation of comparative analyses of the motor-language scores in treated patients versus historical controls.

Study 201's extension, Study 202, is currently ongoing and no efficacy data are available beyond a follow-up period of approximately three years (with the most recent report provided from November 2016). Thus, it is unclear if the suggested benefits of treatment with cerliponase alfa versus no treatment, in terms of language and motor deterioration, would be maintained beyond this time period. In addition, the comparative efficacy of cerliponase alfa versus no treatment, in terms of mortality and of health-related quality of life over the available study period, was not reported.

Harms

The assessment of harms was limited because there were no comparative adverse event (AE) data for the historical control group. CLN2 disease itself is associated with many symptoms and events that would be classified as AEs or serious AEs, making it difficult to assess the relationship with cerliponase alfa for some AEs in the studies (e.g., seizures). According to a clinical expert consulted by CADTH, the risks related to implantation, maintenance, and eventual removal of the ICV system are a significant consideration when deciding whether to initiate therapy with cerliponase alfa. AEs thought to be related to cerliponase alfa and/or its administration were evident in the clinical trials and included hypersensitivity, pyrexia, pleocytosis, device malfunction, and device-related infection. In studies 201 and 202, hypersensitivity events were reported 37 times in 15 patients (63%). In studies 201 and 202, device-related AEs were reported 34 times in 12 patients (50%). Sequelae of these events and treatment of these events related to treatment with cerliponase alfa can have a major impact on the child's quality of life and the child's family, according to a clinical expert consulted by CADTH. However, patients' experiences with the ICV system are not universally negative, according to another clinical expert consulted by CADTH. Eventually, the device needs to be removed and replaced, as occurred in two patients after four years of usage. Two ICV access devices were removed from two patients in Study 202 due to infection after more than four years of device use. This is also an important consideration when weighing the potential benefits and risks of therapy.

[redacted]

Potential Place in Therapy[1]

There is no cure for CLN2 disease and there is no treatment capable of disease modification. Diagnosis is confirmed through genetic testing and verification of TPP enzyme levels. In cases where the pathogenicity of mutations is not certain, additional evidence, such as decreased levels of TPP enzyme activity, are required. Given that the enzyme replacement therapy relies on disease manifestations through abnormally low TPP enzyme activity, the treatment would not be recommended where decreased TPP enzyme activity is not expected from either DNA testing or enzyme testing. Demonstration of storage on biopsy material is not sufficient alone to qualify for enzyme therapy with this product. The clinical features of the disease include seizures, movement disorders, behavioural problems, sleep disturbances, and delays in growth, language, motor, cognitive, and visual function. The seizures may be hard to control with anti-seizure medication, the visual impairment is severe, and the developmental delay is significant, leading to inability to care for self, walk, or mobilize from bed to chair.

An ideal treatment would improve survival, or at least allow children to stay ambulatory for a longer time period, maintain speech, maintain vision, and control seizures.

In the absence of any alternative, cerliponase may delay the progression of motor and language skills in some patients with CLN2 over a period of time, and may be a treatment option for some patients. Patients receiving cerliponase alfa should be managed by a specialist who is familiar with the treatment of CLN2 disease and the treatment of drug- and device-related complications, and has familiarity with other supportive options, palliative care, and nutrition and rehabilitative services.

The clinical experts consulted by CADTH for this review stated that cerliponase alfa may be started in some diagnosed CLN2 cases prior to symptom onset. Decisions to stop treatment would be made in the context of the child's symptoms, treatment goals of the family, and the physician's professional judgment. Continued use of cerliponase in the presence of device complications, such as multiple trials of reservoir insertions, reservoir not allowing cerliponase alpha to be infused, or deterioration of the clinical scoring system to the point of reaching the 0 score on cerliponase alfa treatment, would not be appropriate. Stopping criteria have not been established for cerliponase alfa and the optimal duration of treatment is not known. Patients must meet qualification criteria for treatment using diagnostic testing, and patients and/or families should have the ability to comply with treatment, and have a clear understanding of the risks and potential complications of using implantable devices and the conditions under which cerliponase alfa could be stopped.

[1] This information is based on information provided in draft form by the clinical expert consulted by CADTH Common Drug Review reviewers for the purpose of this review.

Conclusions

Patients treated with cerliponase alfa appear to have less deterioration in motor-language symptoms compared with historical, untreated control groups, and the difference may be clinically meaningful. While cerliponase alfa appears to slow the deterioration of motor-language symptoms in children with CLN2 disease, the use of historical control groups creates uncertainty about the magnitude of possible positive effects on motor-language scores and other outcomes. In studies 201 and 202, the historical control groups used may not be comparable with the cerliponase alfa–treated patients on important prognostic factors. It is also not clear whether cerliponase alfa improves overall health-related quality of life or reduces mortality compared with no treatment. The potential benefits of slower symptom deterioration need to be considered in the context of the known and unknown harms of using cerliponase alfa and the risks associated with ICV drug administration.

Table 1: Summary of Results

	Single-Arm Study	CTD Matched Analysis	
Outcome	**Studies 201 and 202 N = 23**	**Studies 201 and 202 Treated** [redacted]	**Study 901 Historical Control** [redacted]
Absence of an unreversed **2-point** decline or score of 0 in the motor-language score at 48 weeks, n (%)	20 (87)		
Absence of an unreversed **2-point** decline or score of 0 in the motor-language score at 96 weeks, n (%)	20 (87)	[redacted]	[redacted]
Motor-language score change from baseline to last observation (> 97 weeks) (SD)	Baseline : 3.5 (1.2) Change : –0.7 (0.8)		
Health-Related Quality of Life Score Change From Baseline to Week 97 (SD)			
PedsQL Parent Report for Toddlers	Baseline : 60.7 (12.8) Change : –5.7 (18.9); n = 21		
PedsQL Family Impact Module	Baseline : 61.4 (14.3) Change : –1.1 (19.6); n = 21		
CLN2 Disease-Based QoL Instrument	Baseline : 74.2 (13.8) Change : +3.1 (14.4); n = 21		
EQ-5D-5L EQ VAS from baseline to week 49	Baseline : 79.8 (15.2) Change : –9.9 (24.0); n = 21		

CLN2 = neuronal ceroid lipofuscinosis type 2; CTD = Common Technical Document 2.7.3; EQ-5D-5L= EuroQol 5-Dimensions 5-Levels questionnaire; PedsQL= Pediatric Quality of Life Inventory; QoL = quality of life; SD = standard deviation; VAS = visual analogue scale.

Note: The motor-language score has a range of 0 to 6 points, with a lower score indicating poorer function. All health-related quality of life scoring systems have a range of 0 to 100 points, with a lower score representing worse condition. The baseline for EQ-5D-5L is the start of Study 202, all other baselines are the start of Study 201. The CTD analysis was a matched analysis performed by the manufacturer using a historical control population of 21 patients matched to 21 patients from studies 201 and 202. Matching in the CTD analysis was done on two factors (baseline motor-language score and age).

Introduction

Disease Prevalence and Incidence

Batten disease, also called neuronal ceroid lipofuscinosis, describes a heterogeneous group of lysosomal storage disorders that are the most common inherited progressive neurodegenerative disorders in children. The disease is caused by mutations in the neuronal ceroid lipofuscinoses genes.[7] Neuronal ceroid lipofuscinosis type 2 (CLN2) disease is an autosomal recessive, neurodegenerative lysosomal storage disorder caused by deficient activity of tripeptidyl-peptidase-1 (TPP1). Deficiency in TPP1 activity leads to cell death caused by intracellular accumulation of material that would normally be metabolized by this enzyme in the lysosome.[8] Deficiency in TPP1 activity results in the accumulation of lysosomal storage materials in the central nervous system, which leads to progressive decline in motor function.[1] CLN2 is a pediatric-onset disease that is characterized by motor deterioration, language delay, seizures, ataxia, dementia, blindness, and early death.[2] Estimates of incidence of CLN2 disease per 100,000 live births range from 0.15 (Portugal), 0.46 (West Germany), and 0.78 (UK) to as high as 9 (Newfoundland).[2,3] Estimates of prevalence of CLN2 disease in Canada range from 10 to 40 patients, according to published sources and a clinical expert consulted by CADTH.[4]

The most common initial symptoms are language delay and seizures, which occur in the late infantile period (two to four years of age). Rapid deterioration in motor function and cognitive ability often leads to a loss of voluntary movement and speech by age six. Loss of vision by the ages of 7 to 10 is common in those with CLN2 disease, as is dependence on feeding via gastrostomy because of loss of the ability to swallow. Death often occurs in mid-adolescence.[2] CLN2 disease can be diagnosed by the measurement of white blood cell TTP1 activity and by CLN2 gene sequencing. Many mutations have been identified and there are two mutations that have been identified with greater frequency: c.509-1 G>C and c.622 C>T.[5] Screening for CLN2 disease at birth is not performed routinely in Canada.

Standards of Therapy

There are currently no pharmacologic therapies available that target the underlying cause of CLN2 or that can alter the natural history of the disease. A number of therapeutic approaches have been, or are currently being, investigated, including stem cell therapy, enzyme replacement therapy, lysosome modulation, anti-inflammatory approaches, RNA (ribonucleic acid)/DNA modulating compounds, and gene therapy.[9] Symptomatic pharmacologic treatment is common among children to treat seizures, pain, musculoskeletal symptoms, movement disorders, excessive airway secretions, difficulty sleeping, mood changes, and other symptoms. Patients need to be managed by multidisciplinary teams that include physicians, nurses, therapists (physical, occupational, speech), respiratory therapists, dieticians, psychologists, social workers, and counsellors. A holistic approach would include pharmacologic and non-pharmacologic approaches tailored to optimize the child's ability to participate in school and home activities, maintain ambulation, and maintain means of communication. Because CLN2 is a fatal disease, palliative care goals, such as minimization of pain and discomfort, are important in all stages of the disease.[2]

Drug

Cerliponase alfa is a recombinant human tripeptidyl-peptidiase-1. It is a proenzyme that is taken up by target cells in the central nervous system and is translocated to the lysosomes through the cation independent mannose-6-phosphate receptor. Cerliponase alfa is activated in the lysosome and the activated proteolytic form of recombinant human tripeptidyl-peptidiase-1 cleaves tripeptides from the N-terminus of proteins.[8]

The Health Canada–approved indication is for treatment of CLN2 disease. It is available as a 30 mg/mL solution for intracerebroventricular (ICV) infusion. The recommended dosage in patients two years of age and older is 300 mg administered once every other week via a surgically implanted reservoir and catheter. Aseptic technique must be strictly observed during preparation and administration. Patients in clinical trials were observed in hospital for 48 hours after each dose was administered. Outside of clinical trials in Canada, patients can receive the drug in a hospital day unit over six to eight hours, according to a clinical expert consulted by CADTH. The manufacturer proposes that the treating hospitals will provide the devices and facilitate implantation in accordance with the product monograph, as is done similarly with other therapies that require infusion access devices. One such reservoir system used in some Canadian centres is the Ommaya reservoir. The product monograph specifies the materials and devices that should be used to administer cerliponase alfa.[1,6]

There is no guidance in the product monograph regarding length of therapy or criteria for stopping treatment. Cerliponase alfa is expected to be used for several years in most patients.

Objectives and Methods

Objectives

To perform a systematic review of the beneficial and harmful effects of cerliponase alfa for ICV infusion for the treatment of pediatric patients with CLN2 disease, also known as TPP1 deficiency.

Methods

Studies selected for inclusion in the systematic review included pivotal studies provided in the manufacturer's submission to the CADTH Common Drug Review (CDR) and Health Canada, as well as those meeting the selection criteria presented in Table 2.

Table 2: Inclusion Criteria for the Systematic Review

Patient Population	Pediatric patients with CLN2 disease **Subgroups of interest:** • Disease severity at start of treatment (e.g., mild, moderate, severe) • Time of onset (e.g., later infantile versus juvenile) • Age group (e.g., younger versus older)
Intervention	Cerliponase alfa
Comparators	Placebo or best supportive care
Outcomes	**Efficacy outcomes:** • Survival • Impact on symptoms, including neuropsychological (motor, cognitive, language), vision, seizures, and pain using validated scales or symptom improvement response that brings clear benefit to patients (e.g., improvement of swallowing function to allow removal of feeding tube, return of ability to walk, reduction of seizure incidence to null) • Health-related quality of life • Caregiver burden using validated scales • MRI changes **Harms outcomes:** SAEs, AEs , WDAEs Notable harms: administration-related (e.g., infection), cerebrospinal fluid pleocytosis, gastrointestinal (e.g., vomiting), fever, hypersensitivity
Study Design	Randomized or non-randomized trials

AEs = adverse events; CLN2 = neuronal ceroid lipofuscinosis type 2; MRI = magnetic inference imaging; SAEs = serious adverse events; WDAEs = withdrawal due to adverse events.

The literature search was performed by an information specialist using a peer-reviewed search strategy.

Published literature was identified by searching the following bibliographic databases: MEDLINE All (1946–) via Ovid, Embase (1974–) via Ovid, and PubMed. The search strategy consisted of both controlled vocabulary, such as the National Library of Medicine's MeSH (Medical Subject Headings), and keywords. The main search concept was Brineura (cerliponase alfa).

No methodological filters were applied to limit retrieval to study type. Where possible, retrieval was limited to the human population. Retrieval was not limited by publication year or by language. Conference abstracts were excluded from the search results. See Appendix 2 for the detailed search strategies.

The initial search was completed on August 23, 2018. Regular alerts were established to update the search until the meeting of the CADTH Canadian Drug Expert Committee on January 16, 2019. Regular search updates were performed on databases that do not provide alert services.

Grey literature (literature that is not commercially published) was identified by searching relevant websites from the following sections of the *Grey Matters* checklist (https://www.cadth.ca/grey-matters): health technology assessment agencies, health economics, clinical practice guidelines, drug and device regulatory approvals, advisories and warnings, drug class reviews, and databases. Google and other Internet search engines were used to search for additional Web-based materials. These searches were supplemented by reviewing the bibliographies of key papers and through contacts with appropriate experts. In addition, the manufacturer of the drug was contacted for information regarding unpublished studies.

Two CDR clinical reviewers independently selected studies for inclusion in the review based on titles and abstracts, according to the predetermined protocol. Full-text articles of all citations considered potentially relevant by at least one reviewer were acquired. Reviewers independently made the final selection of studies to be included in the review, and differences were resolved through discussion. Included studies are presented in Table 3; excluded studies (with reasons) are presented in Appendix 3.

Results

Findings From the Literature

A total of three studies were identified from the literature for inclusion in the systematic review (Figure 1). The included studies are summarized in Table 3. A list of excluded studies is presented in Appendix 3.

Figure 1: Flow Diagram for Inclusion and Exclusion of Studies

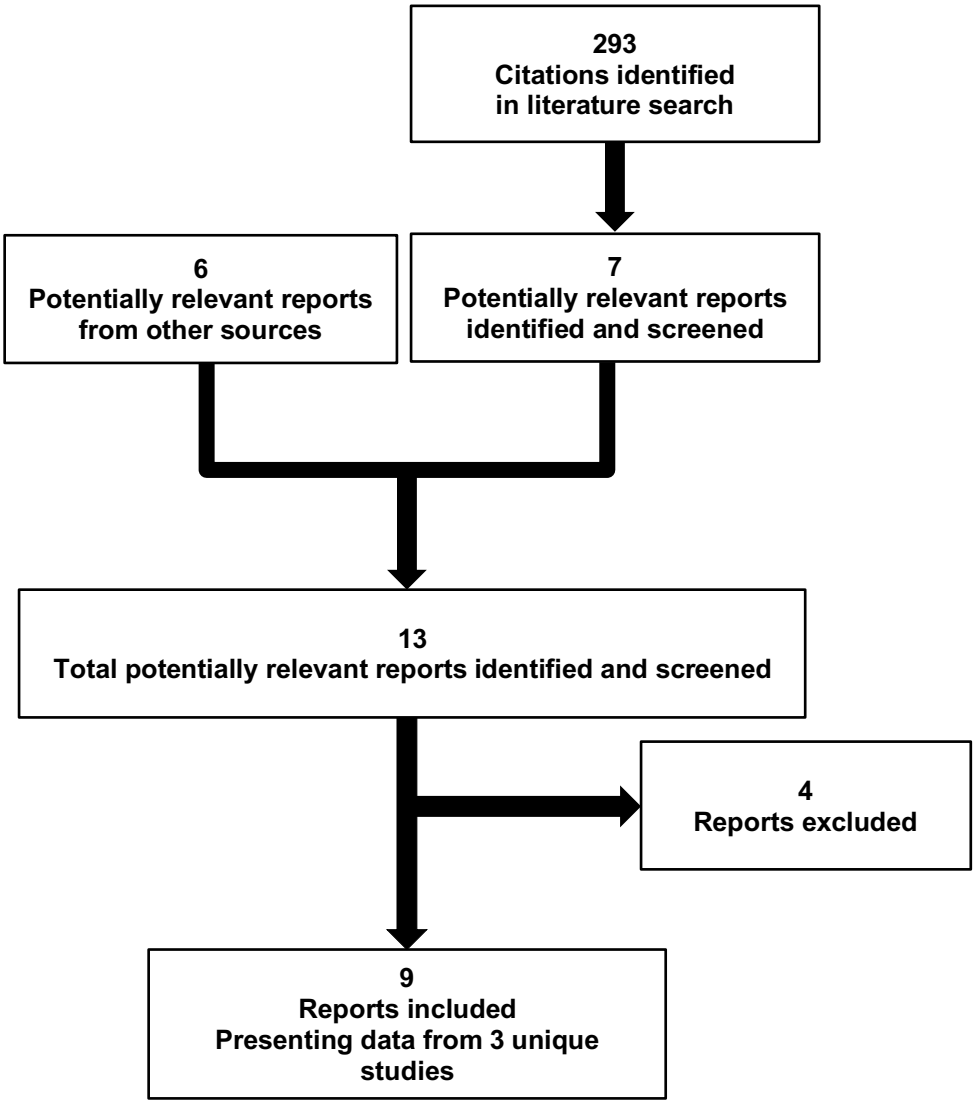


Table 3: Details of Included Studies

		201	202	203
Designs and Populations	**Study design**	Open-label, phase I/II first-in-human, dose escalation, not randomized	Open-label extension of Study 201 (ongoing – last patient visit planned December 2020)	Open-label, phase II, non-randomized (ongoing) sibling study
	Locations	5 centres in 4 countries: Germany, UK, US, Italy	Same as 201	Two centres: Germany and US
	Patients (N)	24	23 (all from Study 201)	[redacted]
	Inclusion criteria	• CLN2 diagnosis by TPP1 enzyme activity and genotype analysis • Score between 3 and 6 out of 6 points on a motor-language scale • ≥ 1 on motor scale • ≥ 1 on language scale • 3 to 15 years of age	Patients who completed 201 with a motor-language score of 3 to 6 and ≥ 1 point in each domain (all were included except one patient who withdrew after the first dose during study 201)	• At least 1 sibling in Study 201 • CLN2 diagnosis by TPP1 enzyme activity • Score between 3 and 6 out of 6 points on a motor-language scale • ≥ 1 year of age
	Exclusion criteria	• Generalized motor status epilepticus or severe infections within four weeks of first drug dose • Previous stem cell, gene therapy, or enzyme therapy for CLN2 • Contraindications for neurosurgery		• Other neurologic disease • Feeding tube • Previous stem cell, gene therapy or enzyme therapy for CLN2 • Generalized motor status epilepticus or severe infections within four weeks of first drug dose
Drugs	**Intervention**	Cerliponase alfa 30 mg/mL by ICV infusion of 2.5 mL/hour for 4 hours. Dose escalation phase: initial doses were 30 mg, 100 mg, or 300 mg Dose maintenance phase: all patients received 300 mg q2w	Cerliponase alfa 300 mg q2w by ICV infusion	[redacted]
	Comparator(s)	No concurrent control group	No concurrent control group. Historical cohort used as a comparison group	No concurrent control group
Duration	**Phase**			
	Dose escalation	Between 4 and 22 weeks	NA	NA
	Stable dose period	48 weeks	NA	Ongoing; planned up to 96 weeks
	Follow-up	NA	Ongoing; planned up to 240 weeks	NA

		201	202	203
Outcomes	**Primary end point**	Responder analysis using motor and language score from start of 300 mg dose period to week 48. Rate of decline also measured	Same response outcome as in Study 201 measured at week 96. Rate of decline also measured	Motor-language score
	Other end points	Brain MRI assessment; Denver II Developmental Test; PedsQL Parent Report for Toddlers; CLN2QL Questionnaire; Parent and Family impact	Same as in Study 201 plus EuroQol EQ-5D-5L	Brain MRI assessment; motor+language+visual+/– seizure scores; Denver II Developmental Test; mUBDRS-Movement scale; OCT; PedsQL; CLN2QL Questionnaire scale; ITQoL; EQ-5D-5L
Notes	**Publications**	Schulz et al[10] (This publication presented comparisons of patients from studies 201 and 202 with two historical control groups)		None

CLN2 = neuronal ceroid lipofuscinosis type 2; CLN2QL = Neuronal Ceroid Lipofuscinosis Type 2 Quality of Life; EQ-5D-5L = EuroQol 5-Dimensions 5-Levels Questionnaire; ICV = intracerebroventricular; ITQoL= infant toddler module; mUBDRS= modified unified Batten disease rating scale seizure inventory; MRI = magnetic resonance imaging; NA = not applicable; OCT = optical coherence tomography; PedsQL = Measurement Model for Pediatric Quality of Life; QoL = quality of life; q2w = every two weeks; TPPI = tripeptidyl-peptidase-1.

Note: Four additional reports were included: CADTH Common Drug Review Submission;[4] FDA Statistical Review for Brineura;[11] European Medicines Agency Report;[12] National Institute for Health and Care Excellence Committee Papers for Cerliponase alfa.[13]

Source: Schulz et al.;[10] Clinical Study Report 190-201;[14] Clinical Study Report 190-202;[15] Clinical Study Report for 190-203;[16] Development Safety Update Report #5.[17]

Included Studies

Description of Studies

Study 201 was a phase I/II, open-label, first-in-human trial. The main objectives of Study 201 were to evaluate the safety and efficacy of cerliponase alfa administered via a surgically implanted ICV device in children with CLN2. A total of 24 patients were enrolled and the first 10 patients underwent a dose escalation phase of 4 to 22 weeks, followed by a stable dose period of 48 weeks. The next 14 patients subsequently enrolled were initiated directly into the stable dose period, which lasted 48 weeks (Figure 2).

All patients from Study 201 were given the opportunity to continue open-label treatment with cerliponase alfa for an additional 240 weeks in Study 202, an ongoing open-label extension study. Study 201 was conducted between 2013 and 2015. Study 202 was initiated in February 2015; according to the manufacturer, the most recent available unpublished efficacy data are from November 2016 and the estimated end date is in December 2020.

Study 203 started in 2016 and is an ongoing phase II open-label trial in children with CLN2 who are siblings of the children in studies 201 and 202. The design and objectives of Study 203 are similar to studies 201 and 202. Upon request, the manufacturer provided the most recent available unpublished efficacy data; a report with a data cut-off of November 2016 based on [redacted].[16]

Some safety data are available for studies 201, 202, and 203, based on the manufacturer's Development Safety Update Report with a data cut-off date of April 2018,[17] but most of the data from these three trials in this CADTH report are from the November 2016 data cut-off unless otherwise indicated.

There was no concurrent control group in studies 201, 202, or 203; however, in response to a request from CDR, the manufacturer provided a description of numerous sources for historical control data,[6] which included the following:

- the manufacturer's unpublished Study 901, a natural history cohort based on patient registries ([redacted])[18]
- the DEM-CHILD data set (an expansion of one of the cohorts from Study 901, n = 74) and the Weill Cornell Medical College data set (n = 66); published by Nickel et al.[19]
- the FDA Statistical Review, which refers to two historical control populations (N = 42 and N = 17)[11]
- Schulz et al., who used a subset of the DEM-CHILD data set and performed unmatched (N = 42) and matched (N = 17) comparisons with the treated patients from studies 201 and 202[10]
- the manufacturer's Common Technical Document (CTD) section 2.7.3 ([redacted]).[4]

The comparative analyses mentioned in this CADTH report mainly come from the latter two documents: Schulz et al. and the CTD section 2.7.3. These documents contain one analysis with patients who were not matched (Schulz et al., 42 controls versus 24 treated patients from studies 201 and 202) and two analyses in which patients were matched by several prognostic factors (Schulz et al. –17 controls versus 17 treated patients from studies 201 and 202; [redacted]).

Populations

Inclusion and Exclusion Criteria

Study 201 included patients aged 3 to 15 years with a confirmed diagnosis of CLN2 disease, by TPP1 enzyme activity (using a dried blood spot) and CLN2 genotype analysis. Patients had mild-to-moderate CLN2 disease, as measured by a two-domain (motor and language) score of 3 to 6 on the adapted CLN2 disease rating scale; patients were required to have a score of at least 1 in each of the two domains.

Patients were excluded if they had had previous stem cell, gene therapy, or enzyme replacement therapy for CLN2 disease. The patients must have had no contraindications for neurosurgery (such as congenital heart disease or severe respiratory impairment). There must have been no generalized motor status epilepticus or severe infections within four weeks before the first scheduled drug dose. Patients could have no underlying condition that would make the patient prone to complications from using an ICV shunt (such as hydrocephalus or ventricular shunts).

Patients who completed 48 weeks of treatment in Study 201 were eligible to enroll in Study 202 if they had a motor-language score of 3 to 6 and at least one point in each domain.

Children enrolled in Study 203 could be as young as one year of age and had to be a sibling of a child who was participating in studies 201 and 202. Patients were diagnosed using TPP1 enzyme analysis in Study 203. Genotype analysis was not required. Other inclusion and exclusion criteria for Study 203 were similar to the criteria used for studies 201 and 202.

Baseline Characteristics

The baseline characteristics of the patients in studies 201 and 202 and Study 203 are summarized in Tables 4 and 5. In the treated cohort of studies 201 and 202, 23 out of 24 patients were white (one was Asian), 38% of patients were male, with a median age of three years at disease onset and a median of four years of age at study enrolment. The most common symptoms present at the beginning of Study 201 were epilepsy, ataxia, and language disorder.

There were very few baseline data available from the four patients enrolled in Study 203. All were [redacted].

In the unmatched comparison, there were several imbalances between the patients in studies 201 and 202 and the historical control (N = 42, Table 4). There was an imbalance in sex between the treated patients in studies 201 and 202 and the historical cohorts, but sex is not likely to affect prognosis in patients with CLN2, according to the clinical experts involved in this CDR review. There was an imbalance in the number of key mutations and this could reflect significant prognostic differences between the study groups. There was also a difference in the decades in which the patients were born. No patients were born prior to 2000 in studies 201 and 202. Birth decade could be a significant prognostic factor because standard care in recent decades has evolved and is not the same as it was prior to 2000. In the historical cohort (N = 42), there were no data on motor-language score or disease symptoms.

[redacted].[4] The N = 17 historical control group in Schulz et al. was matched to the treated population using three factors: exact motor-language score, age within three months, and genotype (equal number of common alleles).[10]

[redacted] There was an imbalance in sex between the historical cohort ([redacted]) and the populations in studies 201 and 202. There was a similar distribution of known mutations between the two groups, but there were many patients in both groups for whom the mutation status was “other” or “missing”; therefore, it is not known whether the matching achieved a similar genotypic profile across the two groups.

There were no baseline data reported for the N = 17 historical control group reported by Schulz et al.[10]

Table 4: Summary of Baseline Characteristics: Studies 201 and 202 and Study 203 Unmatched Comparison

	Historical Control Used for Unmatched Comparison N = 42	Studies 201 and 202 N = 24	Study 203 [redacted]
Male, n (%)	25 (60)	9 (38)	[redacted]
Genotype			[redacted]
Two common allele	24 (57)	9 (38)	
One common allele and one uncommon allele	11 (26)	8 (33)	
Two uncommon alleles	7 (17)	7 (29)	
Decade born			[redacted]
Pre-1980	4 (10)	0	
1980s	2 (5)	0	
1990s	19 (45)	0	
2000s	16 (38)	13 (54)	
≥ 2010	1 (2)	11 (46)	
Median age at disease onset (range), years	NR	[redacted]	[redacted]
Mean age at enrolment, years (SD)	NA	5 (1.2)	[redacted]
Median age at enrolment, years (range)	NA	[redacted]	[redacted]
Mean M-L at screening (SD)	NA	3.7 (1.0)	[redacted]
Mean M-L at study baseline (SD)	NA	[redacted]	[redacted]
Mean M-L at 300 mg baseline (SD)	NA	[redacted]	[redacted]
M-L score at screening, n (%)	NA		
6		2 (8)	[redacted]
5		2 (8)	[redacted]
4		7 (29)	[redacted]
3		13 (54)	[redacted]
2		0	[redacted]
1		0	[redacted]
Symptoms present in history, n (%)	NR		
Ataxia		[redacted]	[redacted]
Dysarthria		[redacted]	[redacted]
Epilepsy		[redacted]	[redacted]
Hypotonia		[redacted]	[redacted]
Language disorder		[redacted]	[redacted]
Psychomotor impairment		[redacted]	[redacted]
Seizure		[redacted]	[redacted]
Speech disorder		[redacted]	[redacted]

M-L = motor-language; NA = not applicable; NR = not reported; SD = standard deviation.

Note: In studies 201 and 202, "screening" occurred before baseline; "study baseline" is at start of treatment, regardless of dose; "300 mg baseline" is start of 300 mg treatment period. Symptoms history lists symptoms that occurred in 15% or more of patients. As of November 2016, four patients had enrolled. As of April 2018, 11 patients had enrolled, but there are no baseline data or efficacy data available for these patients.

Source: FDA Statistical Reviewt,[11] Schulz et al.,[10] Clinical Study Reports for studies 201, 202, and 203.[14-16]

Table 5: Summary of Baseline Characteristics — Studies 201 and 202 Matched Comparison ([redacted])

	Historical Control Used for Matched Comparison in CTD [redacted]	Studies 201 and 202 [redacted]
Age at enrolment in 201, years		
Mean (SD), years	[redacted]	[redacted]
Median (range), years	[redacted]	[redacted]
Male, n(%)	[redacted]	[redacted]
M-L score at baseline, n(%)		
6	[redacted]	[redacted]
5	[redacted]	[redacted]
4	[redacted]	[redacted]
3	[redacted]	[redacted]
2	[redacted]	[redacted]
1	[redacted]	[redacted]
Genotype		
c.622C>T homozygous	[redacted]	[redacted]
c.509-1G>C homozygous	[redacted]	[redacted]
c.622C>T and c.509-1G>C	[redacted]	[redacted]
c.622C>T, other	[redacted]	[redacted]
c.622C>T, missing	[redacted]	[redacted]
c.509-1G>C, other	[redacted]	[redacted]
Other homozygous	[redacted]	[redacted]

CTD = Common Technical Document; M-L = motor-language; SD = standard deviation.

Note: In this matched comparison, the matching factors were motor-language score and age difference of 12 months or less.

Source: Common Technical Document Summary section 2.7.3 Clinical Efficacy.[4]

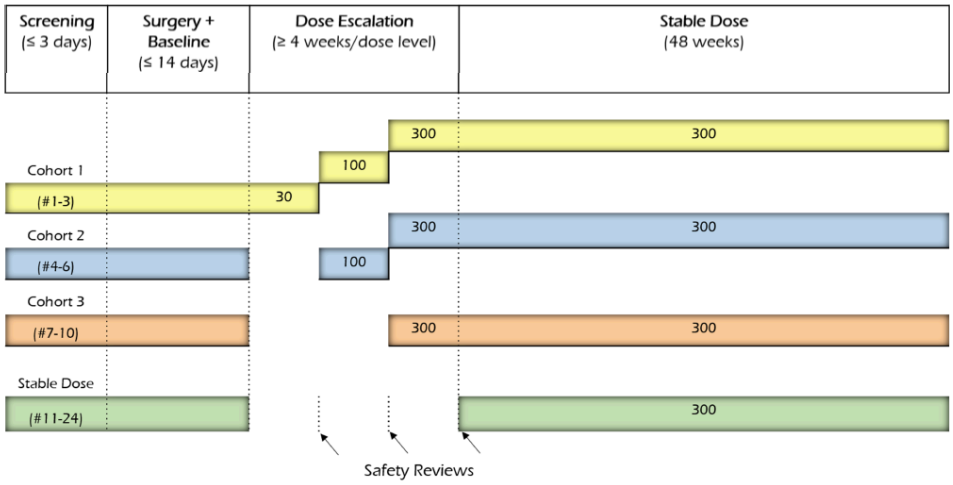
Interventions

In studies 201 and 202, surgical implantation of the ICV access device took place prior to drug administration. Patients were admitted to hospital for every infusion of cerliponase alfa. Prior to infusion, the drug is diluted and then administered by continuous infusion of 2.5 mL per hour over approximately four hours. The total volume administered was approximately 10 mL, and this was estimated by the manufacturer to be approximately 10% of the total cerebrospinal fluid volume for an average patient with CLN2. Patients were observed in an inpatient setting for 48 hours after each infusion. Study 201 was a first-in-human study and doses were based on pre-clinical canine and primate studies. Eligible patients were enrolled sequentially into three dose cohorts and later into a stable dose period.

During the dose escalation period, three dose cohorts were enrolled. The first cohort of three patients started cerliponase alfa 30 mg every 14 days. Once that dose was determined to be safe, an additional three patients were enrolled starting at 100 mg every 14 days, and the three patients from Cohort 1 were escalated to 100 mg. [redacted] Each patient participating in the dose escalation period received at least four weeks (two infusions) of treatment at a dose level before being moved on to the next higher-dose level.

One patient who enrolled in Cohort 3 withdrew from the study after a single dose. During the stable dose period, all patients from the three dose escalation cohorts remained at 300 mg every 14 days. In addition, 14 new patients were enrolled to start dosage at 300 mg every 14 days (Figure 2).

Figure 2: Dose Cohorts in Study 201



Note: Cohort 1 (N = 3), Cohort 2 (N = 6), Cohort 3 (N = 4), Stable Dose cohort (N = 14).

Source: Clinical Study Report 201.[14]

In Study 201, all patients participating in the stable dose period received 48 weeks of treatment (24 infusions) at 300 mg every 14 days. Patients participating in the dose escalation period received, in addition to the 48 weeks of stable dose period treatment, an additional 4 to 22 weeks of treatment (30 mg, 100 mg, and/or 300 mg) during their participation in one of the dose cohorts.

[redacted]

[redacted]

Outcomes

Primary Outcome

The primary efficacy outcome of studies 201 and 202 was the adapted CLN2 motor-language rating scale. The primary efficacy end point was a responder analysis during the 300 mg dose period based on the motor-language scale in the intention-to-treat (ITT) population. The original protocol specified that rate of decline of motor-language scores was the primary end point and this was modified after the study began (cited in the statistical analysis plan for Study 201, March 2016, Supplemental Appendices to Schulz et al.[10]). Response was defined as the absence of an unreversed two-point decline or score of zero in the adapted CLN2 score by week 48 (Study 201) or week 96 (Study 202). For this analysis, the baseline was defined as the start of the 300 mg infusion (referred to as the 300 mg baseline), and the end point was measured 48 weeks after this time. As a result of this definition, patients who were enrolled in the dose escalation cohorts received more than 48 weeks of treatment with cerliponase alfa at the time of observation for the primary end point.

An unreversed two-point decline was defined as a decline that had not returned to within one point of baseline at the time of the final motor-language assessment. For patients with a 300 mg baseline score of 1 (n = 1), a responder was defined as a patient who did not progress to a 0 in the treatment period. For individual motor and language domains, responders were identified as patients who did not lose a point on that domain at time of last assessment.

The rate of decline of the patient's score on the motor-language scale was also assessed and reported as a point loss per 48-week 300 mg dose period.

In Study 203, the manufacturer stated that the "efficacy endpoint of primary interest" was the motor-language scale.

The CLN2 scale used in studies 201, 202, and 203 used four components of the Hamburg (motor, language, vision, seizures) and four components of the Weill Cornell (gait, language, myoclonus, feeding) scoring systems. For the primary outcome analyses, the CLN2 scale excluded four of the domains in these two scales (Hamburg seizure and vision; Weill Cornell feeding and myclonus).[14] The remaining four domains were combined to create the adapted motor-language scoring system with a range of 0 to 6 points (0 to 3 for motor plus 0 to 3 for language [Table 6]). The remaining two domains of the Hamburge scale (vision and seizures) and the Weill Cornell scale (myoclonus and feeding) were collected as secondary and exploratory end points.

The adapted motor-language scale was administered at screening, baseline, and every eight weeks during the 300 mg stable dose period of the study. Investigators were encouraged to use the same rater for each patient throughout the study, but this practice was not followed for all patients.[11,14] Raters at the study sites were trained in the use of the scales, but it was not stated if the raters were the treating physicians or another member of the site staff.

In addition, there were several other related scales used in the 201, 202, and 203 studies, including a nine-point adapted CLN2 motor-language-vision scale and a 12-point adapted CLN2 motor-language-vision-seizure scale (total CLN2 scale). Higher scores represented better patient function (Table 6).

The motor-language CLN2 scale was used to assess patients in studies 201, 202, and 203, and the natural history cohort. The scoring criteria and definitions for each step were different in studies 201, 202, and 203 compared with the natural history study (Table 6). In addition, in the natural history cohort, the scale assessments were made both prospectively and retrospectively, whereas in studies 201, 202, and 203, motor-language assessments were prospectively done. A comparison of the scoring criteria is provided in Table 6.

Table 6: Motor-Language Scale Used in Studies 201, 202, and 203 and Comparison With the Definitions Used in the Natural History Studies

Score		
Motor	**Historical Cohort Study Definition**	**Studies 201, 202, and 203**
3	Walks normally	Grossly normal gait. No prominent ataxia, instability, shortened stride, pathologic falls
2	Frequent falls, clumsiness obvious	Independent gait, as defined by ability to walk without support for 10 autonomous steps. May have obvious instability, and may have intermittent falls
1	No unaided walking or crawling only	Requires external assistance to walk, and without support can only take fewer than 10 steps or crawl
0	Immobile, mostly bedridden	Can no longer independently walk or crawl
Language	**Historical Cohort Study Definition**	**Studies 201, 202, and 203**
3	Normal	Apparently normal language. Intelligible and grossly age-appropriate
2	Has become recognizable abnormal	Language has become recognizably abnormal: some intelligible words, but does not form sentences to convey concepts, requests, or needs
1	Hardly understandable	Hardly understandable. Few intelligible words in the context of unintelligible vocalizations
0	Unintelligible or no language	No intelligible words or vocalizations
Vision	**Historical Cohort Study Definition**	**Studies 201, 202, and 203**
3	Recognizes desirable object, grabs at it	Grossly normal. Appears to recognize multiple objects and reacts appropriately (reaches for a toy, etc.)
2	Grabbing for objects uncoordinated	Apparent difficulty seeing some objects. May be able to discern large objects, some TV, moving objects, but clearly is impaired
1	Reacts to light	Reacts only to light or threat
0	No reaction to visual stimuli	No reaction to light or (visual) threat
Seizure	**Historical Cohort Study Definition**	**Studies 201, 202, and 203**
3	No seizure in 3 months	No seizures in a 12 week period
2	1 to 2 seizures in 3 months	1 to 2 seizures in a 12 week period
1	1 seizure per month	3 seizures in a 12 week period (1 per 4 weeks)
0	> 1 seizure per month	> 3 seizures in a 12 week period (> 1 per 4 weeks)

Source: Clinical Study Report for Study 203;[16] FDA Statistical Review.[11]

Other Outcomes

In studies 201 and 202, the secondary outcome measures included magnetic resonance imaging (MRI) assessments, including whole brain volume, volume of cerebrospinal fluid, volume of total cortical grey matter, total white matter volume, and whole brain apparent diffusion coefficient. MRI was performed at week 1, 9, 25, and then approximately every 24 weeks.

Pediatric Quality of Life Inventory 4.0 Generic Core Scales and Family Impact Module

The Pediatric Quality of Life Inventory (PedsQL) Generic Core Scales are intended to be administered in both healthy and patient pediatric populations and, together with disease-specific modules, measure pediatric health-related quality of life (HRQoL).[20] Each item is scored on a five-point Likert scale (three-point scale for ages two to four) with each score linearly transformed to a scale of 0 to 100.[20] To generate domain and total scores, the transformed item scores are summed and then divided by the number of items (range 0 to 100).[20] The Generic Core Scales consist of the following scales: Physical Functioning (eight items), Emotional Functioning (five items), Social Functioning (five items), and School Functioning (five items), with higher scores indicating better HRQoL.[20] The Psychosocial Health Summary Score is the sum of the items in the Emotional, Social, and School Functioning Scales.[20] Higher scores indicate better HRQoL.[20] No evidence regarding a minimal clinically important difference (MCID) was identified for the CLN2 population.

The PedsQL Family Impact Module is a parent self-reporting instrument that is comprised of 36 items measuring parental functioning. The domains that are examined include Emotional Functioning (five items), Social Functioning (four items), Physical Functioning (six items), Communication (three items), Cognitive Functioning (five items), Worry (five items), along with two other scales that measure family functioning: Family Relationships (five items) and Daily Activities (three items).[21] A five-point Likert scale is used for all items with 0 meaning never a problem and 4 meaning always a problem. A 0 to 100 scale is used (after reverse scoring and linear transformation of items), with higher scores indicating better functioning.[21] In order to compute the domain scores, the items are summed and then divided by the number of items that were reported; however, if more than 50% of the items in a domain are missing, then the individual Domain Score is not calculated, per the developers instructions. The Family Impact Module Total Score is obtained by summing the 36-items and dividing by the number of items answered.[21] No evidence regarding an MCID was identified for the general chronically ill patient population or the CLN2 population.

Denver II Developmental Screening Test

The Denver II Developmental Screening Test (Denver II) is a revision and update of the Denver Developmental Screening Test, which aims to monitor and assess infant and preschool-aged children's development. It is comprised of 125 performance-based and parentally reported items (with 20 new items added on to the original Denver Developmental Screening Test) split between four functional groups: Personal-Social (e.g., smiling), Fine Motor-Adaptive (e.g., grasping, drawing), Language (e.g., combining words), and Gross Motor (e.g., walking).[11,14,22] Instead of using a scoring system, pass/fail/refusal scores are assigned to each item, which is then reinterpreted in terms of caution, delay, no opportunity, normal, or advanced performance based on the child's age. An abnormal overall test score is obtained when two or more delays are noted and a questionable score is obtained when one delay and/or two or more cautions are noted.[22] Only trained professionals are permitted to administer the Denver II.[22]

The sensitivity and specificity of the Denver II to identify children with developmental problems were assessed and high sensitivity and low specificity of the Denver II was also noted across all age groups examined (age range from 0 to 72 months).[22]

EuroQol 5-Dimensions 5-Levels Questionnaire

The EuroQol 5-Dimensions Questionnaire (EQ-5D) is a generic quality-of-life instrument developed by the EuroQol Group.[23] It may be applied to a wide range of health conditions

and treatments.[23] As a generic measure of HRQoL that can capture the net effect of treatment benefits and harms, the EQ-5D provides valuable information from a patient perspective. It consists of an EQ-5D descriptive system and the EQ visual analogue scale (VAS). The descriptive system is comprised of the following five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, each with five levels: a level 1 response represents "no problems," level 2 "slight problems," level 3 "moderate problems," level 4 "severe problems," and level 5 "extreme problems" or "unable to perform," which is the worst response in the dimension. A score of 0 represents the health state "dead" and 1.0 reflects "perfect health."

The EQ VAS records the respondent's self-rated health on a vertical VAS where the end points are labelled 0 ("the worst health you can imagine") and 100 ("the best health you can imagine"). The respondents are asked to mark an X on the point of the VAS that best represents their health on that day. The EQ-5D index and VAS scores can be summarized and analyzed as continuous data.[23,24]

MCID estimates for the index score in the general Canadian population were generated by simulating the effects of single-level transitions in each dimension.[25] The results yielded MCIDs with a summarized mean of 0.056 (standard deviation [SD] of 0.011), and a summarized median of 0.056 (interquartile range, 0.049 to 0.063).[25] There was no evidence of validity, reliability, responsiveness, or an MCID specific to CLN2 disease or any neuronal ceroid lipofuscinosis disease identified in the literature for the EQ-5D-5 Levels (5L).

CLN2 Quality of Life Questionnaire

The Neuronal Ceroid Lipofuscinosis Type 2 Quality of Life (CLN2QL) Questionnaire is a CLN2 disease-specific questionnaire designed by Biomarin to examine the HRQoL in patients with CLN2 disease undergoing treatment with Brineura.[11] Family feedback from two focus groups (one in Europe and one in the US) was ascertained in order to help with the development of this questionnaire.[14] The CLN2QL Questionnaire is comprised of 28 items that are part of six domains related to CLN2 disease. The items are partitioned as follows:

- seizures (six items)
- feeding (with no gastrostomy tube) (four items)
- feeding (with a gastrostomy tube) (three items)
- sleep (five items)
- behaviour (six items)
- daily activities (four items).[11]

Harms

Treatment emergent adverse events (AEs) were defined as those that newly appeared, increased in frequency, or worsened in severity following ICV surgery, whether or not the patient had yet received any study drug. In studies 201, 202, and 203, [redacted]

AEs of special interest to the manufacturer in studies 201 and 202 and [redacted]

Statistical Analysis

Analyses Without Comparison Group

In Study 201, to estimate sample size requirements, the primary end point was assumed to be the within-patient estimate of slope reflecting the rate of decline in the seven-point (0 to 6) Hamburg CLN2 rating scale over time. Based on a review of 30 patients from a natural history data set, it was assumed that, untreated, patients decline a mean of [redacted]. If treatment resulted in a mean rate of decline of 0.5, with an SD of 1.8, then 18 evaluable patients would be required to achieve 90% power to reject the null hypothesis, assuming a two-sided test with a significance level of alpha equals 0.05. This was increased to 22 patients to allow for a discontinuation rate of approximately 20% (retrieved from the Schulz et al. supplemental appendix).[10]

The statistical analysis plan for Study 201 states that for the primary efficacy analysis (responder) of CLN2 ratings, missing data would be handled by the Kaplan-Meier method. For the analysis of rates, missing data were not expected to be an issue given that [redacted]

There were some subgroup analyses performed [redacted] It was not clear if these analyses were specified a priori.

Analyses With Comparison Group

There are several historical cohorts referred to in the main publication for studies 201 and 202,[10] the FDA Statistical Review,[11] and the CTD.[4] CADTH requested the manufacturer's opinion on which of the matched analyses achieved the best minimization of bias. Ideally, a response would have included evidence that a particular method of matching patients was defined a priori, and subsequently applied, or a detailed table summarizing baseline

characteristics and potential baseline prognostic factors of this best match historical cohort, and the corresponding population from studies 190-201 and 202. The manufacturer stated that the criteria for matching historical controls to patients in studies 201 and 202 was specified prior to performing efficacy analysis;[6] however, they did not explain why the main publication matched analysis (N = 17)[10] differed from the matched analysis in the CTD ([redacted]).[4] The manufacturer considered the [redacted] historical control group to be the most appropriate matched analysis but gave no explanation why an alternative matched control group of N = 17 was used by Schulz et al. in the main publication for studies 201 and 202.[6]

The Clinical Study Report for Study 201 states, "These study data are to be compared with historical natural history data from a registry initiated at University Medical Center Hamburg-Eppendorf in Hamburg, Germany. Within the present report, the rate of decline in the [motor-language] scale score observed within the study is compared to a value estimated from those historical data. A more extensive, and more direct, comparison will be presented in a separate report." CADTH requested this "separate report" from the manufacturer. The manufacturer responded by providing some data from [redacted] is the "separate report" referred to in the Clinical Study Report for Study 201.

Analysis Populations

Most efficacy analyses with the uncontrolled data were performed on the ITT population (n = 23), which included all patients who received cerliponase alfa, and reported any efficacy results, excluding one patient who withdrew from the study after a single infusion of the study drug during the dose escalation period. This is not a true ITT population given that it did not contain all patients who were assigned to treatment. The efficacy population (n = 21) was the ITT population excluding two patients who enrolled with a baseline motor-language scale score of 6 (the maximum score) and who showed no decline in their motor-language scale scores at the end of 48 weeks of treatment in the 300 mg dose period. The enrolled population (n = 24) was all patients who provided informed consent. The safety population (n = 24) included all patients who had an ICV access device implanted.

Patient Disposition

Twenty-four patients were enrolled in Study 201 after screening (Table 7). The number of screened patients was not reported. One patient who was enrolled into the dose escalation phase was withdrawn from Study 201 after one dose was administered. Twenty-three patients completed Study 201 and all of them enrolled in Study 202. As of November 2016, the mean (SD) duration on treatment with cerliponase alfa in studies 201 and 202 was 117 (33) weeks for all doses (range: 0.1 to 161) and 115 (30) weeks during the 300 mg dose period (range: 0.1 to 145). At the time of data cut-off for efficacy analysis (November 2016), 23 patients remained in Study 202. [redacted]

In Study 203, four patients were enrolled at the time of data cut-off for efficacy (November 2016). [redacted]

[redacted]

Table 7: Patient Disposition in Studies 201 and 202

	Cohort 1	Cohort 2	Cohort 3	Stable Dose Cohort	Total
Enrolled in Study 201	3	3	4	14	24
Completed Study 201	3	3	3	14	23
Transferred from Study 201 to Study 202	3	3	3	14	23
Completed at least 96 weeks of treatment with 300 mg cerliponase alfa in Studies 201 and 202	3	3	3	14	23

Note: Data in table are number of patients.

Source: Clinical Study Report for studies 201 and 202.[14,15]

Exposure to Study Treatments

[redacted]Patients received 99% of the planned dose (measured in mg) over both the entire dosage period and the 300 mg dose period. [redacted]

Critical Appraisal

Internal Validity

There was no concurrent control group for studies 201, 202, or 203. There were several historical cohorts mentioned in the reports for studies 201 and 202 (Schulz et al. and the CTD). Schulz et al. and the FDA Statistical Review cite historical control groups that were unmatched (N = 42) and matched (N = 17).[10,11] The manufacturer's CTD reported a matched historical control group ([redacted]).[4] The [redacted] historical control group described in the CTD was matched to the treated population using two factors: motor-language score and age difference of 12 months or less. [4] The N = 17 historical control group described by Schulz et al. was matched to the treated population using three factors: exact motor-language score, age within three months, and genotype (equal number of common alleles).[10] The manufacturer considered the [redacted] historical control group to be the most

appropriate matched analysis but gave no explanation why an alternative matched control group of N = 17 was used in the main publication for studies 201 and 202.[6]

While there may have been challenges to implementing a prospective, untreated control group within studies 201 and 202, there are a number of limitations associated with the use of historical control groups in the matched and unmatched comparisons:

- In the matched analyses, the number of criteria used for matching was two or three; motor-language score and age were balanced in the analysis, which used two criteria ([redacted]Table 5). Matching on baseline prognostic factors is a technique that would be expected to result in a reduction of bias, but would not eliminate bias. There are many other prognostic factors, known and unknown, that may have been unbalanced between the natural history cohort and the treated patients in studies 201 and 202.
- There was very little information available on the baseline characteristics and demographics of the historical cohorts. The data that were available suggest that the historical cohort may not be comparable with the treated patients. [redacted] The standards of care were different during those decades relative to current care. [redacted], there were differences in the severity of disease at diagnosis, the pharmacologic options for seizure control were far fewer, and the threshold for gastronomy tube insertion was different, relative to current treatment scenarios, according to a clinical expert consulted by CADTH.
- Several important baseline characteristics were not available or not reported for the historical control group and/or the treatment group. There were no data provided for either group on TPP enzyme activity, which is an important indicator of disease activity. In the historical control groups, there was no information for age of the child at time of disease onset. These gaps in the data hinder the assessment of the comparability of the treatment group and the historical control groups.
- The manufacturer did not provide evidence that the methods to select an appropriate historical control group were appropriate, or that the statistical analysis plan for comparing the treated patients with historical controls was pre-specified.

The scales used to assess symptom severity (motor, language, seizure, vision) applied the same numerical gradations in the treated patients and the historical control groups, but the definitions corresponding to the severity levels (0, 1, 2, 3) were not the same (Table 6). This was highlighted by the FDA reviewers, who stated that there was insufficient comparability in scale definitions used in treated patients compared with the definitions used for the historical controls.[11] The FDA reviewers stated that the definitions for the motor scale were more similar across treatment and control groups, but the definitions for the language scale were less comparable. The indication granted by the FDA states that the cerliponase alfa is indicated to “slow the loss of ambulation” but does not mention impact on language, vision, or seizures. In addition, the scale assessments were made prospectively in the treated cohorts, but many assessments were made retrospectively in the historical cohorts via parental recall interview and medical chart review.[11] Retrospective application of symptom scales may not provide an accurate depiction of the severity of these historical control patients. In addition, in the historical cohort described by the FDA Statistical Review (N = 17), data were imputed to planned visit because the historical data did not correlate with the time points used in studies 201 and 202. Retrospective assessment and imputation of data increase uncertainty in the results of the historical cohorts.

The adapted motor-language scale was used as the primary outcome in studies 201, 202, and 203. An unreversed decreased of two points or more was selected for the responder analyses. In some patients, a decrease of two points, or even one point, could be considered clinically significant, but this assessment is dependent upon the baseline value and the subscore changes. For example, a change in motor score from 2 to 1 point might not be considered as significant a deterioration to the patient as a change in language score from 2 to 1 point. According to Wyrwich et al.,[26] a measurable and clinically important difference is one point on the motor or language rating scale. However, no information on how they obtained that number was identified; therefore, the MCID of the motor and language rating scales remains uncertain.

An important omission from this submission is the lack of data on AEs and serious AEs (SAEs) in the historical cohorts. Without control group data on AEs, it is difficult to assess the overall risk of harms associated with cerliponase alfa.

The quality of life data and the data on family and caregiver impact only included change from baseline and there were no comparisons with a historical cohort. The SDs for the mean score change were larger than the score change values, indicating high variability.

External Validity

The baseline characteristics that were reported for patients treated with cerliponase alfa in studies 201, 202, and 203 were similar to patients who would be eligible to receive cerliponase alfa in Canada, according to the clinical experts for this CDR review. However, there was important information not reported regarding TPP enzyme activity at baseline and mutation status. [redacted]

The maintenance dosage of 300 mg every two weeks used in the trials is the same dose proposed by the manufacturer for use in Canada.

The study populations were small but the planned follow-up time for patients in the ongoing Study 202 [redacted]. This long follow-up time appropriately reflects the long duration of therapy that is expected in patients who begin treatment with cerliponase alfa; however, patients may take cerliponase alfa for longer time periods and there are no efficacy data yet available beyond a follow-up of approximately [redacted].

The study was conducted in centres that likely had high familiarity with using intraventricular catheters and this may not represent real-world experience, according to a clinical expert consulted by CADTH.

Efficacy

Only those efficacy outcomes identified in the review protocol (Table 2) are reported in the following.

No data were available for the following outcomes: caregiver burden, improvement of swallowing function to allow removal of feeding tube, return of ability to walk, and reduction of seizure incidence to null.

There were very few efficacy data available for Study 203 because of the short time frame of drug exposure and the low number of patients ([redacted]).

Survival

There were no deaths reported in patients treated with cerliponase alfa during studies 201, 202, or 203 as of the safety cut-off date of April 2018.

There were no mortality data provided for the historical controls described in the [redacted] and by Schulz et al. (N = 42, N = 17).[4,10]

Symptom Scales

Impact on Symptoms

Motor and Language and Total Neuronal Ceroid Lipofuscinosis Type 2 Assessment (Analyses Without a Control Group)

The motor-language scoring assessment had a total range of 0 to 6 points. There were several types of analyses presented for the motor-language scores: responder analyses (Table 8 and Figure 3), mean rate of decline, numbers of patients who improved/worsened (Table 9), and mean score change at different time points (Table 11).

Of patients taking cerliponase alfa, 20 out of 23(87%) had an absence of an unreversed two-point decline or score of 0 in the motor-language score after 48 weeks and also at 96 weeks (Table 8). At the 96-week time point, 12 out of 23(52%) patients met the stricter definition of response, an absence of an unreversed one-point decline or score of 0 in the motor-language score.

The mean rate of decline of motor-language score points per 48-week period was 0.27(SD of 0.35) for the ITT population (N = 23) and [redacted].[14,15]

[redacted], the addition of seizures to the motor-language-vision score resulted in less worsening (i.e., addition of seizures resulted in a more favourable score at week 97, relative to the score without seizures).

The manufacturer provided additional data regarding the seizure domain changes in the CLN2 scale (Table 10) in its comments on the CADTH clinical review. These data suggested that seizure scores worsened for four patients, did not change for seven patients, and improved for 12 patients at week 96 relative to baseline.

Table 8: Response Based on the Motor and/or Language Score During the 300 mg Dose Period in Studies 201 and 202

Response Criteria	Studies 201 and 202 N = 23
Absence of an unreversed **2-point** decline or score of 0 in the motor-language score at 48 weeks[a]	20 (87)
Absence of an unreversed **1-point** decline or score of 0 in the motor-language score at 48 weeks	15 (65)
Absence of an unreversed **1-point** decline or score of 0 in the motor score at 48 weeks	16 (70)
Absence of an unreversed **1-point** decline or score of 0 in the language score at 48 weeks	[redacted]
Absence of an unreversed **2-point** decline or score of 0 in the motor-language score at 96 weeks	[redacted]
Absence of an unreversed **1-point** decline or score of 0 in the motor-language score at 96 weeks	[redacted]
Absence of an unreversed **1-point** decline or score of 0 in the motor score at 96 weeks	[redacted]
Absence of an unreversed **1-point** decline or score of 0 in the language score at 96 weeks	[redacted]

Note: Data are n (%); total population for studies 201 and 202 in this table excludes the patient who withdrew from the trial after the first dose.

[a] Primary outcome of Study 201.

Source: Clinical Study Reports for studies 201 and 202.[14,15]

Figure 3: Studies 201 and 202 — Kaplan-Meier Time to First Unreversed Two-Point Decline or Score of 0 in Motor and Language

Figure removed based on manufacturer's redaction request.

Note: An unreversed two-point decline is any decline of two points or more that had not reversed to a one-point decline (or better) at the last recorded observation. An unreversed score of 0 is a decline of 0 that had not reverted back to greater than 0 at last recorded observation. These data use the day of first 300 mg infusion in Study 201 as "Day 1."

Source: Clinical Study Report for Study 202.[15]

Table 9: Improvement/Worsening of Scores for Studies 201 and 202 and Study 203

	Studies 201 and 202 [redacted]			Study 203 [redacted]		
Score change	**Motor-Language Score, n (%)**	**Motor Score, n (%)**	**Language Score, n (%)**	**Motor-Language Score, n (%)**	**Motor Score, n (%)**	**Language Score, n (%)**
3 (improvement)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
0 (no change)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
–1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
–2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
–3 (decline)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Note: Change scores are from beginning of 300 mg dose period to last assessment for studies 201 and 202. Change scores for Study 203 are from baseline to last assessment. One child in Study 203 was not assessable for language because of autism.

Source: Clinical Study Reports for studies 201 and 202,[14,15] and Study 203.[16]

Table 10: Ordinal Analysis of Change From Baseline in Neuronal Ceroid Lipofuscinosis Type 2 Clinical Rating Scale Seizure Domain Scores in Brineura Clinical Studies (190-201and 202)

Seizure CLN2 Score Change From Baseline	Total Weeks of Treatment[a,b]			
	Week 24 N = 23 n	Week 48 N = 23 n	Week 72 N = 23 n	Week 96 N = 23 n
+3 (Improvement)	[redacted]	[redacted]	[redacted]	[redacted]
+2	[redacted]	[redacted]	[redacted]	[redacted]
+1	[redacted]	[redacted]	[redacted]	[redacted]
0 (No change)	[redacted]	[redacted]	[redacted]	[redacted]
–1	[redacted]	[redacted]	[redacted]	[redacted]
–2	[redacted]	[redacted]	[redacted]	[redacted]
–3 (Worsening)	[redacted]	[redacted]	[redacted]	[redacted]

CLN2 = neuronal ceroid lipofuscinosis type 2.
[a] Weeks of treatment at 300 mg dose.
[b] Includes patients receiving more than one dose of treatment.
Source: Manufacturer's Comments on the CADTH Clinical Review.[27]

Table 11: Mean Scores Over Time From Beginning of 300 mg Dose for Studies 201 and 202 (N = 23)

Time Point	Motor-Language Score Mean (SD)	Motor-Language-Vision Score, Mean (SD)	Total CLN2 Scale: Motor-Language-Vision-Seizure Score, Mean (SD)
Total points possible	0 to 6	0 to 9	0 to 12
Baseline	3.5 (1.2)	6.3 (1.3)	8.0 (1.8)
Week 49	3.0 (1.3)	5.7 (1.6)	7.8 (2.1)
Week 97	[redacted]	[redacted]	[redacted]
Last observation	[redacted]	[redacted]	[redacted]
Change in mean (SD) from 300 mg baseline to last observation	[redacted]	[redacted]	[redacted]
Change in median from 300 mg baseline to last observation	[redacted]	[redacted]	[redacted]

CLN2 = neuronal ceroid lipofuscinosis type 2; SD = standard deviation.
Source: Common Technical Document section 2.7.3.[4]

Denver II Developmental Screening Test

Over the entire studies 201 and 202 dosage period, [redacted]

Health-Related Quality of Life

HRQoL was assessed using the PedsQL (including both a Parent Report for Toddlers and a Family Impact Module) and a CLN2 disease-based HRQoL instrument (Table 12). Scores on these instruments range from 0 to 100, with higher scores relating to better function. In addition, the EQ-5D-5L instrument was assessed in Study 202 only.

Mean scores increased (improved) for the CLN2 Disease-Based Quality of Life Instrument. Scores decreased (worsened) for the PedsQL Parent Report for Toddlers, the PedsQL Family Impact Module (which measures impact on caregiver and family), and the EQ-5D-5L EQ VAS. The standard deviations for the mean score change were larger than the score change values, indicating high variability.

For the EQ-5D-5L domains, of the 23 patients with data at baseline and week 97, [redacted]

Table 12: Summary of Quality of Life Data

Scale	Mean Baseline Score (SD) N = 23	Mean Change From Baseline to Week 97 (SD)
PedsQL Parent Report for Toddlers	60.7 (12.8)	–5.7(18.9); n = 21
PedsQL Family Impact Module	61.4 (14.3)	–1.1(19.6); n = 21
CLN2 Disease-Based QoL Instrument	74.2 (13.8)	+3.1(14.4); n = 21
EQ-5D-5L EQ VAS (baseline to week 49)[a]	[redacted]	[redacted]

CLN2QL = Neuronal Ceroid Lipofuscinosis Type 2; EQ-5D-5L = EuroQol 5-Dimensions 5-Levels Questionnaire; PedsQL = Pediatric Quality of Life; QoL = quality of life ; SD = standard deviation ; VAS = visual analogue scale.

Note: The PedsQL Family Impact Module describes impact on family and caregivers.

[a] Baseline for EQ-5D-5L is the start of Study 202, all other baselines are the start of Study 201.

Source: Clinical Study Reports for studies 201 and 202.[14,15]

The manufacturer provided some information regarding the subscale change from baseline in the CLN2 disease-based HRQoL instrument. These data indicated improvements from baseline for the seizure domain; deterioration in the feeding/G-tube domain; and no change in the sleep, behaviour, and activity domains.

Magnetic Resonance Imaging Changes

There were decreases in whole brain volume (–4.3%), cortical grey matter (–12.4%), and white matter (–2.7%) over the course of studies 201 and 202. There was a corresponding increase in cerebrospinal fluid of 6.4% (Table 13).

Table 13: Changes in Magnetic Resonance Imaging Parameters in Studies 201 and 202

MRI Parameter	Baseline, Mean (SD)	Last Observation, Mean (SD)	Change From Baseline, cm^3	% Change From Baseline
Whole brain volume, cm^3	1,105 (189)	[redacted]	[redacted]	[redacted]
Cortical grey matter volume, cm^3	452 (88)	[redacted]	[redacted]	[redacted]
White matter volume, cm^3	342 (55)	[redacted]	[redacted]	[redacted]
Cerebrospinal fluid volume, cm^3	310 (72)	[redacted]	[redacted]	[redacted]

MRI = magnetic resonance imaging; NR = not reported; SD = standard deviation.

Note: Changes are from 300 mg baseline to last observation. The last observation occurred at least 96 weeks after the start of the 300 mg stable dose period.

Source: Clinical Study Reports for studies 201 and 202,[14,15] Schulz et al.,[10] Common Technical Document section 2.7.3.[4]

Motor and Language Assessment (With Historical Control Comparisons)

This section describes available data that compares the treated population in studies 201 and 202 with historical control populations. There were several historical control populations referenced in the study reports and publications with N = 42, N = 17, and [redacted]. The manufacturer stated that the matched analysis using a historical control of [redacted] was the most appropriate comparison, but did not explain why a matched comparison with N = 17 was used in the main publication of these studies.

As reported by Schulz et al., the median time until a two-point decline in the motor-language scale score among the unmatched historical controls (N = 42) was 345 days (49.3 weeks), but was not reached among the treated patients in studies 201 and 202; 9% of the treated patients had a decline of two points at 345 days (see Figure 4A).[10] The unadjusted mean rate of decline in the motor-language scale score per 48-week period was 0.27 ± 0.35 points among the 23 treated patients as compared with 2.12 ± 0.98 among the 42 unmatched historical controls, a difference of 1.85 ± 0.21 points (95% confidence interval, 1.51 to 2.18; $P < 0.001$). [10] The hazard ratio comparing the treated population in studies 201 and 202 with the unmatched historical control (N = 42) was 0.08 (95% confidence interval, 0.02 to 0.23) favouring treated patients for the time to unreversed two-point or greater decline in the motor-language score (see Table 14 and Figure 4A). The hazard ratio comparing the treated population in studies 201 and 202 with the matched historical control ([redacted]) for the same outcome was [redacted] favouring treated patients (see Table 14).

Schulz et al. also reported on a matched analysis based on three factors (baseline motor-language score, age within three months, and genotype [equal number of common alleles]), and this resulted in 17 treated patients for which matching was achieved (Table 15). The mean decrease from baseline in motor-language scale at week 96 was –0.20 (SD of 0.67) in the treated population, versus –1.90 (SD of 1.23) in the historical control group. [redacted]

Table 14: Response Data at Week 96 Based on the Motor and/or Language Score in Studies 201 and 202 — Analyses With Control Groups

Source	Schulz et al. Unmatched Analysis			FDA Matched Analysis			Common Technical Document Matched Analysis		
Response outcome	**Studies 201 and 202 Treated N = 23**	**Historical Control N = 42**	**HR (95% CI)**	**Studies 201 and 202 Treated N = 17**	**Historical Control N = 17**	**OR (95% CI)**	**Studies 201 and 202 Treated** [redacted]	**Study 901 Historical Control** [redacted]	**HR (95% CI)**
Absence of an unreversed **2-point** decline or score of 0 in the **motor-language** score at last observation (≥ 96 weeks)	NR	NR	0.08 (0.02 to 0.23) *P* < 0.001	NR	NR	NR	[redacted]	[redacted]	[redacted]
Absence of an unreversed 1**-point** decline or score of 0 in the **motor-language** score at last observation (≥ 96 weeks)	NR	NR	NR	NR	NR	NR	[redacted]	[redacted]	[redacted]
Absence of an unreversed **2-point** decline or score of 0 in the **motor** score at last observation (≥ 96 weeks)	NR	NR	0.04 (0.00 to 0.29) *P* = 0.002	16 (94)	6 (35)	0.09 (0.002 to 0.63)	[redacted]	[redacted]	[redacted]
Absence of an unreversed **2-point** decline or score of 0 in the **language** score at last observation (≥ 96 weeks)	NR	NR	0.15 (0.04 to 0.52) *P* = 0.003	NR	NR	NR	[redacted]	[redacted]	[redacted]

CI = confidence interval; HR = hazard ratio; NR = not reported; OR = odds ratio.

Note: Matching in Schulz et al. was done on three factors (baseline motor-language score, age within three months, and genotype [equal number of common alleles]). Matching in the Common Technical Document analysis was done on two factors (baseline motor-language score and age). It is not clear if the 17 patients in the FDA analysis are the same as the 17 patients reported by Schulz et al.

Source: Clinical Study Reports for studies 201 and 202;[14,15] Schulz et al.;[10] Common Technical Document section 2.7.3.[4]

Table 15: Mean Change in Symptom Scores — Analysis With Matched Control Group

	Schulz et al. Matched Analysis	
Outcome	**Studies 201 and 202 Treated Population N = 17**	**Historical Control N = 17**
Mean change from baseline in motor-language scale at week 48 (SD)	–0.20 (0.67)	–1.90 (1.23)
Mean change from baseline in motor-language scale at week 96 (SD)	–0.50 (0.71)	–2.80 (1.10)
Mean change from baseline in motor-language-vision-seizure scale at week 48	+0.30 (1.70)	–2.80 (2.04)
Mean change from baseline in motor-language-vision-seizure scale at week 96	+0.40 (2.08)	–4.30 (2.26)

SD = standard deviation.

Note: Baseline values and statistical test results for these data were not reported by Schulz et al.

Source: Schulz et al.[10]

Harms

Only those harms identified in the review protocol are reported below (see section 2.2.1, Protocol).

The following harms data are from the treated populations. There were no AE data available for any of the historical cohorts.

Adverse Events

In studies 202 and 203, all patients reported at least one AE (Table 16). The most frequently reported AEs included pyrexia (71%), vomiting (63%), hypersensitivity (63%), seizure/epilepsy (≥ 58%), upper respiratory tract infection (54%), nasopharyngitis (42%), rhinitis (42%), and constipation (33%).

[redacted]

Table 16: Adverse Events

	Studies 201 and 202 N = 24
Patients with ≥ 1 adverse event	24 (100)
Adverse Event	
Pyrexia	17 (71)
Vomiting	15 (63)
Hypersensitivity	15 (63)
Seizure	14 (58)
URTI	13 (54)
Epilepsy	12 (50)
Generalized tonic-clonic seizure	12 (50)
Nasopharyngitis	10 (42)
Rhinitis	10 (42)
Constipation	8 (33)
Viral infection	8 (33)

	Studies 201 and 202 N = 24
Myoclonus	8 (33)
Cough	8 (33)
Diarrhea	7 (29)
Dysphagia	7 (29)
Gait disturbance	7 (29)
Gastroenteritis	7 (29)
Pharyngitis	7 (29)
Fall	7 (29)
Tremor	7 (29)
Petit mal epilepsy	7 (29)
Viral URTI	6 (25)
Tonsilitis	5 (21)
Dystonia	5 (21)
Extensor plantar response	5 (21)
Needle issue	5 (21)
Insomnia	5 (21)

URTI = upper respiratory tract infection.

Note: This table contains specified adverse events occurring in 20% or more of patients in studies 201 and 202 up to November 2016.

Source: Clinical Study Report 202.[15]

Serious Adverse Events

In studies 201 and 202, a total of 56 events were reported for 20 patients (Table 17). Of these, 13 patients (54%) experienced SAEs that were rated Grade 3 or Grade 4 for severity. The most common SAEs included hypersensitivity (29%), upper respiratory tract infection (17%), and gastroenteritis (13%).

[redacted]

Table 17: Serious Adverse Events

Category	Studies 201 and 202 N = 24
	Incidence, n(%)
Any SAE	20 (83)
Infections and infestations	13 (54)
Nervous system disorders	8 (33)
Immune system disorders	7 (29)
Hypersensitivity	7 (29)
URTI	4 (17)
Gastroenteritis	3 (13)
Bacterial pharyngitis	2 (8)
Pyrexia	2 (8)
Propionibacterium infection	1 (4)
Infusion related reaction	1 (4)
Patients with (severity grade) :	
Grade 1 SAE	3 (13)
Grade 2 SAE	4 (17)

Category	Studies 201 and 202 N = 24
Grade 3 SAE	12 (50)
Grade 4 SAE	1 (4)
SAE considered related to the study drug by investigator	NR

NR = not reported; SAE = serious adverse event; URTI = upper respiratory tract infection.

Note: This table contains specified SAEs occurring in more than one patient in studies 201 and 202 up to the data cut-off date.

Source: Clinical Study Report 202.[15]

Withdrawals Due to Adverse Events

There were no withdrawals due to AEs from studies 201, 202, or 203 as of the most recent safety analysis in April 2018.

Mortality

[redacted]

Notable Harms

In studies 201 and 202, hypersensitivity events were reported 37 times in 15 patients (63%). These included stomatitis (one event in one patient), seasonal allergy (two events in one patient), conjunctivitis (five events in four patients), dermatitis (five events in four patients), rash (seven events in two patients), and urticarial (two events in two patients).

In Study 203, [redacted].

In studies 201 and 202, device-related AEs were reported 34 times in 12 patients (50%). These included vomiting (one event in one patient), device complication (one event in one patient), pain on administration (one event in one patient), Propionibacterium infection (two events in one patient), device-related infection (one event in one patient), post procedural hematoma (one event in one patient), wound complication (one event in one patient), cerebrospinal fluid test abnormal (one event in one patient), systemic vascular resistance abnormal (one event in one patient), fluid retention (one event in one patient), pleocytosis (four events in four patients), needle issue (six events in five patients), device leakage (eight events in two patients), device malfunction (one event in one patient), device connection issue (one event in one patient), device deployment issue (one event in one patient), device infusion issue (one event in one patient), and device occlusion (one event in one patient).

[redacted]

Other Harms Information; From the Development Safety Update Report (Until April 2018)

The manufacturer's Development Safety Update Report summarized cumulative AE data across all clinical studies up to April 2018.[17] This report provided the following safety information on SAEs and notable harms. There were no denominator data provided.

[redacted]

[redacted]

Two ICV access devices were removed from two patients in Study 202 due to infection after more than four years of device use.

[redacted]

[redacted]

Discussion

Summary of Available Evidence

Three studies met the inclusion criteria for this review. Study 201 and its extension, Study 202, were phase I/II, multi-centre, open-label, single-arm studies of cerliponase alfa in 24 patients with CLN2. The age range at study entry was between 3.0 and 8.0 years. Doses were escalated during Study 201 until a stable dose of 300 mg every two weeks was reached. The primary objectives of these trials were to evaluate the safety of cerliponase alfa and the impact on motor-language scores. Post hoc comparisons were made between the treated patients in studies 201 and 202 and historical controls. Study 203 had a similar design as studies 201 and 202 and administered a stable dose of 300 mg cerliponase alfa every two weeks to siblings of the children who had enrolled in studies 201 and 202. Study 202 is ongoing (N = 23) and Study 203 is ongoing ([redacted]). The most recent efficacy data from these studies is from November 2016. Key limitations in both trials were the small sample sizes, lack of concurrent control groups, and inconsistencies in the scales used to compare treated patients with historical controls.

Interpretation of Results

Efficacy

Compared with the historical control data, patients receiving cerliponase alfa appear to experience a two-point or greater decline in the CLN2 motor-language scale score at a slower rate. However, comparison with a historical control group cannot produce results that are as reliable as those from a randomized study. The investigators attempted to overcome this limitation by matching the treated patients with historical control patients using several known prognostic factors (e.g., age, genotype, and baseline motor-language score). These methods would be expected to create populations that are more similar than unmatched populations, but the amount of residual confounding and the degree to which the treated and untreated patients differ remains unknown. There are other drawbacks of using historical controls, including the difficulties in assessing motor, language, seizure, or vision symptoms retrospectively from patient records.

Notwithstanding the uncertainties described regarding the methods for selection of the historical control group, the magnitude of treatment effect was large and statistically significant for all comparisons made between treated patients versus historical controls for the motor-language analyses. A large treatment effect is an important consideration for assessing the likelihood that the observed effect of cerliponase alfa on motor-language scores is attributable to the drug treatment. The hazard ratios are statistically significant for the motor-language response results comparing treated patients with unmatched or matched controls (see Table 14). The analysis of mean change from baseline in motor-language scores showed differences of approximately two points on the motor-language scale at week 96 for treated patients versus a matched historical control (Table 15; statistical significance is unknown). One point could represent a clinically meaningful difference for some patients, according to a clinical expert consulted by CADTH for this review, but the MCID has not been definitively established (Appendix 5).

The clinical experts for this review stated that there is no reason to believe that cerliponase alfa can impact visual deterioration because no data were presented for how cerliponase alfa impacts storage and visual deterioration in the eyes. Differences were observed in the

CLN2 total score, which included vision and seizure components for treated patients compared with historical controls, but these differences were likely related to the impact on symptoms other than vision (Figure 5B). There were no comparisons between treated and untreated patients for HRQoL, caregiver burden, developmental assessment, or brain MRI changes. According to one of the clinical experts consulted by CADTH, the decreases observed in whole brain volume, cortical grey matter, and white matter observed over the course of treatment with cerliponase alfa are consistent with worsening disease, but these were not measured in the control group and so the relative effects of cerliponase alfa on brain changes is not known. The clinical significance of the MRI changes are unclear, according to a clinical expert consulted by CADTH.

The manufacturer provided additional data regarding the seizure domain changes in the CLN2 scale in its comments on the CADTH clinical review. Approximately half of the patients had improved seizure scores at week 96 relative to baseline. This analysis was not specified as a primary or secondary outcome of studies 201 and 202 and these data did not appear in the main publication of this study. The FDA indication makes reference to motor benefits, but not seizure benefits. Reviewers for the National Institute for Health and Care Excellence stated that "improvement in the seizure domain does not necessarily reflect a halt in the deterioration of seizures. The seizure domain of the Hamburg reflects only the frequency of tonic-clonic seizures, and does not take into account the activity of other movement disorders."[13] There were no comparative analyses provided for the seizure data, either through a matched or an unmatched analysis. The CLN2 scale has four domains (motor, language, vision, and seizure). Of these four domains, only seizures have pharmacologic options that can be used to alleviate these symptoms. It is reasonable to expect that seizures could improve during the study in some patients and that this would be attributable to the anti-seizure medications prescribed to the children. Most children were taking anti-seizure medications during the study. Therefore, it is not clear that the improvements in seizure suggested by these data are attributable to cerliponase.

Based on input from the clinical experts consulted by CADTH, outcomes of interest in this review included improvement of swallowing function to allow removal of feeding tube, return of ability to walk, and reduction of seizure incidence to null. There were no data on these outcomes from the cerliponase alfa trials. Study 202 is ongoing; however, the available efficacy data are from November 2016, when the all patients in the efficacy population had received at least 96 weeks of treatment. Longer exposure is needed to better understand the long-term effects on disease progression and survival and how this compares with an untreated patient population. It is not known if maintenance of motor and language functions correlates with improved survival in patients with CLN2 disease.

The Health Canada–approved indication is for the "treatment of CLN2 disease" with no qualification or comment on the symptomatology. In contrast, the indication for cerliponase alfa in the US is "to slow the loss of ambulation in symptomatic pediatric patients 3 years of age and older with late infantile CLN2." This indication suggests that the drug slows the loss of motor ability but does not acknowledge benefit in other symptoms. This indication also reflects the fact that patients taking cerliponase alfa in the studies did not show improvement in motor or language symptoms (Table 9).

Several subgroups of interest were specified in the protocol for study selection, including disease severity, age of disease onset, and age. There were no data available to assess these subgroups and the number of patients in the studies was too small to allow assessment of these potential prognostic factors on treatment efficacy.

There were very few efficacy data available for Study 203 because the study did not accrue a long follow-up time prior to the data cut-off. The efficacy data from Study 203 did not contradict the observations in studies 201 and 202.

Harms

The assessment of harms is limited because there were no comparative AE data for the historical control group. CLN2 disease itself is associated with many symptoms and events that would be classified as AEs or SAEs, making it difficult to assess the relationship with cerliponase alfa for some AEs in the studies. According to a clinical expert consulted by CADTH, the risks related to implantation, maintenance, and eventual removal of the ICV system are a major consideration when deciding to initiate therapy with cerliponase alfa. AEs thought to be related to cerliponase alfa and/or its administration were evident in the clinical trials and included hypersensitivity, pyrexia, pleocytosis, device malfunction, and device-related infection. Sequelae of these events and treatment of these events related to cerliponase can have a major impact on the child's quality of life and the child's family, according to the clinical experts consulted by CADTH. Eventually, the device needs to be removed and replaced, as occurred in two patients after four years of usage. This is also an important consideration when weighing the potential benefits and risks of therapy.

[redacted]17

Study 201 had a dropout rate of one out of 24 (4%). This would be considered a low rate of attrition in most disease conditions; however, should not be ignored because the child was withdrawn after ICV implant that was subsequently not used. Rates of withdrawal in clinical practice where conditions are not as tightly controlled are unknown.

Potential Place in Therapy[2]

There is no cure for CLN2 disease and there is no treatment capable of disease modification. Diagnosis is confirmed through genetic testing and verification of TPP enzyme levels. In cases where the pathogenicity of mutations is not certain, additional evidence, such as decreased levels of TPP enzyme activity, are required. Given that the enzyme replacement therapy relies on disease manifestations through abnormally low TPP enzyme activity, the treatment would not be recommended where decreased TTP enzyme activity is not expected from either DNA testing or enzyme testing. Demonstration of storage on biopsy material is not sufficient alone to qualify for enzyme therapy with this product. The clinical features of the disease include seizures, movement disorders, behavioural problems, sleep disturbances, delays in growth, language, motor, cognitive, and visual function. The seizures may be hard to control with anti-seizure medication, the visual impairment is severe, and the developmental delay is significant, leading to inability to care for self, walk, or mobilize from bed to chair.

An ideal treatment would improve survival, or at least allow children to stay ambulatory for a longer time period; maintain speech and vision; and control seizures.

In the absence of any alternative, cerliponase may delay the progression of motor and language skills in some patients with CLN2 over a period of time, and may be a treatment

[2] This information is based on information provided in draft form by the clinical expert consulted by CDR reviewers for the purpose of this review.

option for some patients. Patients receiving cerliponase alfa should be managed by a specialist who is familiar with the treatment of CLN2 disease and the treatment of drug and device-related complications, and has familiarity with other supportive options, palliative care, and nutrition and rehabilitative services.

The clinical experts consulted by CADTH for this review stated that cerliponase alfa may be started in some diagnosed CLN2 cases prior to symptom onset. Decisions to stop treatment would be made in the context of the child's symptoms, treatment goals of the family, and the physician's professional judgment. Continued use of cerliponase in the presence of device complications, such as multiple trials of reservoir insertions, reservoir not allowing cerliponase alpha to be infused, or deterioration of the clinical scoring system to the point of reaching the 0 score on cerliponase alfa treatment, would not be appropriate. Stopping criteria have not been established for cerliponase alfa and the optimal duration of treatment is not known. Patients must meet qualification criteria for treatment using diagnostic testing, and patients and/or families should have the ability to comply with treatment, and have a clear understanding of the risks and potential complications of using implantable devices and the conditions under which cerliponase alfa could be stopped.

Conclusions

Patients treated with cerliponase alfa appear to have less deterioration in motor-language symptoms compared with historical, untreated control groups, and the difference may be clinically meaningful. While cerliponase alfa appears to slow the deterioration of motor-language symptoms in children with CLN2 disease, the use of historical control groups creates uncertainty about the magnitude of possible positive effects on motor-language scores and other outcomes. In studies 201 and 202, the historical control groups used may not be comparable with the cerliponase alfa–treated patients on important prognostic factors. It is also not clear whether cerliponase alfa improves overall HRQoL or reduces mortality compared with no treatment. The potential benefits of slower symptom deterioration need to be considered in the context of the known and unknown harms of using cerliponase alfa and the risks associated with ICV drug administration.

Appendix 1: Patient Input Summary

This section was summarized by CADTH Common Drug Review staff based on the input provided by patient groups.

1. Brief Description of Patient Group(s) Supplying Input

No patient group input was received by July 24, 2018, following CADTH's call for patient input. CADTH is unaware of a Canadian patient group for neuronal ceroid lipofuscinosis type 2 (CLN2) who could provide patient input. Given the rarity of the condition and invasive role of administration, CADTH saw patient input as important to contribute to the understanding of the disease and drug. CADTH accepted a description of the experiences of a Canadian family with a child with Batten disease CLN2, whose physician provided the family with CADTH's contact information for the purpose of providing input to this review.

2. Condition-Related Information

Cerliponase alfa (Brineura) is indicated for patients with CLN2 or Batten disease. The patient's mother described the disease as having a profound and quick progression on her son's body and intellect. Initially, the disease presented as seizures leading to an initial diagnosis of epilepsy. Experiencing seizures was described as something seen in most patients with Batten disease. Other physical issues followed, such as those related to the patient's gait. Per the report, the patient began to have difficulty walking and running, which progressed to falls while walking or standing; this eventually led to the need to use a walker full-time, followed by the use of a wheelchair while in public. The parents reported the patient went from writing with a pencil to barely being able to write his name. Further, difficulty coordinating movements (ataxia), and assistance with activities of daily living were also attributed to the progression of the disease.

As for the decline of the patient's intellectual ability, the patient was initially diagnosed with dyslexia at the age of six or seven based on an educational assessment by an educational psychologist, although it was indicated that this may be related to family history as both parents had dyslexia. At the age of eight (the same time that gait issues began), school teachers identified a significant change in the patient's cognitive ability; the patient progressed from dyslexia to a "significant intellectual disability" over a two-year period. Difficulty with speech and remembering concepts were also reported, as the patient became increasingly difficult to understand and could not remember concepts learned in school from the week before.

As Batten disease is both debilitating and a disease that presents during childhood, it requires assistance from a caregiver for things such as transfers and bathing, and food preparation and safety measures. The patient's mother reported quitting her full-time job to become a caregiver. Arrangements for transportation to and from school were also described, along with an individualized educational plan and attending school in a special needs classroom.

3. Current Therapy-Related Information

The parents state that the drug under review is the "only medication available for children with CLN2." They mentioned attending physiotherapy and occupational therapy to address and maintain strength and dexterity. They also described other patients who were "too far gone" and used a gastric tube for feeding.

4. Expectations About the Drug Being Reviewed

The family has had experience with cerliponase alfa. They stated that currently this is the only treatment for Batten disease and that the only thing that would be an improvement would be gene replacement therapy, which isn't available at this point. According to the parents, the only challenge experienced regarding the drug under review was gaining access to treatment in Canada. Before it was available, they travelled to the US every two weeks, and had to sell their house in order to be able to afford to do that; however, they expressed not having any regret doing so, "because the thing with Batten Disease is that every single day without medication makes a HUGE impact on the child's life.. each day lost means more brain cells die - more damage done."

The child is now able to receive treatment every two weeks at a hospital in Canada and the family indicated that they knew of one other Canadian patient who has access as well. The visit takes about a full day in the hospital, beginning with a wellness check followed by thawing of the medication (stored frozen). The patient is given Tylenol, Benadryl, and Ondansetron prior to infusion of cerliponase alfa, which is administered through a port in the patient's head. The infusion process takes approximately 4.5 hours and is followed by an assessment by a neurosurgeon. The caregiver reported that the medication is very well tolerated and other than fatigue caused by the Benadryl and Ondansetron and the day at the hospital, there are no side effects. Overall, they describe this treatment as being effective, claiming that they have witnessed the halting of the "frighteningly fast decline of [their] son" since treatment began.

Appendix 2: Literature Search Strategy

OVERVIEW	
Interface:	Ovid
Databases:	Embase 1974 to present MEDLINE All 1946 to present **Note:** Subject headings have been customized for each database. Duplicates between databases were removed in Ovid.
Date of Search:	August 23, 2018
Alerts:	Weekly search updates until January 16, 2019
Study Types:	No search filters were applied
Limits:	No date or language limits were used Human filter was applied Conference abstracts were excluded

SYNTAX GUIDE	
/	At the end of a phrase, searches the phrase as a subject heading
.sh	At the end of a phrase, searches the phrase as a subject heading
MeSH	Medical Subject Heading
fs	Floating subheading
exp	Explode a subject heading
*	Before a word, indicates that the marked subject heading is a primary topic; or, after a word, a truncation symbol (wildcard) to retrieve plurals or varying endings
#	Truncation symbol for one character
?	Truncation symbol for one or no characters only
adj#	Adjacency within # number of words (in any order)
.ti	Title
.ab	Abstract
.ot	Original title
.hw	Heading word; usually includes subject headings and controlled vocabulary
.kf	Author keyword heading word (MEDLINE)
.kw	Author keyword (Embase)
.pt	Publication type
.rn	CAS registry number
.nm	Name of substance word
medall	Ovid database code; MEDLINE All 1946 to present
oemezd	Ovid database code; Embase 1974 to present, updated daily

MULTI-DATABASE STRATEGY	
1	X8R2D92QP1.rn,nm.
2	(brineura* or cerliponase alfa* or cerliponase alpha* or tripeptidyl peptidase 1 or tripeptidyl peptidase I or tripeptidyl peptide hydrolase* or tripeptidylpeptidase* or tripeptidylpeptide hydrolase* or bmn 190 or bmn190 or immature cell growth-inhibiting gene 1 protein* or Immature human tripeptidyl-peptidase 1 or Immature lysosomal pepstatin-insensitive protease* or Immature tripeptidyl-peptidase I or tripeptidyl aminopeptidase* or pepinase* or rhtpp1).ti,ab,ot,kf,hw,rn,nm.
3	or/1-2
4	3 use medall
5	*tripeptidyl peptidase I/
6	(brineura* or cerliponase alfa* or cerliponase alpha* or tripeptidyl peptidase 1 or tripeptidyl peptidase I or tripeptidyl peptide hydrolase* or tripeptidylpeptidase* or tripeptidylpeptide hydrolase* or bmn 190 or bmn190 or immature cell growth-inhibiting gene 1 protein* or Immature human tripeptidyl-peptidase 1 or Immature lysosomal pepstatin-insensitive protease* or Immature tripeptidyl-peptidase I or tripeptidyl aminopeptidase* or rhtpp1).ti,ab,kw.
7	or/5-6
8	7 use oemezd
9	4 or 8
10	conference abstract.pt.
11	9 not 10
12	exp animals/
13	exp animal experimentation/ or exp animal experiment/
14	exp models animal/
15	nonhuman/
16	exp vertebrate/ or exp vertebrates/
17	or/12-16
18	exp humans/
19	exp human experimentation/ or exp human experiment/
20	or/18-19
21	17 not 20
22	11 not 21
23	remove duplicates from 22

OTHER DATABASES		
PubMed	A limited PubMed search was performed to capture records not found in MEDLINE. Same MeSH, keywords, limits, and study types used as per MEDLINE search, with appropriate syntax used.	
Trial registries (Clinicaltrials.gov and others)	Same keywords, limits used as per MEDLINE search.	

Grey Literature

Dates for Search:	August 2018
Keywords:	Brineura (cerliponase alfa), neuronal ceroid lipofuscinosis type 2
Limits:	No date or language limits used

Relevant websites from the following sections of the CADTH grey literature checklist *Grey Matters: a practical tool for searching health-related grey literature* (https://www.cadth.ca/grey-matters) were searched:

- health technology assessment agencies
- health economics
- clinical practice guidelines
- drug and device regulatory approvals
- advisories and warnings
- drug class reviews
- databases (free)
- Internet search.

Appendix 3: Excluded Studies

Table 18: Excluded Studies

Reference	Reason for Exclusion
28	Not an interventional study
29	Not an interventional study
30	Review
31	Review

Appendix 4: Detailed Outcome Data

Table 19: Summary of Symptom Score Changes Over Time

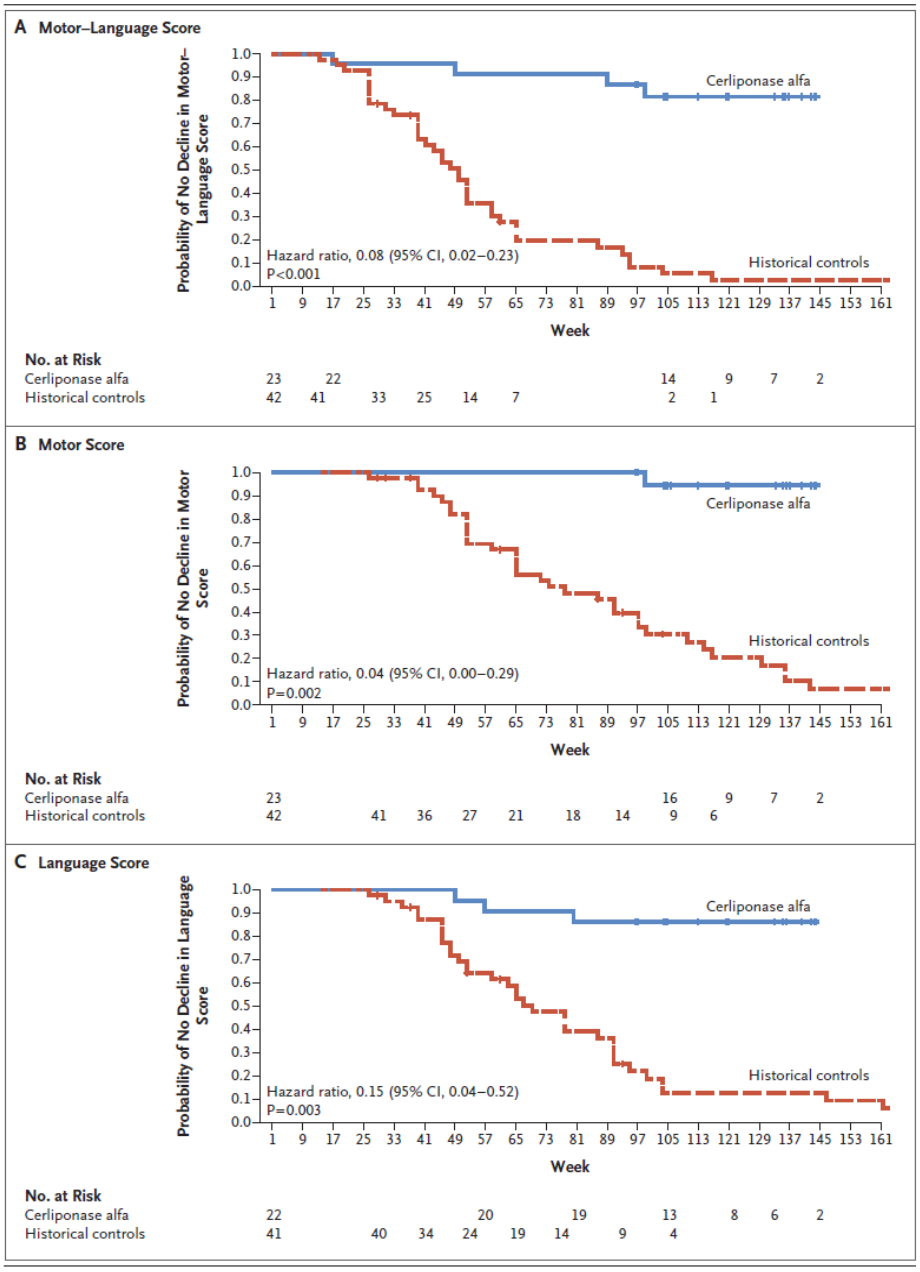
	ML Scale (0-6)		MLV Scale (0-9)		Total CLN2 Scale (0-12)	
	190-901	190-201/202	190-201	190-201/202	190-201	190-201/202
Baseline						
n	42	23	42	23	42	23
Mean (SD)	4.5 (0.77)	3.5 (1.20)	7.2 (1.12)	6.3 (1.34)	9.3 (1.80)	8.0 (1.83)
Week 49						
n	39	23	39	23	39	23
Mean (SD)	2.3 (1.11)	3.0 (1.33)	4.7 (1.47)	5.7 (1.56)	7.0 (2.38)	7.9 (2.07)
Mean (SD) Change from Baseline	-2.2 (1.09)	-0.4 (0.79)	-2.6 (1.29)	-0.6 (1.03)	-2.3 (1.89)	-0.1 (1.93)
Week 73						
n	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Mean (SD)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Mean (SD) Change from Baseline	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 97						
n	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Mean (SD)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Mean (SD) Change from Baseline	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

CLN2 = neuronal ceroid lipofuscinosis type 2; ML= motor-language; MLV = motor-language-vision; SD = standard deviation.

Note: 190-201/202 (N = 23) are the treated patients. 190-901 (N = 42) are unmatched historical controls.

Source: Common Technical Document section 2.7.3.[4]

Figure 4: Kaplan-Meier Plots of Motor and/or Language Scale — Responder Analyses (Time Until the First Two-Point Decline)

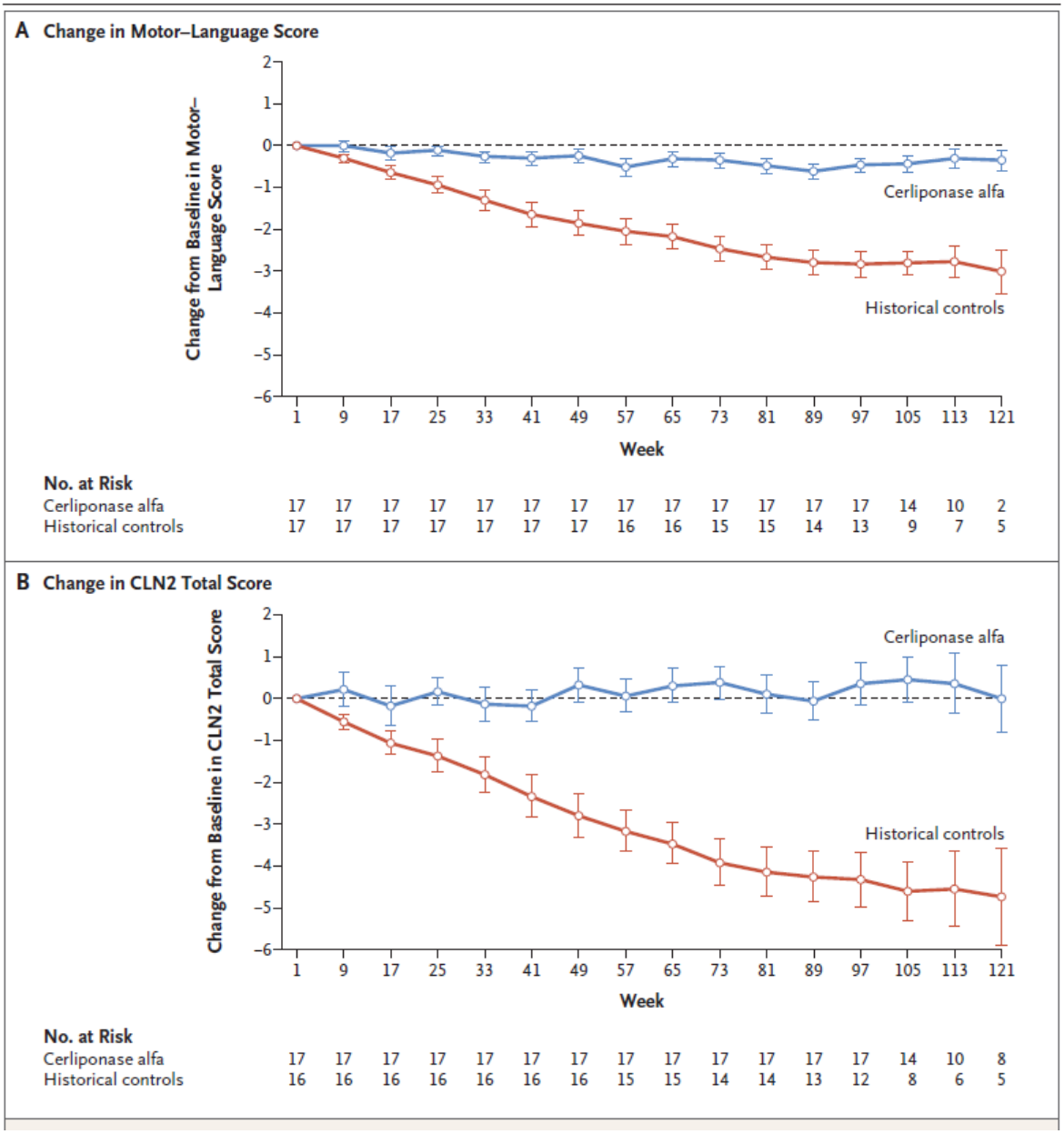


No. = number.

Note: These Kaplan-Meier plots depict the time until the primary outcome for studies 201 and 202; i.e., the time to an unreversed decrease of two or more points in the combined score for motor and language function or a combined score of 0. Baseline was the beginning of the 300 mg dose period. Plot A depicts the motor-language scores; Plot B depicts the motor score; Plot C depicts the language score.

Source: The New England Journal of Medicine, Schulz A, et al., Study of intraventricular cerliponase alfa for CLN2 disease, 378, 1898-907. Copyright © (2018) Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.[10]

Figure 5: Mean Change From Baseline in Motor-Language and Total Neuronal Ceroid Lipofuscinosis Type 2 Scores



CLN2 = neuronal ceroid lipofuscinosis type 2; No. = number.

Note: These plots depict the change from baseline in the score for motor and language function (Plot A) and in the four domains of the total CLN2 score (Plot B: motor, language, vision and seizure). Vertical bars represent standard errors.

Source: The New England Journal of Medicine, Schulz A, et al., Study of intraventricular cerliponase alfa for CLN2 disease, 378, 1898-907. Copyright © (2018) Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.[10]

Appendix 5: Validity of Outcome Measures

Aim

To summarize the validity of the following outcome measures:

- Hamburg Scale (the Clinical Scoring System for Late Infantile Neuronal Ceroid Lipofuscinoses [LINCL])
- Weill Cornell LINCL Scale
- Neuronal Ceroid Lipofuscinosis Type 2 (CLN2) Clinical Rating Scale
- CLN2 Quality of Life (CLN2QL) Questionnaire
- Pediatric Quality of Life Inventory (PedsQL) 4.0 Generic Core Scales and Family Impact Module
- Denver II Developmental Screening Test (Denver II)
- European Quality of Life 5 Dimensions 5 Levels (EQ-5D-5L).

Findings

Table 20: Validity and Minimal Clinically Important Differences of Outcome Measures

Instrument	Type	Evidence of Validity	MCID	References
Hamburg Scale (Clinical Scoring System for LINCL)	• Disease-specific (CLN2 disease), four category scale (includes motor, language, and visual scales plus an additional seizure score, which is not included in the Total Disability Score); with higher scores indicating lower disability • Total Disability Score has a maximum score of 9 (as the Motor Function, Visual Function, and Language functional categories are scored between 0 and 3)	No literature identified	Not identified	Steinfeld, 2002[31]
Weill Cornell LINCL Scale	• Disease-specific (CLN2 disease), four category scale (includes swallowing, gait, motor, and language); with higher scores indicating lower disability • Total Disability Score has a maximum score of 12 (as the Swallowing Function, Gait Function, Motor Function, and Language functional categories are scored between 0 and 3)	No literature identified	Not identified	Worgall, 2007[29]
CLN2 Clinical Rating Scale (Adapted) ML Domain	• Disease-specific (CLN2 disease), adapted version of the CLN2 Clinical Rating	Yes for construct validity	Decrease of one point[a]	Wyrwich, 2018[26]

Instrument	Type	Evidence of Validity	MCID	References
	Scale that includes motor, language, and combined ML scores • The total score of the ML combined scale is 6 (scores ranging between 0 and 6), with higher scores indicating better function			
CLN2QL Questionnaire	CLN2QL is a CLN2 disease-specific questionnaire that is comprised of 28 items in six domains related to CLN2 disease	Not identified	Not identified	FDA Statistical Review[11] CSR 201[14]
PedsQL 4.0 Generic Core Scales and Family Impact Module	Surveys consisting of 23 items in the generic core scales for measuring health-related quality of life in healthy and patient pediatric populations and 36 items in the family impact module for assessing parental function. Items in both are scored with a five-point Likert scale	Yes (with limitations)	Generic Core Scales in the general pediatric population: 4.4 points for self-report and 4.5 points for proxy-report Family Impact Module: Not identified	Varni, 2003,[20] Varni, 2002[32] Varni, 2004[21]
Denver II Developmental Screening Test	• A revision and update of the Denver Developmental Screening Test, which aims to monitor and assess infant and preschool-aged children's development • Four general functions are assessed: Personal-Social, Fine Motor-Adaptive, Language, and Gross Motor	Not identified	Not identified	Glascoe, 1992[22] CSR 201[14] FDA Statistical Review[11]
EQ-5D-5L	• EQ-5D is a general, non–disease-specific health-related quality of life questionnaire	Yes	Index score: Summarized mean, 0.056 (SD 0.011), summarized median, 0.056 (IQR 0.049, to 0.063) (general use)	Health Quality Council of Alberta, 2014[24] McClure, 2017[25]

CNL2 = neuronal ceroid lipofuscinosis type 2; CLN2QL = Neuronal Ceroid Lipofuscinosis Type 2 Quality of Life; CSR = clinical study report; EQ-5D-5L = EuroQol 5-Dimensions 5-Levels; IQR = interquartile range; LINCL = late infantile neuronal ceroid lipofuscinoses; MCID = minimal clinically important difference; ML = motor-language; PedsQL = Pediatric Quality of Life Inventory; QoL = quality of life; SD = standard deviation.

[a] There remains a lot of uncertainty surrounding this number as there was no evidence of how the authors obtained it.[26]

Hamburg Scale (Clinical Scoring System for Late Infantile Neuronal Ceroid Lipofuscinoses)

The Hamburg Scale was designed to ascertain a quantitative disease-specific description of the clinical course of LINCL that could be obtained using a simple assessment system; namely, input made by the families of patients.[31] It is a clinician-reported scale that consists of four main functional categories: motor function, seizures (grand mal), visual function, and language. Within each of these functional categories, the scores are assigned per ability, with higher scores representing better functioning (see the following). A Total Disability Score (maximum score of 9) can also be obtained by summing up the motor, visual, and language function scores, with lower scores indicating greater disability.

Motor Function

- Walks normally (score = 3)
- Frequent falls, clumsiness evident (score = 2)
- No unaided walking or only crawling (score = 1)
- Immobile, mostly bedridden (score = 0)[31]

Seizures (Grand Mal)

- No seizure in a three-month period (score = 3)
- One to two seizures (per three-month period) (score = 2)
- One seizure per month (score = 1)
- Greater than one seizure per month (score = 0)[31]

Visual Function

- Recognizes and grabs at desirable object (score = 3)
- Uncoordinated grabbing for objects (score = 2)
- Reacts to light (score = 1)
- No reaction to any visual stimuli (score = 0)[31]

Language

- Normal (according to the individual's maximum) (score = 3)
- Has become recognizably abnormal (score = 2)
- Hardly understandable (score = 1)
- Unintelligible or no language (score = 0)[31]

Within the motor function category, the "walks normally" has a score of 3; however, many children with CLN2 will never completely reach this milestone. In addition, "normal" (individual maximum), score of 3 of the language function category, may also never be reached in children with CLN2; therefore, the suggestion by the authors would be to take the best performance observed by the child as the starting point and score that as 3 (knowing that there will eventually be a decline in function).[31] Only grand mal seizures are recorded for the assessment of epileptic activity; however, given that seizure frequency can vary and will be impacted by the use of anti-seizure medications, this score is omitted from the Total Disability Score.[31]

Steinfeld et al.[31] developed the Hamburg Scale to examine the natural course of CLN2 disease and then assessed this scale by interviewing 22 patients in Germany and Switzerland with a clinical and genetic diagnosis (of two mutant alleles) of CNL2 (or LINCL) disease. In general, the scale can be administered using an interview setting; the interview can take between one to two hours. Of these 22 patients, 16 were identified as standard patients (referring to the exclusivity of their mutations [R208X and IVS 5-1G>C]) and followed a similar decline in function within a relatively narrow time frame. Rapid declines in function were observed to start at the age of three and were typically followed by a complete loss of function in these patients by the age of five.[31] The usual appearance of epileptic seizures occurred toward the end of the third year; however, frequency was observed to be variable (presumably due to differing anti-convulsant treatment). In addition, while blindness was typically observed after three years, there were cases where it took up to ten years to fully observe. There were some similarities in functional decline noted when the standard patients were compared with others other patients with different mutations; however, there were also some substantially different rates of decline in other patients who displayed with less rapid decline.[31] Therefore, while this study shows that there is variability within the disease course that appears to be dependent on genetic variation, this study did not provide evidence of reliability, validity, or responsiveness of the Hamburg Scale in patients with CLN2. In addition, there was no minimal clinically important difference (MCID) identified for the Hamburg Scale for patients with CLN2 disease.

Weill Cornell Late Infantile Neuronal Ceroid Lipofuscinoses Scale

The Weill Cornell LINCL Scale[29] is an expansion of the modified Hamburg LINCL scale that aims to identify a broad range of phenotypic differences in children with CLN2 disease. It is comprised of an expanded neurological assessment compared with the modified Hamburg LINCL scale and includes four main functional categories (swallowing, gait, motor, and language), which are scored from 0 to 3 (with higher scores indicating better function). In order to obtain an index of disability, the functional category scores are summed to achieve a score of 0 (indicating severe disability) and 12 (indicating normal function).[29] Ratings for the functional categories are as follows.

Swallowing Scale

- No swallowing dysfunction (score = 3)
- Mild swallowing dysfunction (score = 2)
- Moderate swallowing dysfunction (score = 1)
- Gastronomy dependent (score = 0)[29]

Gait Scale

- Normal (score = 3)
- Abnormal (spastic, bradykinetic, or ataxic) but ambulates independently (score = 2)
- Abnormal (spastic, bradykinetic, or ataxic) and unable to independently ambulate (score = 1)
- Non-ambulatory (score = 0)[29]

Motor Scale

- No myoclonus, chorea/tremor/athetosis, or positive Babinski reflex (score = 3)
- One of myoclonus, chorea/tremor/athetosis, or positive Babinski reflex (score = 2)
- Two of myoclonus, chorea/tremor/athetosis, or positive Babinski reflex (score = 1)
- Myoclonus and one of chorea or tremor or athetosis and positive Babinski reflex (score = 0)[29]

Language Scale

- Normal speech (score = 3)
- Abnormal speech (abnormal articulation or decreased vocabulary) (score = 2)
- Barely understandable speech (severe dysarthria or very few meaningful words) (score = 1)
- Unintelligible words or no speech (score = 0)[29]

In a study of 18 children from the US, Canada, Germany, Australia, and England (nine boys and nine girls who followed the classic CLN2 disease pattern in terms of the time of onset of language abnormalities, seizures, and motor abnormalities) with clinically and genetically confirmed CLN2 disease, Worgall et al.[29] examined the usefulness of the Weill Cornell LINCL Scale in comparison with a modified Hamburg Scale (unclear how modified) and MRI/magnetic resonance spectroscopy imaging. Both the modified Hamburg Scale and the Weill Cornell LINCL Scale were moderately correlated with the age of the children (r^2 of 0.55 and 0.44, respectively) and were also moderately correlated with the time since initial clinical presentation (r^2 of 0.53 and 0.35, respectively).[29] When comparing the age progression with the modified Hamburg Scale, a steeper rate of decline in Weill Cornell LINCL Scale scores was observed when examining the patients at progressively older ages (y of –0.85 and –0.36, respectively). No differences in scores using either the modified Hamburg Scale or the Weill Cornell LINCL Scale were observed when assessing patients[29] Brain imaging, however, revealed brain volume declines with age (based on cortical N-acetylaspartate, gray matter volume, per cent gray matter volume of total brain volume). The authors surmised that this was likely due to the higher ratings on the seizure scales using the modified Hamburg Scale (presumably due to either the optimization of anti-convulsant treatments or from the loss of seizure activity due to the massive neurodegeneration associated with age).[29] This study did not provide evidence of reliability, validity, or responsiveness of the Weill Cornell LINCL Scale in patients with CLN2. In addition, there was no MCID identified for the Weill Cornell LINCL Scale for patients with CLN2 disease.

Neuronal Ceroid Lipofuscinosis Type 2 Clinical Rating Scale (Two-Item Version) Motor-Language Domain

The CLN2 Clinical Rating Scale Motor-Language Domain is a scale that was specifically adapted from the Hamburg and Weill Cornell CLN2 clinical rating subscales for use in the cerliponase alfa clinical trials (studies 201 and 202). It includes motor and language anchor points that were derived "from existing natural history databases"[26] as the results of the children treated with cerliponase alfa in clinical trials 201 and 202 were to be compared with historical data (and the adaptation of the motor and language function would help with coding).[26] The authors' reasons for only including the motor and language function in this scale are that motor and language function decline rapidly in the natural history of CLN2,

standard of care does not seem to affect either motor or language function, and their declines are fundamental to the disease.[26] Per the Hamburg and Weill Cornell CLN2 clinical rating scales, the scoring algorithm is similar (albeit the descriptions were slightly modified), with higher scores indicating better function. A rating of 3 indicates "a normal condition," 2 is "a slight or just noticeable abnormality," 1 is "a severe abnormality," and 0 indicates a "complete loss of functioning."[26] The combined motor-language function score ranges from 0 (most severe disability) to 6 (normal function).[26] The descriptions of the specific scale scores are described as follows:

Motor Function

- Grossly normal gait. No prominent ataxia, no pathologic falls (score = 3)
- Independent gait defined by the ability to walk without support for 10 steps. Patients will have obvious instability and may have intermittent falls (score = 2)
- Requires external assistance to walk or can only crawl (score = 1)
- Can no longer walk or crawl (score = 0)[26]

Language Function

- Apparently normal language. Intelligible and grossly age-appropriate. No decline noted yet (score = 3)
- Language has become recognizably abnormal: some intelligible words may form short sentences to convey concepts, requests, or needs. Denotes a decline from a previous level of ability (from the individual maximum reached by the child) (score = 2)
- Hardly understandable. Few intelligible words (score = 1)
- No intelligible words or vocalizations (score = 0)[26]

In order to determine the construct validity, inter-rater reliability, and responsiveness of the combined motor-language domain, the PedsQL Generic Core Scales and Family Impact Module, and the Hamburg Scale, were also administered in person at eight-week intervals throughout the two trials (studies 201 and 202). Sessions were also videotaped.[26]

Inter-rater reliability was assessed using videotaped analysis by measuring the moto-language scale motor rating, the motor-language scale language rating, and the combined motor-language scale scores at baseline (n = 12), week 25 (n = 11), week 49 (n = 10), and week 73 (n = 3) in Hamburg Clinic patients in 2016. A non-study scale trainer assessed the videotapes and these ratings were compared with the treating clinician's assessments using the Landis and Koch classifications of agreement (0.00 to 0.20 indicates slight agreement; 0.21 to 0.40 indicates fair agreement; 0.41 to 0.60 indicates moderate agreement; 0.61 to 0.80 indicates substantial agreement; 0.81 to 0.99 indicates almost perfect agreement; 1.00 indicates perfect agreement).[26] Weighted kappa statistics ranged between substantial agreement (0.76 at baseline) and perfect agreement (1.00 for all the other time points) for the motor-language scale motor ratings, almost perfect agreement (0.93 at baseline) to substantial agreement (0.67 to 0.80 for the other time points) for the motor-language scale language motor ratings, and almost perfect agreement (0.92 at baseline) to almost perfect agreement (0.89 to 0.93 for the other time points) for the combined motor-language scale score. When examining all four time points (n = 36) almost perfect agreement (range of 0.82 to 0.93) was observed for the motor, language, and combined motor-language scores.[26] However, the small sample size (especially at week 73 [n = 3]) calls into question the precision of these findings. Spearman correlation coefficients

were used to assess the construct validity of the motor, language, and combined motor-language scale at baseline with other similar health measures, including the Physical Functioning Domain and the Total Scale score of the PedsQL Generic Core Scales, V.4.0, and the Total Score of the Family Impact Module.[26] The impact of the child's motor function on the caregiver's role was assessed by comparing with PedsQL three key domain scores (Physical, Emotional, and Social Functioning) and the total score of the PedsQL Family Impact Module. The language motor-language score was compared with PedsQL Family Impact Module domain scores for Social, Cognitive, and Communication Functioning. In addition, the combined motor-language score was compared with the PedsQL Total Score, the Total Family Impact Module Score, and the CLN2 QL Questionnaire Daily Activities Score (subsequently described).[26] The motor domain and combined motor-language scores were moderately correlated with the PedsQL Physical Functioning Domain (r = 0.64) and Total Score (r = 0.61), while the language domain and the combined motor-language score was moderately correlated with the PedsQL Family Impact Module Communication Domain (r = 0.65), Social Domain (r = 0.57), and Cognitive Function Domain (r = 0.47). Weak-to-moderate correlations were observed between the combined motor-language score and the PedsQL Total Score (r = 0.35), the PedsQL Family Impact Module Total Score (r = 0.51) and the CLN2 Daily Activities Score (r = 0.35).[26]

Spearman correlation coefficients were used to assess the responsiveness of the motor, language, and combined motor-language scores from baseline to week 49 when compared with the change score on the CLN2QL Questionnaire Daily Activities Score.[26] Weak-to-moderate correlations were observed in the responsiveness (from baseline to week 49 of the trial) of the motor and combined motor-language scores when compared with the CLN2QL Daily Activities Score (r = 0.37 and 0.41, respectively).[26]

According to Wyrwich et al.,[26] a measurable and clinically important difference is one point on the motor or language rating scale. However, no information on how they obtained that number was identified and therefore, the MCID of the motor and language rating scales remain uncertain.

Upon examining the evidence from studies 201 and 202 regarding the inter-rater reliability of the combined motor-language score, the FDA noted that[33] during the trial, only one clinician assessed the patients at each time point and the same clinician was not consistently used through the trial for assessments at the various time point; hence, there were challenges faced by the FDA in its evaluation of the inter-rater reliability.[11]

Neuronal Ceroid Lipofuscinosis Type 2 Quality of Life Questionnaire

The CLN2QL Questionnaire is a CLN2 disease-specific questionnaire designed by Biomarin to examine the health-related quality of life (HRQoL) in patients with CLN2 disease undergoing treatment with Brineura.[11] Family feedback from two focus groups (one in Europe and one in the US) was ascertained in order to help with the development of this questionnaire.[14] The CLN2QL Questionnaire is comprised of 28 items that are part of six domains related to CLN2 disease. The items are partitioned as follows:

- seizures (six items)
- feeding (with no gastrostomy tube) (four items)
- feeding (with a gastrostomy tube) (three items)
- sleep (five items)
- behaviour (six items)

- daily activities (four items).[11]

Pediatric Quality of Life Inventory 4.0 Generic Core Scales and Family Impact Module

The PedsQL Generic Core Scales is intended to be administered in both healthy and patient pediatric populations and, together with disease-specific modules, measures pediatric HRQoL.[20] The Generic Core Scales is available in formats for child self-report and parent proxy-report for ages 5 to 7, 8 to 12, and 13 to 18, along with a parent proxy-report format for ages 2 to 4.[20] Each item is scored on a five-point Likert scale (three-point scale for ages 2 to 4) with each score linearly transformed to a scale of 0 to 100.[20] To generate domain and total scores, the transformed item scores are summed and then divided by the number of items (range: 0 to 100).[20] The Generic Core Scales consist of the following scales: Physical Functioning (8 items), Emotional Functioning (5 items), Social Functioning (5 items), and School Functioning (5 items), with higher scores indicating better HRQoL.[20] The Psychosocial Health Summary Score is the sum of the items in the Emotional, Social, and School Functioning Scales.[20] Higher scores indicate better HRQoL.[20]

The PedsQL Family Impact Module is a parent self-reporting instrument that is comprised of 36 items measuring parental functioning. The domains that are examined include Emotional Functioning (5 items), Social Functioning (4 items), Physical Functioning (6 items), Communication (3 items), Cognitive Functioning (5 items), Worry (5 items), along with two other scales that measure family functioning: Family Relationships (5 items) and Daily Activities (3 items).[21] A five-point Likert scale is used for all items with 0 meaning never a problem and 4 meaning always a problem. A 0 to 100 scale is used (after reverse scoring and linear transformation of items) with higher scores indicating better functioning.[21] In order to compute the domain scores, the items are summed and then divided by the number of items that were reported; however, if more than 50% of the items in a domain are missing then the individual Domain Score is not calculated, as per the developers instructions. The Family Impact Module Total Score is obtained by summing the 36-items and dividing by the number of items answered.[21]

Pediatric Quality of Life Generic Core Scales

In order to validate the PedsQL Generic Core Scales, a sample of chronically ill (as reported by their parents in a specialty clinic [n = 683]), acutely ill (parents reported no presence of chronic illness and attended a specialty clinic [n = 207]), and healthy children (identified at their physician's office during regular visits or using telephone calls [n = 730]) between the ages of two to 18 years were included.[34] Construct validity was assessed using the known-groups method, whereby scale scores were compared across groups that are known to differ in the specific health constructs being examined (in this case, healthy versus acute or healthy versus chronic conditions). In addition, potentially confounding factors such as age, gender, and ethnicity were also examined across health states. Hypothesizing that healthy children would have a higher HRQoL, Varni et al. noted that the PedsQL differentiated between the different health states (healthy, acute, and chronically ill) and it also correlated with illness burden and morbidity measures.[34] Internal consistency reliabilities generally exceeded the Cronbach's coefficient alpha of 0.70. The internal consistency reliability for the Total Scale Score across the ages for the self-report and proxy-report were 0.88 and 0.90, respectively, indicating this as an appropriate primary analysis summary score. The internal consistency reliability for the Physical Health and Psychosocial Health Summary Scores were greater than 0.8 for the self-report and the

proxy-report; hence, the authors determined they were best for secondary analyses. The Emotional, Social, and School Functioning Subscales generally obtained Cronbach's coefficient alpha around 0.70; therefore, the authors suggested these be used for descriptive or exploratory analyses.[34]

Varni et al.[32] then examined three studies in order to determine the sensitivity and responsiveness of the PedsQL 4.0 Generic Core Scales. The population included pediatric patients (age range: two to 18 years) with acute or chronic health conditions (n = 115 presenting to a cardiology clinic; n = 47 presenting to an orthopedic clinic; n = 127 presenting to a rheumatology clinic) and their parents. Statistically significant differences were observed between pediatric patients defined as Class II/IV New York Health Assessment and Classes I and II, suggesting that the PedsQL was likely to be sensitive to detecting differences between the two classes.[32] Likewise, statistically significant changes between the initial and follow-up visit of patients attending the orthopedic clinic (to treat their fractures) were observed (and the follow-up visit results also corresponded to that of healthy children responses), demonstrating the responsiveness of the PedsQL.[32] In another study by Desai et al.,[35] patients admitted to medical or surgical units were administered the PedsQL 4.0 upon admission (64.5%; n = 4,637 out of 7,184) and during follow-up (58.1%; n = 2,694 out of 4,637). The responsiveness of the PedsQL was demonstrated upon examination of the mean differences between admission and follow-up; 22.1 (standard deviation [SD] of 22.7) for the total score, 29.3 (SD of 32.4) for the physical domain, and 17.1 (SD of 21.0) for the Psychosocial Domain. Moderate variability in responsiveness was observed by age and minimal variability in responsiveness was observed for patients having been admitted for medical or surgical reasons.[35] Construct validity was further demonstrated as patients with no chronic illness (and their parents) scored higher on the Total Score, Physical Domain, and Psychosocial Domain when compared when patients with either complex or non-complex chronic illness.[35]

In a study by Varni et al.,[20] the authors mailed out a survey to residents in California (of which 10,241 [51%] completed and returned the survey) to further examine the validity, reliability, and feasibility of the PedsQL 4.0 Generic Core Scales. In this study, the authors also explored the MCID by calculating the Standard Error of Measurement in the survey responses. They determined the MCID for the Total Scale Score of the child self-report is a change of 4.4, while the MCID for the Total Scale Score for parent proxy-report is a change of 4.5.[20] However, an anchor-based approach using a valid patient-reported scale would be a preferable approach to determining the MCID.

No evidence regarding an MCID was identified for the CLN2 population.

Pediatric Quality of Life Family Impact Module Scale Score

In order to assess the preliminary validity and reliability of the PedsQL Family Impact Module, the investigators tested this module in families with "medically fragile children with complex chronic medical conditions"[21] whose children were either being cared for in a convalescent hospital (n = 12) or at home (n = 11). Using the known-groups method, the construct validity of the Family Impact Module was established by obtaining means and accompanying t-tests. Medium-to-large means were observed in both groups, with means ranging between 69.17 and 88.75 for the parents of the children residing in convalescent care and 51.89 to 78.95 for the parents of the children residing at home. This supported the author's hypothesis that those parents whose children were being cared for in a convalescent hospital would have better HRQoL than those caring for their children at home.[21]

Cronbach's alpha was calculated to determine the internal consistency reliability of the PedsQL Family Impact Module, with reliabilities of 0.70 or higher indicating appropriateness of comparing across groups and reliabilities of 0.90 or higher indicating appropriateness for examining individual patient scale scores.[21] All individual domain scores and the Total Score obtained the minimum of 0.70 or higher (ranging between 0.79 to 0.98 in the convalescent group and ranging between 0.83 and 0.96 for the home group). In addition, most domain scores reached or exceeded the minimum criteria of 0.90 for measuring individual scale scores in six domains in both the convalescent and home groups.[21] However, generalizability of these findings may be limited by the small sample size used in this study.[21]

No evidence regarding an MCID was identified for the general chronically ill patient population or the CLN2 population.

Denver II Developmental Screening Test

The Denver II is a revision and update of the Denver Developmental Screening Test, which aims to monitor and assess infant and preschool-aged children's development. It is comprised of 125 performance-based and parentally reported items (with 20 new items added onto the original Denver Developmental Screening Test) split between four functional groups: Personal-Social (e.g., smiling), Fine Motor-Adaptive (e.g., grasping, drawing), Language (e.g., combining words), and Gross Motor (e.g., walking).[11,14,22] Instead of using a scoring system, pass/fail/refusal scores are assigned to each item which is then reinterpreted in terms of caution, delay, no opportunity, normal, or advanced performance based on the child's age. An abnormal overall test score is obtained when two or more delays are noted and a questionable score is obtained when one delay and/or two or more cautions are noted.[22] Only trained professionals are permitted to administer the Denver II test.[22]

Sensitivity and specificity of the Denver II to identify children with developmental problems were assessed in one study of 104 children of different demographic profiles who were enrolled in five daycare centres in Nashville, Tennessee. The Denver II and a battery of other recognized tests that were age-specific (e.g., but not limited to, Kaufman Assessment Battery for Children Achievement Subsets, Stanford-Binet Intelligence Scale, Fluharty Preschool Speech and Language Screening Test, and the Peabody Picture Vocabulary Test) were administered to the children in order to obtain a developmental diagnosis. Children with developmental issues based on the previously mentioned tests were observed to be categorized as abnormal, questionable, or untestable on the Denver II; however, it was also noted that suspect scores were also observed in nearly half of the children with no developmental issues. Hence, the Denver II was associated with high over-referral rates.[22] Upon further examination, Glascoe et al.[22] noted that those children who were most over-referred received either delays or cautions in only their Fine Motor-Adaptive, Gross-Motor, or Personal-Social domains. The high sensitivity and low specificity of the Denver II were also noted across all age groups examined (age range from 0 to 72 months).[22]

No evidence regarding the validity, reliability, or responsiveness of this test were identified in patients with CLN2 disease.

EuroQol 5-Dimensions 5-Levels Questionnaire

EQ-5D is a generic quality-of-life instrument developed by the EuroQol Group.[23] It may be applied to a wide range of health conditions and treatments.[23] As a generic measure of HRQoL that can capture the net effect of treatment benefits and harms, the EQ-5D provides valuable information from a patient perspective. In addition to this purpose, the EQ-5D is used in clinical trials to obtain utility weights for economic models.[24] The EQ-5D 5-Levels version (EQ-5D-5L) was introduced in 2005 based on the earlier 3-Levels version (EQ-5D-3L).[23] It consists of an EQ-5D descriptive system and the EQ visual analogue scale (VAS). The descriptive system comprises the following five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, each with five levels: a level 1 response represents "no problems," level 2 "slight problems," level 3 "moderate problems," level 4 "severe problems," and level 5 "extreme problems" or "unable to perform," which is the worst response in the dimension. Respondents are asked to choose the level that reflects their health state for each of the five dimensions. In total, there are 3,125 possible unique health states defined by the EQ-5D-5L, with 11111 and 55555 representing the best and worst health states. The numerical values assigned to levels 1 to 5 for each dimension reflect rank order categories of function. In terms of measurement properties, these are ordinal data; they do not have interval properties and therefore should not be summed or averaged to, for example, produce an individual dimension "score." Results from the EQ-5D-5L descriptive system can be converted into a single index score using a scoring algorithm taking the local patient and population preferences into account. Therefore, the index score is a country-specific value and a major feature of the EQ-5D instrument.[24] The range of index scores will differ according to the scoring algorithm used; however, in all scoring algorithms of the EQ-5D-5L, a score of 0 represents the health state "dead" and 1.0 reflects "perfect health." Negative scores are also possible for those health states that society (not the individual patient) considers to be "worse than dead."

The EQ VAS records the respondent's self-rated health on a vertical VAS where the end points are labelled 0 ("the worst health you can imagine") and 100 ("the best health you can imagine"). The respondents are asked to mark an X on the point of the VAS that best represents their health on that day. The EQ-5D index and VAS scores can be summarized and analyzed as continuous data.[23,24] Hence, the EQ-5D produces three types of data for each respondent:

- a profile indicating the extent of problems on each of the five dimensions represented by a five-digit descriptor, such as 11121 or 21143
- a population preference-weighted health index score based on the descriptive system
- a self-reported assessment of health status based on the EQ VAS.

The EQ-5D-5L has been validated in terms of feasibility, ceiling effects, discriminatory power, and convergent validity in a diverse patient population from six countries with chronic conditions.[23] MCID estimates for the index score in the general Canadian population were generated by simulating the effects of single-level transitions in each dimension.[25] The results yielded MCIDs with a summarized mean of 0.056 (SD of 0.011), and a summarized median of 0.056 (interquartile range, 0.049 to 0.063).[25] There was no evidence of validity, reliability, responsiveness, or an MCID specific to CLN2 disease or any neuronal ceroid lipofuscinosis disease identified in the literature for the EQ-5D-5L.

Conclusion

The Hamburg Scale was designed in order to ascertain a quantitative disease-specific description of the clinical course of LINCL that could be obtained using a simple assessment system; namely, input made by the families of patients.[31] It is a clinician-reported scale. No MCID was identified for the Hamburg Scale in the CLN2 disease population.

The Weill Cornell LINCL Scale is an expanded neurological assessment compared with the modified Hamburg LINCL scale and is a clinician-reported scale.[29] No MCID was identified for the Weill Cornell LINCL Scale in patients with CLN2 disease.

The CLN2 Clinical Rating Scale Motor-Language Domain is a scale that was specifically adapted from the Hamburg and Weill Cornell CLN2 clinical rating subscales for use in the cerliponase alfa clinical trials (studies 201 and 202). It includes motor and language anchor points that were derived "from existing natural history databases" as the results of the children treated with cerliponase alfa in clinical trials 201 and 202 were to be compared with historical data (and the adaptation of the motor and language function would help with coding).[26] The authors stated that a change of one point was a clinical important difference in patients with CLN2 disease; however, there was no information provided on how this MCID was estimated.

The CLN2QL Questionnaire is a CLN2 disease-specific questionnaire that is comprised of six domains that reflect the nature of CLN2 disease. It was developed by Biomarin. No validity, reliability, responsiveness, or an MCID have been identified in patients with CLN2 disease.

The PedsQL 4.0 Generic Core Scales and Family Impact Module are surveys for assessing HRQoL in pediatric patients, each with multiple scales and the parent proxy-report formats, with the Generic Core Scale Score also including child self-reports.[20,21] MCIDs of 4.4 and 4.5 for the self-report and proxy-report formats have been estimated for the Generic Core Scales Total Score in the general pediatric population, by calculating the Standard Error of Measurement;[21] however, an anchor-based approach using a valid patient-reported scale would be preferable. An MCID was not identified for the Generic Core Scales or Family Impact Module in the CLN2 population.

The Denver II is a revision and update of the Denver Developmental Screening Test, which aims to monitor and assess infant and preschool-aged children's development. It is comprised of 125 performance-based and parentally reported items split between four functional groups: Personal-Social, Fine Motor-Adaptive, Language, and Gross Motor. Overall, it has high sensitivity but low specificity for identifying children with developmental problems in all age groups examined (age range of 0 to 72 months) and tends to over-refer children to having developmental issues. No evidence was identified pertaining to its validity, reliability, responsiveness, or MCID in patients with CLN2 disease.

The EQ-5D is a generic quality-of-life instrument that may be applied to a wide range of health conditions and treatments. As a generic measure of HRQoL that can capture the net effect of treatment benefits and harms, the EQ-5D provides valuable information from a patient perspective. The EQ-5D-5L consists of an EQ-5D descriptive system and the EQ VAS. While it has been validated and deemed reliable, responsive, and associated with an MCID in the general population, there was no evidence identified in the literature of these attributes for patients with CLN2 disease or any neuronal ceroid lipofuscinosis diseases.

References

1. Brineura (cerliponase alfa injection): solution for infusion for intracerebroventricular infusion 150 mg/5 mL (30 mg/mL) [product monograph]. Toronto (ON): BioMarin Pharmaceutical (Canada) Inc.; 2018.
2. Williams RE, Adams HR, Blohm M, et al. Management strategies for CLN2 disease. *Pediatr Neurol.* 2017;69:102-112.
3. Moore SJ, Buckley DJ, MacMillan A, et al. The clinical and genetic epidemiology of neuronal ceroid lipofuscinosis in Newfoundland. *Clin Genet.* 2008;74(3):213-222.
4. CDR submission: Brineura (cerliponase alfa) 150mg/5 mL solution administered by Intracerebroventricular (ICV infusion) [**CONFIDENTIAL** manufacturer's submission]. Toronto (ON): BioMarin Pharmaceuticals (Canada) Inc.; 2018 Jul 31.
5. Bennett MJ, Rakheja D. The neuronal ceroid-lipofuscinoses. *Dev Disabil Res Rev.* 2013;17(3):254-259.
6. BioMarin Pharmaceuticals (Canada) Inc. response to October 5, 2018 request for additional information regarding Brineura (cerliponase alfa) CDR review**:** historical control description and matched analyses [**CONFIDENTIAL** additional manufacturer's information]. Toronto (ON): BioMarin Pharmaceuticals (Canada) Inc.; 2018.
7. Simpson NA, Wheeler ED, Pearce DA. Screening, diagnosis and epidemiology of Batten disease. *Expert Opin Orphan Drugs.* 2014;2(9).
8. Markham A. Cerliponase alfa: first global approval. *Drugs.* 2017;77(11):1247-1249.
9. Geraets RD, Koh S, Hastings ML, Kielian T, Pearce DA, Weimer JM. Moving towards effective therapeutic strategies for neuronal ceroid lipofuscinosis. *Orphanet J Rare Dis.* 2016;11:40.
10. Schulz A, Ajayi T, Specchio N, et al. Study of intraventricular cerliponase alfa for CLN2 disease. *N Engl J Med.* 2018;378(20):1898-1907.
11. Center for Drug E, Research. Statistical review(s). *Brineura (cerliponase alfa) injection. Company: BioMarin Pharmaceutical Inc. Application No.: 761052. Approval date: 04/27/2017 (FDA approval package)*. Rockville (MD): U.S. Food and Drug Administration (FDA); 2017: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2017/761052Orig1s000TOC.cfm. Accessed 2018 Aug 20.
12. Committee for Medical Products for Human Use. Assessment report: Brineura (cerliponase alfa). *(European public assessment report)*. London (GB): European Medicines Agency; 2017: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/004065/WC500229800.pdf. Accessed 2018 Sep 12.
13. Cerliponase alfa for treating neuronal ceroid lipofuscinosis type 2 *(Highly specialised technology evaluation report ID943)* London (GB): National Institute for Health and Care Excellence; 2018: https://www.nice.org.uk/guidance/gid-hst10008/documents/committee-papers. Accessed 2018 Nov 13.
14. Clinical Study Report: 190-201. A phase 1/2 open-label dose-escalation study to evaluate safety, tolerability, pharmacokinetics, and efficacy of intracerebroventricular BMN 190 in patients with late-infantile neuronal ceroid lipofuscinosis (CLN2) disease [**CONFIDENTIAL** internal manufacturer's report]. Novato (CA): BioMarin Pharmaceutical Inc.; 2016 Apr 27.
15. Clinical Study Report: 190-202. A multicenter, multinational, extension study to evaluate the long-term efficacy and safety of BMN 190 in patients with CLN2 disease [**CONFIDENTIAL** internal manufacturer's report]. Novato (CA): BioMarin Pharmaceutical Inc.; 2017 Oct 13.
16. Clinical Study Report: 190-203. A phase 2, open-label study to evaluate safety, tolerability, and efficacy of intracerebroventricular BMN 190 in patients with CLN2 disease [**CONFIDENTIAL** internal manufacturer's report]. Novato (CA): BioMarin Pharmaceutical Inc.; 2017 Oct 23.
17. Development safety update report: BMN 190 (cerliponase alfa; recombinant human tripeptidyl peptidase-1) [**CONFIDENTIAL** internal manufacturer's report]. Novato (CA): BioMarin Pharmaceutical Inc.; 2018 Jun 13.
18. Clinical Study Report: 190-901. Natural history of late-infantile CLN2 disease: quantitative assessment of disease characteristics, rate of progression, and magnetic resonance imaging findings [**CONFIDENTIAL** internal manufacturer's report]. Novato (CA): BioMarin Pharmaceutical Inc.; 2015 Jun 30.
19. Nickel M, Simonati A, Jacoby D, et al. Disease characteristics and progression in patients with late-infantile neuronal ceroid lipofuscinosis type 2 (CLN2) disease: an observational cohort study. *Lancet Child Adolesc Health.* 2018;2(8):582-590.
20. Varni JW, Burwinkle TM, Seid M, Skarr D. The PedsQL 4.0 as a pediatric population health measure: feasibility, reliability, and validity. *Ambul Pediatr.* 2003;3(6):329-341.
21. Varni JW, Sherman SA, Burwinkle TM, Dickinson PE, Dixon P. The PedsQL Family Impact Module: preliminary reliability and validity. *Health Qual Life Outcomes.* 2004;2:55.
22. Glascoe FP, Byrne KE, Ashford LG, Johnson KL, Chang B, Strickland B. Accuracy of the Denver-II in developmental screening. *Pediatrics.* 1992;89(6 Pt 2):1221-1225.
23. van Reenen M, Janssen B. EQ-5D-5L user guide: basic information on how to use the EQ-5D-5L instrument. Rotterdam (NL): EuroQol Research Foundation; 2015 Apr: https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-5L_UserGuide_2015.pdf. Accessed 2018 Nov 13.
24. Health Quality Council of Alberta. 2014 Alberta population norms for EQ-5D-5L. Calgary (AB): Health Quality Council of Alberta; 2014 Oct: https://d10k7k7mywg42z.cloudfront.net/assets/542f01f2edb2f37083002e54/2014_EQ_5D_5L_report_FINALFINAL.pdf. Accessed 2018 Nov 13.
25. McClure NS, Sayah FA, Xie F, Luo N, Johnson JA. Instrument-defined estimates of the minimally important difference for EQ-5D-5L index scores. *Value Health.* 2017;20(4):644-650.

26. Wyrwich KW, Schulz A, Nickel M, et al. An adapted clinical measurement tool for the key symptoms of CLN2 disease. *J Inborn Errors Metab Screen.* 2018;6.

27. BioMarin Pharmaceuticals (Canada) Inc. comments to CDR request for additional information regarding Brineura (cerliponase alfa) CDR review [**CONFIDENTIAL** additional manufacturer's information]. Toronto (ON): BioMarin Pharmaceuticals (Canada) Inc.; 2018 Nov 23.

28. Ju W, Wronska A, Moroziewicz DN, et al. Genotype-phenotype analyses of classic neuronal ceroid lipofuscinosis (NCLs): genetic predictions from clinical and pathological findings. *Beijing Da Xue Xue Bao Yi Xue Ban.* 2006;38(1):41-48.

29. Worgall S, Kekatpure MV, Heier L, et al. Neurological deterioration in late infantile neuronal ceroid lipofuscinosis. *Neurology.* 2007;69(6):521-535.

30. Kohlschutter A, Schulz A. CLN2 disease (classic late infantile neuronal ceroid lipofuscinosis). *Pediatr Endocrinol Rev.* 2016;13 Suppl 1:682-688.

31. Steinfeld R, Heim P, von Gregory H, et al. Late infantile neuronal ceroid lipofuscinosis: quantitative description of the clinical course in patients with CLN2 mutations. *Am J Med Genet.* 2002;112(4):347-354.

32. Varni JW, Seid M, Knight TS, Uzark K, Szer IS. The PedsQL 4.0 Generic Core Scales: sensitivity, responsiveness, and impact on clinical decision-making. *J Behav Med.* 2002;25(2):175-193.

33. Center for Drug E, Research. Medical review(s). *Brineura (cerliponase alfa) injection. Company: BioMarin Pharmaceutical Inc. Application No.: 761052. Approval date: 04/27/2017 (FDA approval package)*. Rockville (MD): U.S. Food and Drug Administration (FDA); 2017: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2017/761052Orig1s000TOC.cfm. Accessed 2018 Aug 20.

34. Varni JW, Seid M, Kurtin PS. PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations. *Med Care.* 2001;39(8):800-812.

35. Desai AD, Zhou C, Stanford S, Haaland W, Varni JW, Mangione-Smith RM. Validity and responsiveness of the pediatric quality of life inventory (PedsQL) 4.0 generic core scales in the pediatric inpatient setting. *JAMA Pediatr.* 2014;168(12):1114-1121.