

REVIEW

Open Access



Reporting preclinical gene therapy studies in the field of Niemann-Pick type C disease according to the ARRIVE guidelines

Charlotte Laurfelt Munch Rasmussen^{1,2*} , Annette Burkhart¹ , Torben Moos¹ and Louiza Bohn Thomsen^{1*}

Abstract

The lack of essential information when reporting animal studies causing lower reproducibility has been stressed for decades. The ARRIVE (Animal Research: Reporting of In Vivo Experiments) guidelines were first published in 2010, to improve reporting of animal research, making in vivo studies more transparent thereby improving the scientific quality. Regardless of an endorsement from the scientific community, there is still a continuous need to improve animal research reporting, which unfortunately also is the case in the field of Niemann-Pick type C disease (NPC). NPC is a lipid storage disorder, caused by mutations in either the *Npc1* or *Npc2* gene. Despite years of research, no cure for this fatal disease exists. In 2020, an updated version of the ARRIVE guidelines (ARRIVE 2.0), was published, describing the ten most essential elements to be included when reporting pre-clinical studies. Here we systematically reviewed the compliance with the ARRIVE guidelines using the “ARRIVE Essential 10” checklist in a series of pre-clinical studies investigating gene therapy as a treatment strategy for NPC. None of the reviewed papers fulfilled the ARRIVE 2.0 guidelines. Information regarding sample size, randomization, blinding, and statistical methodology was lacking. Hopefully, the newly updated ARRIVE guidelines will aid researchers in planning and publishing in vivo experiments in the future. More awareness of the importance of including these essential items is needed, both from editors, reviewers and researchers, for complete endorsement of the ARRIVE guidelines in the scientific community.

Keywords Niemann-Pick type C disease, NPC, Gene therapy, ARRIVE guidelines, ARRIVE 2.0

*Correspondence:

Charlotte Laurfelt Munch Rasmussen
clmrasmussen@health.sdu.dk
Louiza Bohn Thomsen
lbt@hst.aau.dk

¹ Neurobiology Research and Drug Delivery, Department of Health Science and Technology, Aalborg University, Selma Lagerlöfs Vej 249, DK-9260 Gistrup, Denmark

² The Biomedical Laboratory, Department of Molecular Medicine, University of Southern Denmark, J.B. Winsløvs Vej 23, DK-5000 Odense C, Denmark



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

Niemann-Pick type C (NPC) disease is a severe autosomal recessive, neurovisceral disorder characterized by the accumulation of cholesterol and other lipids in lysosomes [1]. The etiology of NPC disease resides in loss-of-function mutations in one of the two genes, *Npc1* (95% of the cases) or *Npc2* (5% of the cases), resulting in impaired cholesterol transport out of the lysosomes, subsequently altering the lipid metabolism, which is a critical event in the pathogenesis [2–4]. The progressive nature of the early onset of symptoms, including ataxia, dysphagia, dystonia, and dementia, leads to premature death [5].

Unfortunately, there is no cure for NPC disease and to date, Miglustat is the only approved therapeutic option used for decelerating disease progression by inhibiting the synthesis of glycosphingolipids [6–8]. Therefore, the development of new treatment strategies is needed, but due to the blood–brain barrier, the delivery of drugs, proteins, or genes to the brain is hindered [9]. Within the last years, gene therapy has been explored as a potential option for the treatment of NPC disease. Several studies using NPC-deficient (*Npc*^{-/-}) mouse models have demonstrated improvements in survival and ameliorating brain pathology after the administration of viral vectors [10–16]. Animal models for studying disease mechanisms and therapeutic effects, e.g., gene therapy, are of great importance for NPC disease, but an ongoing discussion related to animal models, is that we are in a translational and reproducibility crisis, and there may be several possible explanations for this [17].

In June 2010, the “Animal Research: Reporting of In Vivo Experiments (ARRIVE) guidelines” were published to improve the quality of reporting animal studies in scientific publications [18]. Sufficient information should be reported allowing for a constructive review of animal studies, making the in vivo studies reproducible, and thus preventing unnecessary use of animals [18, 19]. The ARRIVE guidelines provide a checklist describing the minimum information, that should be included when publishing research using animals, e.g., information regarding the experimental animals (number, species, strain, sex, and genetic background), experimental design, and statistical and analytical methods [18]. Ten years later, an update regarding the ARRIVE guidelines was published (ARRIVE 2.0) due to a continuous need for improvement in reporting animal experiments related to the inconsistency in adherence to the original guidelines [20].

Systematic reviews of the literature in a field of research are a valuable tool when designing in vivo studies, identifying research gaps, and subsequently justifying

why new animal studies are needed [20]. In designing our gene therapy study in NPC2-deficient mice (unpublished data), a systematic review of gene therapy studies was an essential part of the planning.

The main objective of this review is, therefore, to evaluate studies assessing the efficacy of gene therapy in mouse models of NPC disease according to the “ARRIVE Essential 10”, i.e., the minimum information to be included in a manuscript [20].

Criteria for inclusion and exclusion of publications

ARRIVE guidelines were first published in June 2010; therefore, only studies published after June 2010 have been included to accommodate the scientific adoption of these new guidelines for publishing animal studies [21]. Both studies using non-viral and viral gene therapy were included with the viral gene therapy studies being limited to adeno-associated virus (AAV), which is the most widely used vector in mouse models of NPC disease [11]. Only data obtained from the treatment studies were included and study design and results regarding gene expression analysis in wild-type mice were excluded. The main focus of this review is on the information essential to assessing the reliability of the presented findings, “ARRIVE Essential 10”, which includes details on the study design, sample size, measures to reduce subjective bias, outcome measures, statistical methods, the animals, experimental procedures, and results [20]. With these inclusion criteria stated, eight articles were included. An overview of these studies can be seen in Table 1.

With this in mind, the selected publications’ compliance with the ARRIVE guidelines was evaluated, focusing on the “Material and Methods” section. Each of the 10 points will be discussed separately, highlighting the importance of including these specific items in a manuscript. Importantly, the intention was not to question the results reported in the included papers, nor to criticize the experimental work performed by the authors, but to provide more awareness of the ARRIVE guidelines [21]. In addition, this review can also provide an overview of the studies in the field of NPC in terms of different vectors used, doses, etc. Subsequently, it will be helpful in the planning of future animal studies using NPC disease models.

Study design

According to the ARRIVE guidelines, the minimum information included when reporting study design is details of the different groups being compared as well as the experimental unit. All studies have used a case–control design, where the *Npc*^{-/-} mice are allocated to a control or treatment group. Most of the studies also include a group of untreated wild-type mice. Even though the study

Table 1 Overview of the included studies

Mouse model	Vector	Promoter	Dosage per mouse	Route of administration	Age at treatment	Effect		Reference
						CNS	P	
Viral gene therapy								
<i>Npc1^{ni/ni}</i> <i>Npc1^{nmf164}</i>	AAV9	GAPDH PGK NPC1 CBA hSYN1 SYN-S SYN-D CAG	1.5 × 10 ¹¹ vg	Intracerebroventricular injection	PN 0–1	+	n.a.	[16]
<i>BALB/cNctr-Npc1^{m1N}/J</i>	AAV9 and AAV-PHPB	EF1α	1.84 × 10 ¹² vg AAV9 1.43 × 10 ¹² vg AAV-PHPB	Retro-orbital injection	PN 24–27	+	+	[11]
<i>FVB.C-Npc1^{m1N}/J</i>	AAV9/3	CMV	1.35 × 10 ¹¹ vg	Intracerebroventricular and intracisternal injections	PN 4–5	+	+	[12]
<i>BALB/cNctr-Npc1^{m1N}/J</i>	AAV9	hSYN1	4.6 × 10 ⁹ vg or 2.5 × 10 ¹¹ vg	Bilateral intracerebroventricular injections	PN 0–1	+	–	[15]
<i>Npc2^{tm1Plob}</i>	AAVrh.10	CAG	10 ¹¹ vg	Intracisternal injection	6 weeks	+	+	[14]
<i>BALB/cNctr-Npc1^{m1N}/J</i>	AAV9	CMV	2.5 × 10 ¹¹ vg	Intracardiac injection	PN 4	+	+	[13]
<i>BALB/cNctr-Npc1^{m1N}/J</i>	AAV9	CamKII or EF1α	2.6 × 10 ¹¹ vg (neonates) 1.3 × 10 ¹² vg (juvenile)	Retro-orbital injection	PN 1–3 PN 20–25	+	+	[10]
Non-viral gene therapy								
<i>BALB/cNctr-Npc1^{m1N}/J</i>	TfRMAb targeted THLs	PDGFB	6 μg plasmid DNA and 15 μg TfRMAb	Tail-vein injection	6–7 weeks	–	+	[22]

AAV Adeno-associated virus, GAPDH glyceraldehyde-3-phosphate dehydrogenase, PGK phosphoglycerate kinase, NPC1 small, truncated version of the endogenous human NPC1 promoter, CBA chicken β -actin, hSYN1 human synapsin 1, SYN-D neuron de-targeted synapsin-1, SYN-S shortened synapsin-1, CAG chicken β -actin promoter with CMV enhancer, EF1 α : elongation factor 1 α , CMV cytomegalovirus, CamKII: calcium/calmodulin-dependent protein kinase II, TfRMAB targeted THLs Transferrin receptor-specific monoclonal antibody conjugated to THLs Trojan horse liposomes, PDGFB human platelet-derived growth factor B, vg viral genomes, PN postnatal day, CNS central nervous system, P peripheral, n.a. not assessed

design is comparable, the *Npc*^{−/−} control group differs. In four out of eight studies, the control group consisted of vehicle (phosphate-buffered saline (PBS), HEPES buffer, or saline) treated *Npc*^{−/−} mice [10–12, 22], whereas in two other studies, no information about the *Npc*^{−/−} control group is available [14, 15]. One study includes untreated *Npc*^{−/−} mice [16]. Two studies have used a control vector expressing a reporter construct, green fluorescent protein (GFP), instead of the protein of interest [10, 13]. One of the aforementioned studies includes both a PBS and an AAV-GFP-treated *Npc*^{−/−} control group [10]. None of the studies report how the *Npc*^{−/−} mice are allocated to the different experimental groups. Only one study provides details of the experimental unit [11].

Sample size

The following information about the sample size should be included; 1) the total number of animals used, 2) the total number in each group, and 3) the total number in each experiment, and finally, 4) how the sample size was

decided should be explained [20]. Especially the last part, i.e., number four, is very important concerning the 3Rs, reduction, refinement, and replacement. The sample size should be sufficient to detect the intended effect without using more animals than necessary. A sample size calculation can accommodate this issue, ensuring that the study is neither underpowered nor overpowered [23]. Even though this information is an essential part of reporting animal studies, only two studies clearly state the total number of animals used [11, 22] (Table 2). For additional specifications of experimental groups, see Table 3. Only one study substantiated how the sample size was determined [11].

Inclusion and exclusion criteria

Expectedly, the number of animals in the NPC studies changes at different time points since *Npc*^{−/−} mice are reaching the end stage of the disease as the study progresses. Exclusion criteria can in this case reflect humane endpoints where animals are euthanized before the

Table 2 Study design

References	Sample size	Group size			Randomization/ Blinding	Study period*	Statistics**
		WT	CTRL	T			
[16]	n.s. [66]	6	6	54	—/—	10 weeks	—
[11]	55	13	13	29	+ / +	9 weeks	+
[12]	n.s. [46]	11	23	12	—/—	11 weeks	—
[22]	36	n.s.	18	18	—/—	5 weeks	N
[15]	n.s. [38]	6	8	24	—/—	Up to 17 weeks [#]	—
[14]	n.s.	n.s.	n.s.	n.s.	—/—	10 weeks	N
[13]	n.s.	n.s.	n.s.	n.s.	—/—	8 weeks	—
[10]	n.s. [44]	0	22	22	—/+	9 weeks	—

WT wild-type, CTRL untreated *Npc*^{-/-} mice, T treated *Npc*^{-/-} mice, *refer to the duration of the study used for immunohistochemical evaluation (from injection to euthanasia), **fulfillment of the criteria stated in the ARRIVE guidelines, n.s. = not specified, e.g. the number of animals is not stated clearly in the method sections or elsewhere, N = statistical section is lacking, [calculated], [#]untreated *Npc*^{-/-} were euthanized at end-stage of the disease at 9 weeks of age

Table 3 Experimental groups

References	Experimental groups						Total
	Wild-type		Untreated <i>Npc</i> ^{-/-} mice		Treated <i>Npc</i> ^{-/-} mice		
	Females	Males	Females	Males	Females	Males	
[16]	9		<i>Npc1^{nih}</i> : 6 <i>Npc1^{nmf164}</i> : 3		<i>Npc1^{nih}</i> AAV9-GAPDH: 6 AAV9-PGK: 6 AAV9-NPC1: 6 AAV9-CBA: 6 AAV9-SYN: 6 AAV9-SYN-S: 6 AAV9-SYN-D: 6 AAV9-CBA: 6 AAV9-CAG: 6 <i>Npc1^{nmf164}</i> AAV9-SYN: 3 AAV9-NPC1: 3		78
[11]	9 weeks: 3 End-stage: 5	9 weeks: 2 End-stage: 3	9 weeks: 1 End-stage: 6	9 weeks: 4 End-stage: 2	9 weeks AAV9: 4 AAV-PHPB: 2 End-stage: AAV9: 4 AAV-PHPB: 5	9 weeks AAV9: 3 AAV-PHPB: 2 End-stage: AAV9: 5 AAV-PHPB: 4	55
[22]	n.s.		8	10	8	10	36
[12]	11		Saline-treated, gel-food: 11 Untreated, gel-food:6 Untreated, standard food: 6		12		46
[15]	6		8		Low dose: 8 High dose: 8 Miglustat: 8		38
[14]	14–29	11–21	6–13	5–12	3–11	6–14	n.s.
[13]	5		7–17		8–16		n.s.
[10]	0		Untreated: 16 AAV9-GFP: 6		Neonates: AAV9-CamKII: 6 Juvenile AAV9-CamKII: 9 AAV9-EF1α: 7		44

Mice used for investigating the biodistribution of the vector are not included in the experimental groups. n.s. not stated

Table 4 Humane endpoints

Parameter	Observation	References
Weight loss	15% loss of body weight after a 24 h period	[16]
	30% weight loss of maximum weight	[11]
	> 15% loss of body weight in a week	[14]
	Loss of 1 g of body weight within a 24 h period	[15]
	Rapid weight loss	[10]
Locomotor function	Reluctance to move, repeated falling to the side during forward ambulation	[11]
	Tremor, abnormal gait, unbalanced	[14, 22]
	Severe loss of motor function	[10]
Skin and fur	Ruffled fur, yellow coat	[22]
Eyes	Palpebral closure/eyes appearing dull rather than bright	[11]
General appearance	Penile prolapse	[14]

predetermined endpoints [20]. Six out of the eight studies had included humane endpoints, whereas, in the remaining two studies, mice were allowed to die due to severe neurological symptoms consequently making the mice unable to, e.g., eat solid food [12, 13]. When assessing the efficacy of the gene therapy strategy, survival is an important parameter to measure. Still, it can be challenging to compare the efficiency of the viral vectors when different humane endpoints were implemented, thus different criteria for euthanasia were used, or when no humane endpoints were used at all (Table 4).

Another important aspect when reporting exclusion criteria is to state whether any animals or data points were excluded from the analysis and why. Only one study reported whether any inclusion or exclusion criteria were set or whether any animals or data points were excluded from the study [11]. In most of the studies, there was an inconsistency in the number used in a specific analysis compared to the total number of mice in the experimental group. For each analysis, the number of animals in each group can differ from the total number of animals used. It is, therefore, important to report the exact value of the number of animals in each experimental group for every analysis performed. For most outcome measures, the number of mice per experimental group is clearly stated, e.g., in figure legends. Still, there is an inconsistency when reporting the number of animals in immunohistochemical evaluation where the number of mice is rarely reported. As stated in the ARRIVE guidelines, when measurements are collected at different time points, a full description of which animals undergo measurement and when should be reported [20]. Clearly stating the actual number of animals used for testing the particular hypothesis allows for transparency and gives the reader an idea of the value of the specific result.

Randomization and blinding

Even though the ARRIVE guidelines have been accessible for more than 10 years, details of, e.g., randomization and blinding are still missing from most publications [20, 24, 25], and this was also evident when reviewing the included papers (Table 3). Only one out of eight studies provided details on both randomization and blinding strategy, although the details concerning randomization were limited [11]. Davidson and colleagues argue that they have employed randomization by using multiple cohorts, but this does not depict the type of randomization used to allocate the mice to either the control or treatment group (e.g., simple, stratified, or block randomization) [26]. On the other hand, the authors state that they minimized confounders by including mice within each treatment group in every cohort, which complies with the ARRIVE guidelines. Besides the aforementioned study, only one other study included information about blinding [10]. These findings are worrisome, based on the knowledge that randomization and blinding reduce bias in animal research [24, 27, 28].

Outcome measures

The ARRIVE guidelines clearly state: “define all outcome measures assessed” [20]. The outcome measures refer to those variables recorded during an experiment, e.g., used to assess the effect of an intervention. Therefore, it is important to report all outcome measures assessed, otherwise, there is a risk of only including statistically significant data [20].

The pathological hallmarks of the disease in *Npc*^{-/-} mice are progressive weight loss, hepatosplenomegaly, and neurodegeneration resulting in a fatal outcome of the disease [29–31]. In addition, the cerebellum is severely affected and presents with Purkinje cell loss, reactive

Table 5 Body weight

Peak BW untreated <i>Npc</i> ^{-/-}	Peak BW treated <i>Npc</i> ^{-/-}	Vector	Administration	References
n.a.*				[16]
6 weeks	14 weeks	AAV-PHP	Systemic	[11]
	8 weeks	AAV9		
7 weeks	15 weeks	AAV9/3	CNS	[12]
6 weeks	12 weeks	AAV9 low dose	CNS	[15]
	16 weeks	AAV9 high dose		
8 weeks	18 weeks (f)	AAVrh.10	CNS	[14]
	16 weeks (m)			
n.a.				[13]
6 weeks	8 weeks	AAV9-CamKII	Systemic	[10]
	12 weeks	AAV9-EF1 α		
6 weeks	6 weeks	TFRMAb targeted THLs	Systemic	[22]

BW bodyweight, f females, m males, n.a. not assessed, *peak body weight is not assessed, CNS central nervous system

Table 6 Survival

Median survival untreated <i>Npc</i> ^{-/-}	Median survival treated <i>Npc</i> ^{-/-}	Lifespan extension (%)	Vector	Longest survival	References
76 days	263 days	246	AAV-NPC1	275 days	[16]
72 days	234 days	225	AAV-PHP	> 1 year (AAV-PHP)	[11]
	112 days	56	AAV9		
75 days	205 days	173	AAV9/3	310 days	[12]
75 days	116.5 days	55	AAV9 low dose	126 days	[15]
	158 days	111	AAV9 high dose	168 days	
112 days (f)	280 days (f)	150	AAVrh.10	280 days (f)	[14]
98 days (m)	168 days (m)	71			
71 days	94 days	32	AAV9	100 days	[13]
69 days	103 days (j)	49	AAV9-CamKII	140 days (j)	[10]
	166 days	141	AAV9-EF1 α	300 days	
75 days*	75 days	n.s.	TFRMAb targeted THLs	77 days	[22]

f = females, m = males (euthanized due to the development of penile prolapse, starting at 24 weeks of age), j = juvenile at the time of injection, * = data are reported as mean, n.s. = non-significant

microglia, and astrogliosis [14, 32]. These parameters are, therefore, also the primary focus when evaluating the efficiency of gene therapy in mice models of NPC disease. One of the first visible symptoms in NPC-deficient mice is weight loss, and this parameter is also evaluated in all studies [10–12, 14–16, 22] except one [13] (Table 5). All studies testing the effect of an AAV vector showed a delay in the onset of weight loss. Another important outcome measure assessed is survival, which was also evaluated in all included studies (Table 6). However, there are several important issues to consider when comparing NPC disease studies. Also noted by Hughes and colleagues, differences in vector serotype, vector dose, timing and route of administration, transgene cassettes,

and humane endpoints make direct comparison between studies almost impossible [15].

Statistical methods

To fulfill the requirements stated by the ARRIVE guidelines, details of the statistical methods used as well as statistical software should be provided. Furthermore, whether the data met the assumptions for the statistical approach should also be described [20]. Based on these criteria, only one of the included papers fulfills these requirements [11]. Two of the papers did not include a section about the statistical methods used in the study [14, 22]. The remaining papers do not provide sufficient details about the statistical methods, and none of

these papers provided detail about the assumptions of the underlying data making it difficult for the reader to assess the appropriateness and suitability of the methods used. Assumptions for using parametric tests are, e.g., that the data analyzed are continuous, follow a normal distribution, and the variances between different groups are similar [20]. Usually, parametric tests have higher statistical power compared to non-parametric tests [19, 44]. Nevertheless, they require that the aforementioned assumptions are met. The use of incorrect assumptions can result in false positive results, consequently making invalid conclusions. Inappropriate reporting of statistical methodology in combination with the aforementioned experimental bias is a common cause of poor study design [45].

Experimental animals

Species-appropriate details should be provided including species, strain, sex, age, weight, the origin of the animals, health status, genetic modification, and genotype [20].

Mouse models for NPC

Most NPC disease studies to date use the BALB/cNctr-*NPC1^{m1N}/J* (*Npc^{nih}*) murine model, which was also evident when reviewing the literature about gene therapy as a treatment strategy in *Npc1*^{-/-} mice. Six out of eight studies included in this review used the *Npc^{nih}*, whereas only one used the FVB.C-*NPC1^{m1N/J}* mouse model. One study included both *Npc^{nih}* and *Npc^{nmlf164}* mice to test the effect of the gene therapy [16]. Only one study to date has evaluated viral gene therapy in *Npc2*^{-/-} mice [14]. All included studies provide sufficient data on the species, strain, genotype, genetic modification, the origin of the animals as well as the age of the mice at the beginning of the study, which complies with the ARRIVE guidelines. A single study provided details on health status [13]. An important detail when reporting animal studies is the sex of the species used. In a lot of different research areas, sex-dependent differences related to lifespan, immune activation, and response to therapy have been observed [46–49], which are also evident in studies using NPC-deficient mice [8, 14, 50, 51]. The NPC disease phenotype is, e.g., more severe in *Npc*^{-/-} male mice with shorter survival, more dramatic weight loss, and a more progressive Purkinje cell degeneration compared to female *Npc*^{-/-} mice [37, 52, 53]. Despite these well-known differences, the sex of the animals was only specified in three of the eight articles (Table 3). Sex and species-specific parameters such as strain and age contribute to variance between studies [42]. Therefore,

when certain details, such as sex, are not reported, it may pose challenges to the reproducibility and translatability of the study.

Experimental procedures

The reporting of the experimental procedures should be described in enough detail to make it possible for other researchers to replicate them [20]. For all studies included, the AAV subtypes, including promotor, vehicle, doses, volumes, and route of administration were described in detail (Table 1), as well as the rationale for these decisions. However, the specific procedure for the administration of viral vectors and euthanasia lacks several details. Stereotaxic surgery was used in three of the included studies [12, 14, 15], and a thorough description of the surgical procedure would therefore be expected. According to the ARRIVE guidelines, information about surgery includes a specific description of the procedure (incl. sham surgery), anesthetics used (drug, dose, concentration, route of administration), analgesia (pre-and post), aseptic techniques, and parameters monitored during anesthesia (e.g., assessment of the surgical anesthetic plane), duration of the procedure and anesthesia. All these aspects are another essential part of the previously mentioned 3Rs, specifically refinement. In one of the studies stereotaxic surgeries were performed in mice at 6 weeks of age [14]. The only information provided was the anesthesia used (3% isoflurane) and the injection volume of the viral vector. Unfortunately, e.g., analgesia regimen was not mentioned, although this is an important part of securing animal welfare in research. Two studies used retro-orbital injections in weanlings as the route of vector administration, which requires anesthesia [54], but information about anesthesia is lacking [10, 11]. The same limitations were seen when reporting the use of euthanasia. Four out of eight studies did not provide any details of the anesthetics used and/or the procedure used for euthanasia [13, 15, 16, 22]. Although minimizing pain and distress in laboratory animals is the consensus when performing animal studies, the type and use of analgesics during experimental surgery in rodents are often not reported [55], which was unfortunately also the case in the papers included in this review.

Results

As for the statistical methods, the reporting of results, including descriptive statistics for each experimental group (e.g., mean and standard deviation (SD)) [20], are also limited or inconsequent in the reviewed studies. One of the included studies has not provided the reader with any statistical information about the results, neither

in the text nor in the figure legends [12]. Four of the included studies use SD and standard error of the mean (SEM) interchangeably [10, 14, 15, 22]. Importantly, SEM provides information about how accurate the sample mean is and whether the mean is a precise estimate of the population mean. SD gives information about the variability within the group. Thus, data should be summarized using SD [56].

Furthermore, SD is used for calculating the effect size, the last parameter that should be reported according to the ARRIVE guidelines. None of the studies has included information about the effect size, which unfortunately often is the case in animal research [20]. Nevertheless, the effect size together with its confidence interval is an important part of the statistical information providing the reader with information about the estimated magnitude of differences between the experimental groups, the precision of the estimate, as well as whether the findings are biologically relevant [57].

Summary

At the moment, thousands of journals are recommending the ARRIVE guidelines [20], and more will definitely follow. However, there are still some challenges we need to overcome. Three of the included papers refer specifically to ARRIVE guidelines, and state that the study design is based on the ARRIVE guidelines 2.0 [11], or that the animal studies were conducted according to the ARRIVE guidelines and recommendations [15, 16]. None of the remaining papers provide such information.

Interestingly, four journals clearly state that animal studies should comply with ARRIVE guidelines [10, 14–16, 22]. Two of these journals explicitly state that when preparing the manuscript, the ARRIVE guidelines must be followed [10, 15, 16], whereas the two other journals report that the author must state or confirm that the experiments comply with the ARRIVE guidelines, e.g., in the method Sect. [14, 22]. Two other journals encourage following the ARRIVE guidelines [11, 13]. In one journal, no information is provided in the submission guidelines [12]. It is, however, unknown when this editorial information was last updated. More responsibility from the journals to ensure the needed information is reported is probably needed, which has also been emphasized by, e.g., The Journal of Bone and Joint Surgery with the publication of the editorials; “JBJS Will Require Adherence to ARRIVE Guidelines for Animal Research to Reduce Bias and Improve Quality of Reporting” [58]. In this editorial, they highlight that all manuscripts reporting animal studies will include an annotated checklist of the ARRIVE guidelines from the 1st of January 2020. Furthermore, they recognize the importance of their role as editors, e.g., concerning

the problem of publication bias, and will encourage the publication of both positive and negative results [58]. Even though several journals endorse these guidelines, essential information is still lacking from scientific publications [59], which also was the case in the papers included in this review. Surprisingly, in a survey of Swiss researchers, more than half of the authors whose last publication was in a journal endorsing the ARRIVE guidelines had never heard of these guidelines before [60]. The majority of the researchers included in the survey were engaged in biomedical or medical research [60]. In addition, fulfilling a checklist at submission was not sufficient to adhere to the ARRIVE guidelines [61]. More awareness is needed, and this requires first and foremost adoption of these guidelines by the journals [58], but also an acceptance from researchers of why these guidelines are of importance and the risks associated with introducing bias in animal research when the guidelines are not followed [20, 60].

Conclusions

Even though the first ARRIVE guidelines were published more than 10 years ago, the reporting of animal research is still not following these guidelines. This is unfortunately not only the case for studies in NPC disease but for many other research areas using animals to study, e.g., cancer, stroke, infectious, and cardiovascular diseases [17, 21, 24, 62–65].

Animal models are essential for developing new treatment strategies for incurable diseases and learning more about the pathogenesis of NPC disease. During the last twenty years, the research in the field of NPC disease has increased considerably, even though a cure for the disease is still lacking [8].

When reviewing and evaluating whether the papers (all published from 2017 and forth) were published in compliance with the ARRIVE guidelines, none fulfilled the minimum required information to be included when reporting animal research. Most of the papers lacked information in several important areas including justification of sample size, blinding, randomization, and statistical methodology. Without this information, readers and reviewers cannot assess the reliability of the findings. This emphasizes the need for more comprehensive reporting in animal research and that detailed descriptions of the experimental procedures and animal models must be available to ensure that results can be evaluated, interpreted, and replicated in the future.

Despite these guidelines being recommended by more than a thousand journals [20], there is still a lack of adherence to the ARRIVE guidelines in several research areas [66]. The successful implementation of the ARRIVE guidelines requires increased awareness from the entire

scientific community. With the publication of ARRIVE 2.0 in 2020, where they have highlighted “the essential 10” to help scientists, editors, and reviewers focus on the minimum information needed to be reported when publishing pre-clinical studies [20], future research can hopefully become more transparent. Furthermore, the guidelines can be helpful for the researcher during the planning and conducting of animal studies, ensuring rigorous study design and collecting sufficient information for the preparation of the manuscript [20].

Raising more awareness of the ARRIVE guidelines and the importance of reporting adequate information on study design, including randomization and blinding, in combination with conducting systematic reviews, can hopefully contribute to solving the translational and reproducibility challenges in preclinical research in the future, as also stated by Alstrup and Sonne [17].

One step at a time, but when “the essential 10” is fully implemented in the backbone of animal research, all 21 items included in the ARRIVE guidelines can hopefully also be implemented.

Abbreviations

AAV	Adeno-associated virus
ARRIVE	Animal research: reporting of In vivo experiments
GFP	Green fluorescent protein
NPC	Niemann-Pick type C
PBS	Phosphate-buffered saline
SD	Standard deviation
SEM	Standard error of the mean

Acknowledgements

Not applicable.

Author contributions

Conceptualization: CLMR and LBT; investigation: CLMR, visualization: CLMR, writing—original draft preparation: CLMR, writing—review and editing: CLMR, AB, TM, and LBT; supervision, AB, TM, and LBT.

Funding

This work was kindly funded by Lundbeck Foundation—Research Initiative on Blood–Brain Barriers and Drug Delivery (Grant no. 2013-14113), Hørslev-Fonden, Læge Sophus Carl Emil Friis og Hustru Olga Doris Friis Legat.

Availability of data and materials

Data sharing is not applicable as no new data were created or analyzed in this review.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 20 March 2024 Accepted: 21 November 2024

Published online: 06 May 2025

References

- Ko DC, Milenkovic L, Beier SM, Manuel H, Buchanan JA, Scott MP. Cell-autonomous death of cerebellar purkinje neurons with autophagy in niemann-pick type C disease. *PLOS Genet*. 2005;1(1):81–95.
- Vanier MT. Niemann-Pick disease type C. *Orphanet J Rare Dis*. 2010;5:16.
- Patterson MC, Hendriksz CJ, Walterfang M, Sedel F, Vanier MT, Wijburg F. Recommendations for the diagnosis and management of Niemann-Pick disease type C: An update. *Mol Genet Metab*. 2012;106(3):330–44.
- Vanier MT, Gissen P, Bauer P, Coll MJ, Burlina A, Hendriksz CJ, et al. Diagnostic tests for Niemann-Pick disease type C (NP-C): A critical review. *Mol Genet Metab*. 2016;118(4):244–54.
- Wraith JE, Baumgartner MR, Bembi B, Covanis A, Levade T, Mengel E, et al. Recommendations on the diagnosis and management of Niemann-Pick disease type C. *Mol Genet Metab*. 2009;98(1–2):152–65.
- Patterson MC, Garver WS, Giugliani R, Imrie J, Jahnova H, Meaney FJ, et al. Long-term survival outcomes of patients with Niemann-Pick disease type C receiving miglustat treatment: a large retrospective observational study. *J Inher Metab Dis*. 2020;43(5):1060–9.
- Patterson MC, Vecchio D, Prady H, Abel L, Wraith JE. Miglustat for treatment of Niemann-Pick C disease: a randomised controlled study. *Lancet Neurol*. 2007;6(9):765–72.
- Pallottini V, Pfriger FW. Understanding and treating niemann-pick type C disease: models matter. *Int J Mol Sci*. 2020;21(23):8979.
- Pardridge WM. The blood-brain barrier: bottleneck in brain drug development. *NeuroRx*. 2005;2(1):3–14.
- Chandler RJ, Williams IM, Gibson AL, Davidson CD, Incao AA, Hubbard BT, et al. Systemic AAV9 gene therapy improves the lifespan of mice with Niemann-Pick disease, type C1. *Hum Mol Genet*. 2017;26(1):52–64.
- Davidson CD, Gibson AL, Gu T, Baxter LL, Deverman BE, Beadle K, et al. Improved systemic AAV gene therapy with a neurotrophic capsid in Niemann-Pick disease type C1 mice. *Life Sci Alliance*. 2021;4(10):e202101040.
- Kurokawa Y, Osaka H, Kouga T, Jimbo E, Muramatsu K, Nakamura S, et al. Gene therapy in a mouse model of Niemann-Pick Disease Type C1. *Hum Gene Ther*. 2021;32(11–12):589–98.
- Xie C, Gong X-M, Luo J, Li B-L, Song B-L. AAV9-NPC1 significantly ameliorates Purkinje cell death and behavioral abnormalities in mouse NPC disease. *J Lipid Res*. 2017;58(3):512–8.
- Markmann S, Christie-Reid JJ, Rosenberg JB, De BP, Kaminsky SM, Crystal RG, et al. Attenuation of the Niemann-Pick type C2 disease phenotype by intracisternal administration of an AAVrh.10 vector expressing Npc2. *Exp Neurol*. 2018;306:22–33.
- Hughes MP, Smith DA, Morris L, Fletcher C, Colaco A, Huebeker M, et al. AAV9 intracerebroventricular gene therapy improves lifespan, locomotor function and pathology in a mouse model of Niemann-Pick type C1 disease. *Hum Mol Genet*. 2018;27(17):3079–98.
- Hughes MP, Nelvagal HR, Coombe-Tennant O, Smith D, Smith C, Mas-saro G, et al. A novel small NPC1 promoter enhances AAV-mediated gene therapy in mouse models of Niemann-Pick type C1 disease. *Cells*. 2023;12(12):1619.
- Alstrup AKO, Sonne C. 3Rs as part of preclinical neuropsychiatric translational crisis, and ARRIVE guidelines as part of solution. *Acta Neuropsychiatr*. 2019;31(6):348–9.
- Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLOS Biol*. 2010;8(6):e1000412.
- Festing MFW, Altman DG. Guidelines for the design and statistical analysis of experiments using laboratory animals. *ILAR J*. 2002;43(4):244–58.
- Percie du Sert N, Hurst V, Ahluwalia A, Alam S, Avey MT, Baker M, et al. The ARRIVE guidelines 2.0: updated guidelines for reporting animal research. *J Physiol*. 2020;598(18):3793–801.
- Gulin JEN, Rocco DM, García-Bournissen F. Quality of reporting and adherence to ARRIVE guidelines in animal studies for chagas disease pre-clinical drug research: a systematic review. *PLoS Negl Trop Dis*. 2015;9(11):e0004194.
- Jiang D, Lee H, Pardridge WM. Plasmid DNA gene therapy of the Niemann-Pick C1 mouse with transferrin receptor-targeted trojan horse liposomes. *Sci Rep*. 2020;10:13334.
- Richter V, Muche R, Mayer B. How much confidence do we need in animal experiments? Statistical assumptions in sample size estimation. *J Appl Anim Welf Sci*. 2018;21(4):325–33.

24. Kousholt BS, Praestegaard KF, Stone JC, Fick Thomsen A, Johansen TT, Ritskes-Hoitinga M, et al. Reporting quality in preclinical animal experimental research in 2009 and 2018: a nationwide systematic investigation. *PLoS ONE*. 2022;17(11): e0275962.
25. Macleod MR, Lawson McLean A, Kyriakopoulou A, Serghiou S, de Wilde A, Sherratt N, et al. Risk of bias in reports of in vivo research: A focus for improvement. *PLoS Biol*. 2015;13(10):1–12.
26. Kang M, Ragan BG, Park JH. Issues in outcomes research: an overview of randomization techniques for clinical trials. *J Athl Train*. 2008;43(2):215.
27. Hirst JA, Howick J, Aronson JK, Roberts N, Perera R, Koshiaris C, et al. The need for randomization in animal trials: an overview of systematic reviews. *PLoS ONE*. 2014;9(6): e98856.
28. Van Eenige R, Verhave PS, Koemans PJ, Tiebosch IACW, Rensen PCN, Kooijman S. RandoMice, a novel, user-friendly randomization tool in animal research. *PLoS ONE*. 2020;15(8): e0237096.
29. Sleat DE, Wiseman JA, El-Banna M, Price SM, Verot L, Shen MM, et al. Genetic evidence for nonredundant functional cooperativity between NPC1 and NPC2 in lipid transport. *Proc Natl Acad Sci U S A*. 2004;101(16):5886–91.
30. Beltroy EP, Richardson JA, Horton JD, Turley SD, Dietschy JM. Cholesterol accumulation and liver cell death in mice with Niemann-Pick type C disease. *Hepatology*. 2005;42(4):886–93.
31. Nielsen GK, Dagnaes-Hansen F, Holm IE, Meaney S, Symula D, Andersen NT, et al. Protein replacement therapy partially corrects the cholesterol storage phenotype in a mouse model of Niemann-Pick type C2 disease. *PLoS ONE*. 2011;6(11): e27287.
32. Li H, Repa JJ, Valasek MA, Beltroy EP, Turley SD, German DC, et al. Molecular, anatomical, and biochemical events associated with neurodegeneration in mice with Niemann-Pick type C disease. *J Neuropathol Exp Neurol*. 2005;64(4):323–33.
33. Sarna JR, Larouche M, Marzban H, Sillitoe RV, Rancourt DE, Hawkes R. Patterned Purkinje cell degeneration in mouse models of Niemann-Pick type C disease. *J Comp Neurol*. 2003;456(3):279–91.
34. Luan Z, Saito Y, Miyata H, Ohama E, Ninomiya H, Ohno K. Brainstem neuropathology in a mouse model of Niemann-Pick disease type C. *J Neurol Sci*. 2008;268(1–2):108–16.
35. Geberhiwot T, Moro A, Dardis A, Ramaswami U, Sirrs S, Marfa MP, et al. Consensus clinical management guidelines for Niemann-Pick disease type C. *Orphanet J Rare Dis*. 2018;13(1):50.
36. Rasmussen CLM, Thomsen LB, Heegaard CW, Moos T, Burkhardt A. The Npc2Gt(LST105)BygNya mouse signifies pathological changes comparable to human Niemann-Pick type C2 disease. *Mol Cell Neurosci*. 2023;126: 103880.
37. Vöikar V, Rauvala H, Ikonen E. Cognitive deficit and development of motor impairment in a mouse model of Niemann-Pick type C disease. *Behav Brain Res*. 2002;132(1):1–10.
38. Whitney ER, Kemper TL, Rosene DL, Bauman ML, Blatt GJ. Calbindin-D28k is a more reliable marker of human Purkinje cells than standard Nissl stains: a stereological experiment. *J Neurosci Methods*. 2008;168(1):42–7.
39. Buard I, Pfrieger FW. Relevance of neuronal and glial NPC1 for synaptic input to cerebellar Purkinje cells. *Mol Cell Neurosci*. 2014;61:65–71.
40. Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The extent and consequences of P-hacking in science. *PLOS Biol*. 2015;13(3): e1002106.
41. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS ONE*. 2008;3(8): e3081.
42. Ten SR. Points to improve reproducibility and translation of animal research. *Front Behav Neurosci*. 2022;16: 869511.
43. Tsilidis KK, Panagiotou OA, Sena ES, Aretouli E, Evangelou E. Evaluation of excess significance bias in animal studies of neurological diseases. *PLoS Biol*. 2013;11(7):1001609.
44. Nahm FS. Nonparametric statistical tests for the continuous data: the basic concept and the practical use. *Korean J Anesthesiol*. 2016;69(1):8.
45. Smith AJ, Clutton RE, Lilley E, Hansen KEA, Brattelid T. PREPARE: guidelines for planning animal research and testing. *Lab Anim*. 2018;52(2):135–41.
46. Borsotti C, Danzi NM, Nauman G, Hölzl MA, French C, Chavez E, et al. HSC extrinsic sex-related and intrinsic autoimmune disease-related human B-cell variation is recapitulated in humanized mice. *Blood Adv*. 2017;1(23):2007–18.
47. Notta F, Doulatov S, Dick JE. Engraftment of human hematopoietic stem cells is more efficient in female NOD/SCID/IL-2R γ c-null recipients. *Blood*. 2010;115(18):3704–7.
48. Karp NA, Mason J, Beaudet AL, Benjamini Y, Bower L, Braun RE, et al. Prevalence of sexual dimorphism in mammalian phenotypic traits. *Nat Commun*. 2017;8(1):1–12.
49. Abd-Elrahman KS, Albaker A, De Souza JM, Ribeiro FM, Schlossmacher MG, Tiberi M, et al. A β oligomers induce pathophysiological mGluR5 signaling in Alzheimer's disease model mice in a sex-selective manner. *Sci Signal*. 2020;13(662):2494.
50. Houben T, Bitorina AV, Oligschläger Y, Jeurissen MLJ, Rensen S, Köhler SE, et al. Sex-opposed inflammatory effects of 27-hydroxycholesterol are mediated via differences in estrogen signaling. *J Pathol*. 2020;251(4):429.
51. Bascuran-Castillo EC, Erickson RP, Howison CM, Hunter RJ, Heidenreich RH, Hicks C, et al. Tamoxifen and vitamin E treatments delay symptoms in the mouse model of Niemann-Pick C. *J Appl Genet*. 2004;45(4):461–7.
52. Bianconi SE, Hammond DI, Farhat NY, Dang Do A, Jenkins K, Cougnoux A, et al. Evaluation of age of death in Niemann-Pick disease, type C: Utility of disease support group websites to understand natural history. *Mol Genet Metab*. 2019;126(4):466–9.
53. Cougnoux A, Fellmeth M, Gu T, Davidson CD, Gibson AL, Pavan WJ, et al. Maternal immune activation modifies the course of Niemann-Pick disease, type C1 in a gender specific manner. *Mol Genet Metab*. 2020;129(2):165–70.
54. Steel CD, Stephens AL, Hahto SM, Singletary SJ, Ciavarrá RP. Comparison of the lateral tail vein and the retro-orbital venous sinus as routes of intravenous drug delivery in a transgenic mouse model. *Lab Anim*. 2014;43(3):96–9.
55. Stokes EL, Flecknell PA, Richardson CA. Reported analgesic and anaesthetic administration to rodents undergoing experimental surgical procedures. *Lab Anim*. 2009;43(2):149–54.
56. Barde MP, Barde PJ. What to use to express the variability of data: Standard deviation or standard error of mean? *Perspect Clin Res*. 2012;3(3):113.
57. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol Rev*. 2007;82(4):591–605.
58. Bauer TW, Bechtold JE, Swiontkowski MF. JBJS will require adherence to ARRIVE guidelines for animal research to reduce bias and improve quality of reporting. *J Bone Jt Surg - Am*. 2019;101(21):1891–3.
59. Leung V, Rousseau-Blass F, Beauchamp G, Pang DSJ. ARRIVE has not ARRIVED: support for the ARRIVE (animal research: reporting of in vivo experiments) guidelines does not improve the reporting quality of papers in animal welfare, analgesia or anesthesia. *PLoS ONE*. 2018;13(5): e0197882.
60. Reichlin TS, Vogt L, Würbel H. The researchers' view of scientific rigor—survey on the conduct and reporting of In vivo research. *PLoS ONE*. 2016;11(12): e0165999.
61. Hair K, Macleod MR, Sena ES. A randomised controlled trial of an intervention to improve compliance with the ARRIVE guidelines (IICARUS). *Res Integr Peer Rev*. 2019;4(1):1–17.
62. Holman C, Piper SK, Grittner U, Diamantaras AA, Kimmelman J, Siegerink B, et al. Where have all the rodents gone? The effects of attrition in experimental research on cancer and stroke. *PLoS Biol*. 2016;14(1):e1002331.
63. Reynolds PS, Garvan CW. Preclinical research reporting in shock: room for improvement. *Shock*. 2021;55(5):573–80.
64. Ramirez FD, Motazedian P, Jung RG, Di Santo P, Macdonald ZD, Moreland R, et al. Methodological rigor in preclinical cardiovascular studies. *Circ Res*. 2017;120(12):1916–26.
65. Kousholt BS, Praestegaard KF, Stone JC, Thomsen AF, Johansen TT, Ritskes-Hoitinga M, et al. Reporting of 3Rs approaches in preclinical animal experimental studies—a nationwide study. *Anim*. 2023;13(19):3005.
66. Avey MT, Moher D, Sullivan KJ, Fergusson D, Griffin G, Grimshaw JM, et al. The devil is in the details: incomplete reporting in preclinical animal research. *PLoS ONE*. 2016;11(11): e0166733.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.