Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Journal Pre-proof

Evaluation of non-clinical toxicity studies of COVID-19 vaccines

N.K.M. Schilder, B. Tiesjema, P.T. Theunissen, K. Oude Rengerink, J.W. van der Laan

PII: S0273-2300(23)00106-X

DOI: https://doi.org/10.1016/j.yrtph.2023.105438

Reference: YRTPH 105438

To appear in: *Regulatory Toxicology and Pharmacology*

Received Date: 23 December 2022

Revised Date: 15 June 2023

Accepted Date: 21 June 2023

Please cite this article as: Schilder, N.K.M., Tiesjema, B., Theunissen, P.T., Rengerink, K.O., van der Laan, J.W., Evaluation of non-clinical toxicity studies of COVID-19 vaccines, *Regulatory Toxicology and Pharmacology* (2023), doi: https://doi.org/10.1016/j.yrtph.2023.105438.

CRediT author statement

N.K.M. Schilder: Methodology, Software, Formal analysis, Validation, Investigation, Data curation, Writing – original draft and - review and editing, Visualization, Project administration. **B. Tiesjema**: Validation, Resources, Writing – review and editing. **P.T. Theunissen**: Validation, Writing – review and editing. **K. Oude Rengerink**: Methodology, Validation, Writing – review and editing. **J.W. van der Laan**: Conceptualization, Methodology, Validation, Resources, Writing – review and editing, Supervision, Project administration, Funding acquisition.

Evaluation of non-clinical toxicity studies of COVID-19 vaccines

N.K.M. Schilder[1,3], B. Tiesjema[1], P. T. Theunissen[1], K. Oude Rengerink[2], J.W. van der Laan[1,3]

1 Section on Pharmacology, Toxicology and Kinetics. Medicines Evaluation Board, Graadt van Roggenweg 500, 3531AH Utrecht, The Netherlands.

2 Section of Methodology Working Group. Medicines Evaluation Board, Graadt van Roggenweg 500, 3531AH Utrecht, The Netherlands.

3 Division of Toxicology, Leiden Academic Center for Drug Research, Einsteinweg 55, 2333CC Leiden, The Netherlands.

Corresponding author

Jan Willem van der Laan

Medicines Evaluation Board

Graadt van Roggenweg 500

3531AH Utrecht

jw.vd.laan@cbg-meb.nl

+31652756424

Contact information other authors:

N.K.M. Schilder: natalia.schilder@hotmail.com

B. Tiesjema: b.tiesjema@cbg-meb.nl

P. Theunissen: p.theunissen@cbg-meb.nl

K. Oude Rengerink: k.ouderengerink@cbg-meb.nl

Word count abstract: 195

Word count text: 5530

Word count references: 729

Abstract

In this study we evaluated the outcomes of non-clinical toxicity studies of various SARS-CoV-2 vaccines produced with different manufacturing technologies, with focus on Repeated Dose Toxicity (RDT) and Developmental and Reproductive Toxicity (DART) studies. We found that RDT and DART studies at doses relevant for human treatment showed no adverse effects while remaining observations were expected findings including local reactogenicity, immune response and macroscopic findings at the injection site. We have also reviewed the European Medicines Agency (EMA) nonclinical assessment reports for market authorization. Regardless of utilized vaccine manufacturing technology EMA assessment of the non-clinical studies consisted most frequently of comments related to study design, species selection and missing data. Sponsors have often submitted platform studies (vaccine studies with the same technology/construct but using other antigens) as supplementary data. Animal model-based toxicity testing has shown rather small effects, which have been never serious adverse effects. The translational value to support clinical development is mainly to inflammatory effects, indicative of the primary action of the vaccines. From a 3R perspective supportive platform technology data consisting of previously executed RDT and DART studies from the same platform technology are encouraged to be implemented in the vaccine assessment process.

Keywords

COVID-19 vaccines

Developmental and Reproductive Toxicity (DART) study

Repeated Dose Toxicity (RDT) study

Supportive platform technology

Vaccine concept

1. Introduction

Since the emergence of SARS-CoV-2, scientists have aimed at rapid and effective vaccine development. Currently, there are five main vaccine technologies on which a vaccine can be based and that have been employed for COVID-19 vaccine development: inactivated, live attenuated, subunit/peptide based, vector based and nucleic acid based vaccines[1].

Viral infections are prototypic species-specific diseases, which makes clinical translation of animal models difficult[2]. In case of SARS-CoV-2, it has been shown that mice and rats are not susceptible due to the inability of the virus to entry rodents cells via the rodent orthologue receptor of the human ACE2 entry receptor[3]. However, they still do show an immunologic response which is sufficient to enable its use for nonclinical safety testing in line with the WHO recommendations[4]. Non-clinical toxicity testing for vaccines is, therefore, routinely performed in rats and rabbits (usually one species is sufficient). Use of these species has the advantage that historical background data are available to aid interpretation of possible findings[4].

Vaccines differ from small molecule medicines and biological therapeutics in that for vaccines applying the intended full human dose to the individual animal is sufficient to ensure a safety margin (at least in respect to adult humans), while for small molecule pharmaceuticals the dose levels are being increased up to at least a minimal toxic level or a sufficient exposure margin, which thus will have a greater chance to lead to adverse effects.

Furthermore, the dosing regimen, including dose selection and interval, route of administration, timing of evaluation of end-points as well as species historical control data is pivotal for the interpretation of the study[4]. The WHO Guideline on Preclinical Testing of Vaccines[4] recommends the use of a dose leading to a maximal antibody responses.

These considerations are relevant as well in relation to the Developmental and Reproductive Toxicity (DART) studies, which have also been conducted with COVID-19 vaccines, as vaccination of the full adolescent and adult population, including women-of-child-bearing potential, is indicated. In line with the current WHO guidelines[4,5] and the revised ICH S5(R3)[6] a DART study with vaccines is designed as

an Embryo-Fetal Developmental study (EFD) with an early premating starting dose to include endpoints from the Fertility and Early Embryonic Developmental design (FEED), as well as with long post-natal survival, sometimes up to 45-60 days to include endpoints of a Pre-Post-Natal Developmental (PPND) study. A single species is usually considered to be sufficient to provide important information on potential toxicity of the vaccine and safety of the product during human pregnancy.

One of the strategies in the compression of SARS-CoV-2 vaccine development timelines consisted of providing supportive platform toxicity studies for first-in-human (FIH) clinical trials prior to submitting product-specific studies belonging to an identical vaccine platform[7]. The EMA has described supportive platform studies as '*a collection of technologies that have, in common, the use of a 'backbone' carrier or vector that is modified with a different active substance or set of active substances for each vaccine derived from the platform*'[8]. This strategy was recommended by the International Coalition of Medicines Regulatory Authorities (ICMRA) in March 2020 and includes the use of toxicology data (repeat dose toxicity (RDT), biodistribution studies) and clinical data from a well characterized platform.

For utilization of supportive platform data justification for the absence of product-specific toxicity studies prior to FIH is needed[9]. In all cases, additional animal data was still considered necessary for characterization of the immune response prior to execution of non-clinical toxicity studies. Animal data on immunogenicity of the candidate vaccines are available prior execution of FIH and non-clinical toxicity studies. Product-specific studies are expected to follow after initiation of Phase I clinical trials, justified by rationale supported by data that is provided by the manufacturer[10].

There has been public pressure to reduce the regulatory load on the use of animal testing for vaccine development and to apply 3R principles in this area[2]. In an recent overview of five COVID-19 vaccines,

it was concluded that usually the packages are reasonably complete in accordance to the current guidelines and not reduced because of time pressure on the development of the vaccine[11]. Our question is a level deeper, i.e. were these nonclinical data rather needed to decide about the safety of the vaccines being developed or could we have saved time and animals without compromising human safety?

With this in mind, the aim of this study is to evaluate RDT and DART (EFD) studies of COVID-19 vaccines considered from a European perspective. In the first part of this manuscript the species selection, study design and outcomes of the RDT and DART studies of COVID-19 will be discussed. In the second part of the manuscript the EMA assessment of the submitted RDT and DART studies and the impact that supportive vaccine platform studies have had on the review process and authorization of the COVID-19 vaccines will be discussed.

2. Materials and Methods

Information was extracted from the Common Technical Documentation of from 8 COVID-19 vaccines submitted for marketing authorization application (MAA) to EMA until 13-01-2022; Comirnathy, COVID-19 mRNA vaccine [BNT162b2] (Biontech-Pfizer); Spikevax, COVID-19 mRNA vaccine [mRNA-1273 LNP](Moderna); Vaxzevria,COVID-19 Vaccine AstraZeneca [ChAdOx1-S, vector] (AstraZenaca); Jcovden, COVID-19 Vaccine [Ad26 vector], (Janssen), Nuvaxovid, recombinant spike protein, adjuvanted [CoV2373] (Novavax); CureVac COVID-19 mRNA vaccine (CVnCoV); Vidprevtyn COVID-19 vaccine, recombinant adjuvanted (Sanofi); Valneva , SARS-CoV-2 virus [beta-propiolactone inactivated, adjuvanted with AS03], Valneva. This information is not always publicly available, and excerpts have been published in the European Public Assessment Reports, which are in the public domain

All, but one, vaccines had been authorized for use in the European Union by December 2022 (n=7, two are mRNA vaccines, two vector based, two subunit based, and one inactivated). One application (CVnCoV) was withdrawn in October 2021. All detailed information has been derived from the internal medicinal product database of the Medicines Evaluation Board (MEB) (Utrecht, the Netherlands), the

MEB being part of the European Network of Regulatory Authorities on Human Pharmaceuticals. Only the RDT and DART studies were included in the evaluation.

The DART studies consisted of Embryo-Fetal Developmental (EFD) toxicity studies with early administration in dams, a few weeks before mating (representing some aspects of a FEED study, at least in females) and a long follow-up postnatally (representing some aspects of a Pre- and Postnatal Developmental (PPND) study). All studies submitted to the EMA by the applicants and EMA assessment documents were reviewed.

Choice of species selection, study design and outcomes

The design and outcomes of the studies were provided by the applicant in the MAA. In total, 10 vaccine-specific RDT studies and 17 supportive RDT studies, 8 vaccine-specific DART studies and 1 supportive DART study were identified. Table 1 provides an overview of the conducted non-clinical studies for the evaluated COVID-19 vaccines. The 8 evaluated vaccines comprise four different vaccine technologies, i.e. nucleic acid based, vector based, inactivated and subunit-based vaccines, indicated by their names.

Nonclinical studies as referred to in the following guidelines were summarized and scored:

1) EMA Guideline on Preclinical Pharmacological and Toxicological Testing of Vaccines[12]
2) WHO Guidelines on non-clinical evaluation of vaccines[4]
3) FDA Developmental and Reproductive Toxicity studies for vaccines[13]
4) WHO Guidelines on the nonclinical evaluation of vaccine adjuvants and adjuvanted vaccines[5]
5) ICH S5 (R3) Guideline on Reproductive toxicity (2020)[6]

In Repeated Dose Toxicity studies, the applicants reported the common general toxicity endpoint such as: clinical observations, body weight, food consumption, body temperature, ophthalmology, haematology, urinalysis, clinical chemistry (incl. liver enzymes), albumin/globulin ratio, metabolic

state and inflammatory markers (e.g. C-reactive protein). The summarized product-specific effects were scored by frequency with the following semi-quantitative scores:

Table 1 Scores of Effects (terms used in the dossiers)
(**0**) No effect
(**1**) Incidental/one/two/(sporadically in) few/some cases/within range of historical control data/within reference vales/within normal physiological range/common in species/strain/age/age-related
(**1**) Very slight to slight/Slightly/marginally/(mostly) mild(ly)/minimal(ly)/minimal to mild/transient
(**2**) Moderate* (only indication of increase and decrease)
(**3**) Highly frequent/(noted for nearly) all cases/most cases
(**3**) Significant (as indicated in the document)

All scores were designated in appropriate context, i.e. based on presence of the words of frequency stated above, and on surrounding data concerning the effect or the prevalence in animals. Time of effect (i.e. during study period or recovery period) was also considered.

'Moderate' refers to effects not accompanied by any word of frequency, solely and indication of increase or decrease of the concerning parameter. Here, the same approach applies to a 'moderate' score being designated in appropriate context.

In DART studies, the applicants reported in addition the developmental toxicity endpoints: (FEED) Mating and fertility performance, estrous cycle evaluation; (EFD) gravid uterine weight, implantation sites, live and dead conceptuses, fetal body weight, and fetal evaluations (variations, malformations); (PPND) F0 Parturition observations, clinical observations, female necropsy/macro- and microscopy, F1 physical development, necropsy and microscopy.

In order to compare the effects in the DART studies observed in the product-specific COVID-19 vaccine to effects observed in RDT studies, these were categorized as well and assigned a semi-quantitative score based on frequency. An overview can be found in Supplementary Table 1.

2.1 Principal Component Analysis

In order to compare the effects for the list of parameters between the different vaccines (as described above), a Principal Component Analysis (PCA) was performed using Python software (version 3.9.12 Scikit-learn package version 1.0.2). PCA shows the inter-individual variability in a two-dimensional way, representing comparison of the effects (or effect scores) of the RDT and DART studies, respectively, as explained above.

First, the attributed scores were selected for a PCA. To do this, in case of a range in scoring value the most dominant and/or principal finding was chosen to represent the concerning category. Additionally, as in multiple product-specific studies some endpoints were not performed or not specified, the Multiple Imputation by Chained Equations (MICE) algorithm was applied to deal with missing data by using available data of all the effect scores[14]. Two vaccine studies; one mRNA (RDT) and one subunit (DART) were upfront excluded from analysis (only for the PCA analysis) due to too much missing data (availability <60%) for correct estimations. Lastly a dimensional reduction was performed in order to visually compare the studies on their similarities in effects. The 2D representation is described by so-called principal components (PC1 and PC2).

2.2 EMA assessment documents

The assessment documents compiled by the EMA were screened for relevant review and assessment information concerning Repeated dose toxicity (RDT) studies and Development and Reproductive toxicity (DART). Comments concerning supportive studies were included. EMA assessment of RDT studies belonging to the vaccine containing whole inactivated viruses was limited to one.

Selected from 144 documents, this information was derived from:

- European Public assessment Reports (EPARs) – publicly available
- (Co-)Rapporteur (Rolling Review) assessments

- CHMP member comments/peer reviews
- Others (e.g. study summaries, overviews, meeting minutes):

Comments of these documents were summarised and categorised based on theme of content using Atlas.ti (22.0.6) software[15]. The following main categories were used: discussion observation, study design, relevance provided data, species selection, study description (clarity/interpretation of description on study design), missing information/data, GLP compliance (only RDT studies), vaccine batch used (only RDT studies). Other, less frequent commentary can be found in Appendix Table A, together with Atlas.ti summaries of EMA assessment.

3. Results

3.1 PART I – Evaluation of species selection and study design

First it will be described how the species selection and study design of the non-clinical toxicity studies of COVID-19 (with a focus on RDT and DART) affected the outcomes of the non-clinical toxicity studies of the various SARS-CoV-2 vaccine technologies.

In this analysis eight COVID-19 vaccines have been included, of which six were fully evaluated with respect to the nonclinical packages in relation to their use in the European Union and two were under rolling review at the final date of data gathering of this study.

Noteworthy, the evaluation of most of the product-specific COVID-19 RDT studies is supported by platform studies with products with similar vaccine technology. The supportive studies are based on previous research on other vaccines including parainfluenza viruses (hMPV, PIV3, RSV), Rabies, Chikungunya, Malaria, HIV and Ebola. Evaluation of local tolerance was conducted in most cases as part of the RDT studies. Additional toxicity studies, such as single dose toxicity studies, genotoxicity, allergenicity and (supplementary) immunogenicity studies, were occasionally provided by the applicants. For some vaccines additional studies were provided to support the use of on novel excipients, such as lipid nanoparticles and recombinant human albumin (rHA). For DART studies

additional platform studies with similar vaccine technology have not been provided as a standard approach.

3.1.2 Selection of Species

The applicants have utilized rats (Sprague Dawley or Wistar Han) and/or rabbits (New Zealand White) for RDT as well as for DART studies (Table 2). One applicant has used mice (CD-1) in this respect.

The WHO guideline (2013) explicitly indicates that it is not necessary to conduct the nonclinical safety study in the same animal species used for proof-of-concept or nonclinical pharmacology studies, although it should be immunologically responsive to the vaccine antigen. Next to the choice of animal species, the study duration and dosing regimen (including dose selection) is pivotal in designing and interpreting animal toxicity studies[4]. Figure 1 shows the study duration and dosing regimen of the conducted product-specific RDT and DART studies.

3.1.3 Outcomes of RDT studies

In general, the RDT studies were consistent and adhered to the proposed use in humans concerning the duration of administration and the intended clinical dosing regimen[5]. To compare the effects observed in the product-specific COVID-19 vaccine RDT studies, these were categorized and assigned a semi-quantitative score based on frequency (see section 2. Materials and Methods). The categories with scores can be found back in Supplementary Table 1. In the RDT studies effects within a score range of -2 to 2 in frequency, indicated as moderate, were not uncommon. Additionally, a considerable number of findings for local reactogenicity, induction of immune response and macroscopic effects at the injection site were designated with score 3, indicated as highly frequent or significant. In some cases spleen organ weight gain were observed to be highly frequent or significant (score 3) as well. The vast majority of these effects were of reversible nature.

RDT

Species	Vaccine concept (total nr. of studies)	Age of animals	Number of animals (females + males)
Wister Han Rat	mRNA1 (2)	9 weeks	90
		54/60 days (~8 weeks)	241
	mRNA3 (1)	9/11 weeks	110
	InAct2 (1)	6/7 weeks	60
Sprague Dawley Rat	mRNA2 (1)	7 weeks	40
New Zealand White Rabbit	Sub1 (2)	12/14 weeks	40
		21 weeks	160
	Sub2 (1)	~15 weeks	90
	Adeno2 (1)	13/15 weeks	40
CD-1 mouse	Adeno (1)	7/8 weeks	170

DART

Species	Vaccine concept (total nr. of studies)	Age of animals	Number of animals (females only)
Wistar Han Rat	mRNA1 (1)	11 weeks	132
	mRNA3 (1)	7 weeks	60
	InAct2 (1)	(at least) 10 weeks	88
Sprague Dawley Rat	mRNA2 (1)	74 days (~11 weeks)	88
	Sub2 (2)	(at least 8 weeks)	4 (pilot study)
		6.6 weeks	150
New Zealand White Rabbit	Sub1 (1)	16/19 weeks	132
	Adeno2 (1)	5/6 months	92
CD-1 mouse	Adeno1 (1)	11 weeks	64

Table 2: Species selection in vaccine specific (A) Repeated dose Toxicity (RDT) studies and (B) Developmental and Reproductive Toxicity (DART) studies of COVID-19 vaccine candidates. Vaccines are anonymized based on their vaccine concept: mRNA(x) (un)modified mRNA-based vaccine encapsulated in lipid nanoparticle (LNP). Adeno(x) = recombinant replication-incompetent adenoviral vector encoding the spike protein antigen of SARS-CoV-2. InAct(x) = inactivated SARS-CoV-2 virus formulated with adjuvant. Sub(x) = subunit: modified spike protein of SARS-CoV-2 formulated with adjuvant.

Clinical chemistry (i.e. amino-acid transferases such as AST and ALT, and the A/G ratio) was affected at a minimal level in several studies with vaccines from different technologies (mRNA, Adeno-vector, adjuvanted vaccin), These effects appear to correlate to the inflammatory response such as the excretion of acute phase proteins (CRP) in the liver.

3.1.4 Outcomes of DART studies

For DART studies, a difference was observed between timing of mRNA vaccine administration during the early gestational period. For mRNA based vaccines 2, 3, and 1 the test-article was administered on gestation days 1, 3 or 9, respectively and in case of mRNA based vaccines 2 and 3 the vaccine was administered outside of the proposed interval during early organogenesis as recommended by the ICH S5 (R3) guideline on reproductive toxicity[6]. Of note, the timing of vaccination during early organogenesis has been a point of discussion during the reviewing

269

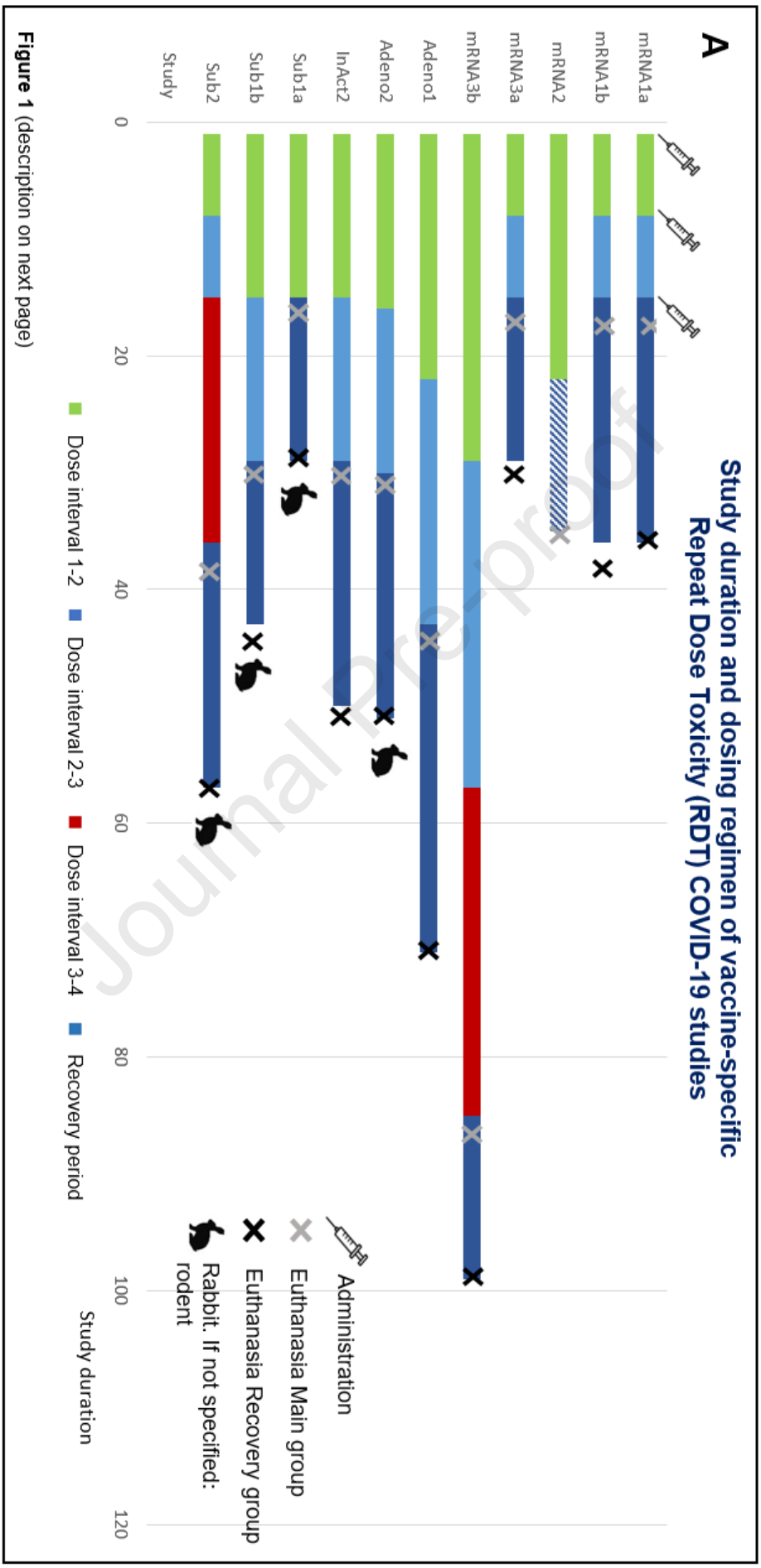


Figure 1 (description on next page)

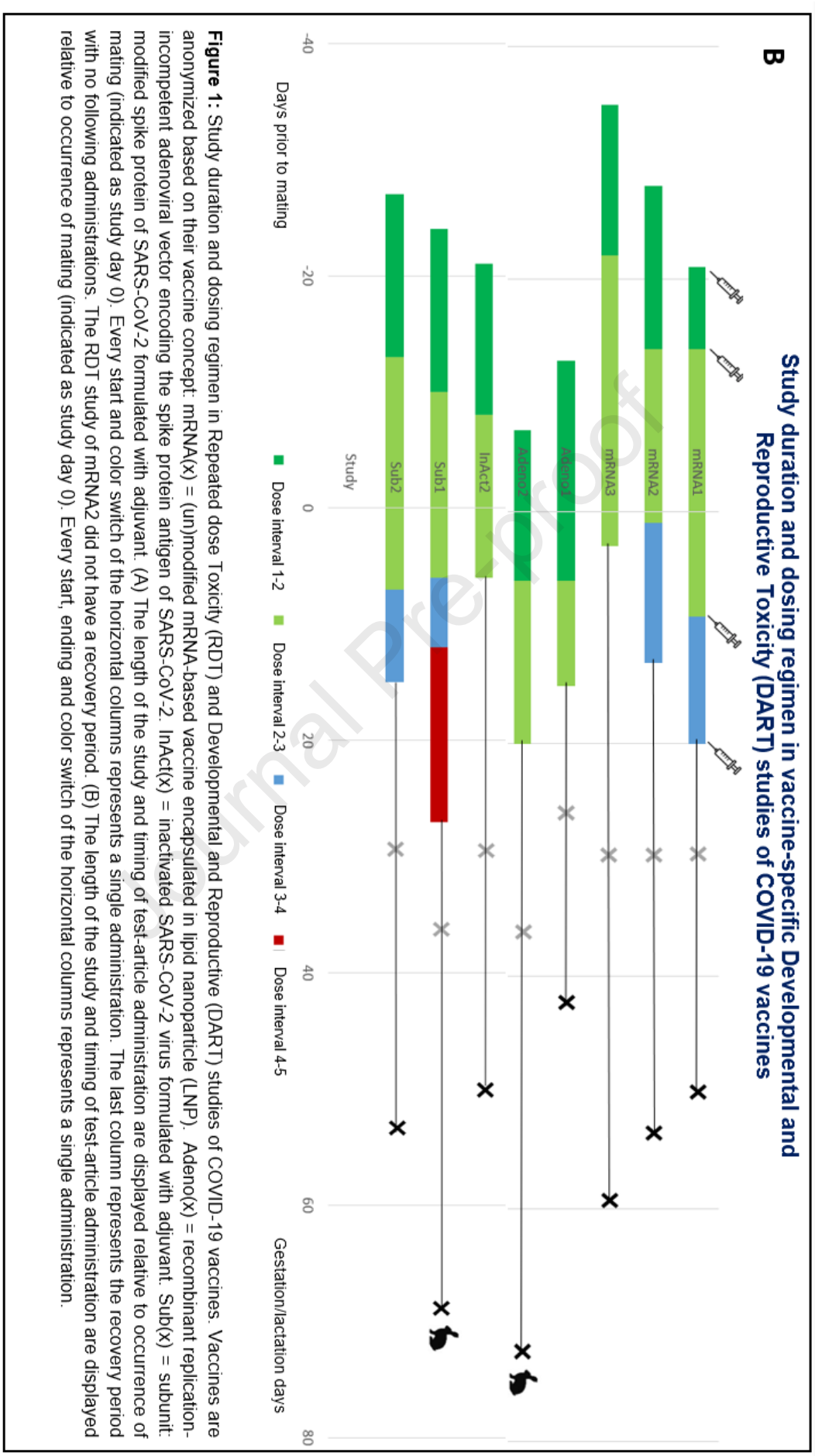


Figure 1: Study duration and dosing regimen in Repeated dose Toxicity (RDT) and Developmental and Reproductive (DART) studies of COVID-19 vaccines. Vaccines are anonymized based on their vaccine concept: mRNA(x) = (un)modified mRNA-based vaccine encapsulated in lipid nanoparticle (LNP). Adeno(x) = recombinant replication-incompetent adenoviral vector encoding the spike protein antigen of SARS-CoV-2. InAct(x) = inactivated SARS-CoV-2 virus formulated with adjuvant. Sub(x) = subunit: modified spike protein of SARS-CoV-2 formulated with adjuvant. (A) The length of the study and timing of test-article administration are displayed relative to occurrence of mating (indicated as study day 0). Every start and color switch of the horizontal columns represents a single administration. The last column represents the recovery period with no following administrations. The RDT study of mRNA2 did not have a recovery period. (B) The length of the study and timing of test-article administration are displayed relative to occurrence of mating (indicated as study day 0). Every start, ending and color switch of the horizontal columns represents a single administration.

process of mRNA vaccines, which will be elaborated on later in the discussion (see 4. Discussion and Conclusion).

The occurrence of effects in DART studies was in most cases within the range -1 and 1, meaning their nature was incidental, slight or common in the species. Some exceptions existed for fetal evaluations including minor variations in the EFD phase, parturition effects and female macro and microscopic observations in F0 females and body weight, physical development and macro and microscopic observations in F1 pups.

3.1.5 Inter-individual variability based on frequency of observations

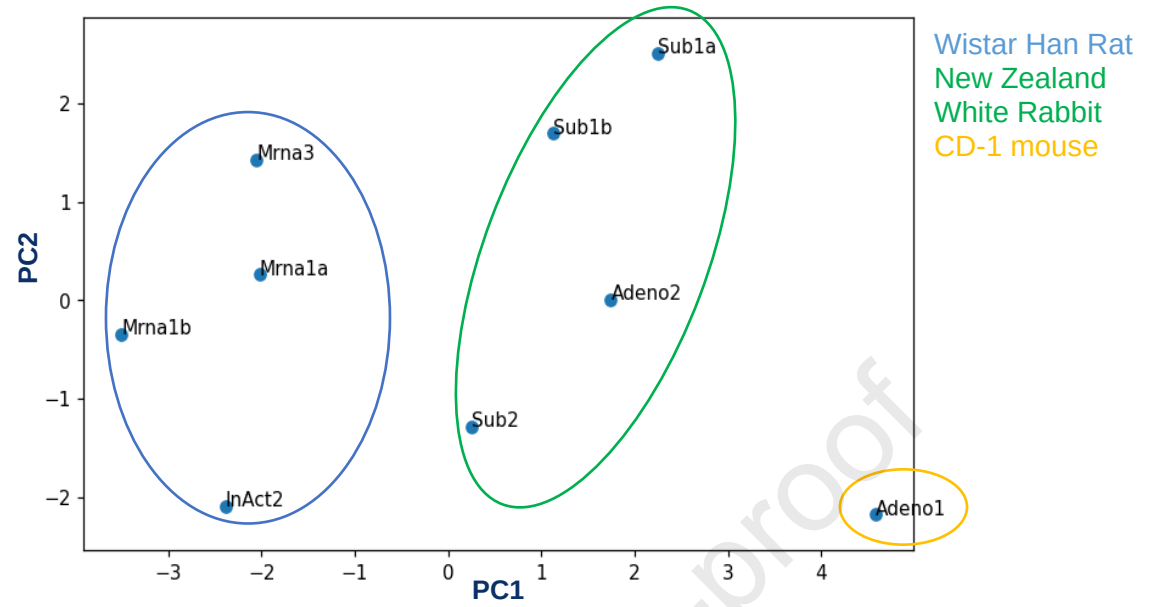
Figure 2 shows the Principal Component Analysis (PCA) based on the semi-quantitative assessment of the COVID-19 vaccine effects. The PCA graphs are displayed in a two-dimensional way representing comparison of the various scores of the RDT studies. The principal components (PC1 and PC2) on the axes describe the largest and second largest variations from the toxicity studies, which together account for approximately 60 -70 % of the varied influences of the original categories[16]. As expected, it was observed that studies conducted with similar vaccine technology showed a high similarity in findings. For instance, Sub1a and Sub1b show little variance for both PC1 and PC2.

Based on the PCA data, differences in dosing regimen between the COVID-19 vaccines did not result in the variation of the effects for the toxicity studies. However, it can be acknowledged that the vaccines for which animal strains were selected high similarity in findings of both the RDT and DART studies regardless of the vaccine technology was observed.

For example, the subunit and adenoviral vector-based vaccines for which the New Zealand White (NZW) rabbit was chosen as species in the RDT studies showed high similarities in observed effects. This high species-determined similarity partly overlapped with the designated study variation, however, the first was stronger.

298

A **Principal component analysis (PCA) of Repeated dose Toxicity (RDT) studies of COVID-19 vaccines**



B **Principal component analysis (PCA) of Developmental and Reproductive Toxicity (DART) studies of COVID-19 vaccines**

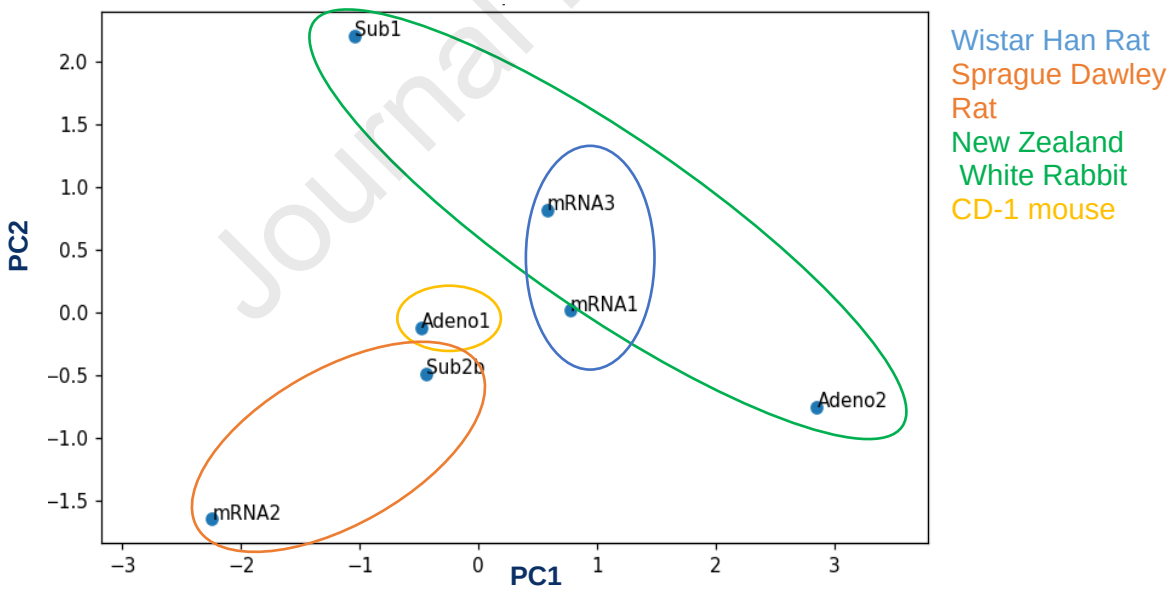


Figure 2: Principal component analysis (PCA) based on the semi-quantitative assessment of the COVID-19 vaccine findings of the Repeated dose Toxicity (RDT) studies (A) and Developmental and Reproductive Toxicity (DART) studies (B). Vaccines are anonymized based on their vaccine concept: mRNA(x) = (un)modified mRNA-based vaccine encapsulated in lipid nanoparticle (LNP). Adeno(x) = recombinant replication-incompetent adenoviral vector encoding the spike protein antigen of SARS-CoV-2. InAct(x) = inactivated SARS-CoV-2 virus formulated with adjuvant. Sub(x) = subunit: modified spike protein of SARS-CoV-2 formulated with adjuvant. One subunit (RDT), one mRNA (DART) and one adenovirus vector based (DART) vaccine are excluded from the analysis due to a high amount of not performed/not specified data. The PC values for explained variance ratio are: RDT studies: PC1 = 44.08%, PC2 = 17.24%, PC3 = 12.02%, DART studies: PC1 = 43.04%, PC2 = 25.05%, PC3 = 17.15%.

299 **3.1.6 RDT Effects contributing to the variability**

300 In order to understand the relevance of PC1 and PC2 in the inter-individual variability of the vaccines,
301 the major contributing categories (i.e. marked by the highest absolute coefficients that exerts an effect
302 on the PCs) were identified. The PC1 in RDT was mostly determined by clinical chemistry parameters
303 (A/G ratio, ion levels and bilirubin), gross necropsy, findings at the injection site and adrenal weights.
304 For PC2 in RDT the variation was determined mostly by haematology (specifically concerning red blood
305 cells), adrenal weights and clinical chemistry (metabolic state/hydration status).
306 The findings that contributed mostly to the variation in the RDT studies were, however, ,
307 immunostimulatory and of reversible nature and other trends were absent.
308

309 **3.1.6 DART Effects contributing to the variability**

310 For the DART studies, the most contributing variation of PC1 was determined by body weight and food
311 consumption, of both the dams during PPND assessment as well as the offspring (pre- and
312 postweaning). For PC2, the variation was determined mostly by parturition effects, female
313 necropsy/macro- and microscopic findings and maternal body weight gain.
314 To investigate the relevance of the DART variation contributors, it was determined how the most
315 prevalent contributors compared to the moment of vaccine administration. It was also examined
316 whether there is a correlation between the dams body weight (gain) and body weight (gain) of the
317 pup, as studies showed that nursing mothers experienced minimal disruption of breastfeeding after
318 COVID-19 vaccination[17].
319 In general, changes in dam body weight (gain) correlated with changes in dam food consumption. More
320 than half of the vaccines showed transient decrease in body weight gain following vaccine
321 administration. For some vaccines changes in pup body weight gain have been observed, these were
322 however of transient nature or present within treated groups with incidentally small litter size resulting
323 in increased access to maternal resources. It is therefore unlikely to be caused by changes in lactation
324 following vaccination.

In general, no adverse findings on offspring development were noted for any of the vaccines.

3.2 PART II – EMA regulatory assessment per vaccine technology

In the second part of the manuscript the EMA assessment of the submitted RDT and DART studies and the impact that supportive vaccine platform studies have had on the review process and authorization of the COVID-19 vaccines will be discussed.

3.2.1 Main themes of RDT and DART study assessment

For analysis of the commentary themes of the study assessments, the assessment documents compiled by the EMA concerning RDT and DART studies were qualitatively analysed. The analysis of the rolling review EMA assessment has been summarized in Table 3.

The RDT studies assessment of all vaccine technologies consisted mainly of commentary related to study design, species selection and missing data (figure 4A). Discussions of observations played a large role in the commentary of RDT studies. The three main categories of the EMA assessment concerning DART studies also mostly consisted of commentary related to study design, species selection and missing data (figure 4B). Overall, the categories deducted from the assessment of the vaccines showed to be less diverse compared to the assessment of RDT studies.

3.2.2 Diverse nature of RDT study assessment

EMA comments were diverse concerning RDT study design. Categories concerning study design recurrent in all vaccine technologies focused on the dosage chosen, lack of inclusion of a control group (e.g. lipid nanoparticle, adjuvant or preservative only group (i.e. 2-phenoxyethanol)) or shortcomings in histopathology examination of tissue samples. EMA assessment related to species selection referred predominantly to total volume of administration feasible in the utilized animal and the induction of the immune response, for example connected with inflammatory markers and the inflammatory response of the innate immune system.

RDT

Vaccine / concept Category	mRNA vaccines	Adenoviral vaccines	Whole Inactivated vaccines	Subunit vaccines
Missing data	6	5	8	2
Relevance	4	4	1	3
Species selection	6	7	1	7
Study design	18	3	8	12
Discussion of observations	17	1	4	5
Used batch	1	-	-	2
Study description	2	-	2	-
GLP	3	-	3	-

DART

Vaccine / concept Category	mRNA vaccines	Adenoviral vaccines	Subunit vaccines
Missing data	8	3	1
Relevance	4	-	1
Species selection	5	5	5
Study design	6	3	3
Discussion of observations	3	4	2

Table 3: EMA assessment per vaccine concept for Developmental and Reproductive Toxicity studies of COVID-19 vaccines. The commentary of the EMA assessment is derived from European Public assessment Reports (EPARs) (publicly available), (Co-)Rapporteur (Rolling Review) assessments, CHMP member comments/peer reviews and others (e.g. study summaries, overviews, meeting minutes). Number refers to number of total comments within the vaccine concept on the concerning categories. Sp. Selection = species selection, Discussion ob. = discussion observations.

Comments of EMA concerning missing information focused on lack of immunogenicity data in the pivotal RDT, whereas companies refer to provided supportive studies. Additionally, other information stated as missing concerned the dose and design of these supportive studies. In the assessment of the RDT studies the lack of data on potential differences between SARS-CoV-2 variants and its consequences for the safety of vaccines based upon the original Wuhan strain was frequently discussed.

Furthermore, comments have been made regarding of the potential effects on liver enzymes (i.e. increased activity of gamma-glutamyl transferase (GGT), increased aspartate transferase (AST) levels together with increased liver weights or hepatocellular vacuolation) in relation to local inflammatory responses.

3.2.3 Divergent views of applicants on assessment of DART studies

Concerning assessment of DART studies, the EMA comments on the study design referred most frequently to either justification (e.g. referral to previous studies) or shortcomings (e.g. number of timepoints included) in the immunogenicity potential and the dose regimen of the studies. Regulatory comments on the dose regimen was mostly intertwined with concerns regarding the timing of administration in relation to mother-pup antibody transfer during lactation, but also placental transfer during gestation. Comments on the dose used being an excess of the human dose was provided to support translation to the clinical situation and in case of deviation in the dose of the administered adjuvant.

Comments on the species selection for DART studies were aimed at susceptibility of the species to SARS-CoV-2, as well as discussion on interspecies differences in placental antibody transfer in the part of gestation representing the second half in human pregnancy. Various applicants had different views on fetal- and pup-maternal IgG ratios during gestation and lactation periods. Additionally, comments also stated that allometric rules do not apply to local immune responses as induced by vaccines, making exposure weight adjusted exposure margins irrelevant for local use.

The EMA assessment on missing information in the DART studies was diverse among the vaccine technologies. In all three mRNA vaccines included, data on mRNA placental transfer and milk excretion was mentioned as important missing information. Comments on the adenoviral vector-based vaccines frequently pointed out that information was lacking regarding potential embryo-fetal toxicity of the adenovirus carrier. With regard to the subunit vaccine, the assessment focused solely on missing information regarding possible effects of the antigen:adjuvant stoichiometry on immunogenicity. Next to the three most discussed effects (study design, species selection and missing data), comments on the relevance of the data and discussion of effects was also profoundly present in the EMA assessment.

3.2.4 Comment categories regarding supportive platform studies

As indicated before, applicants of the various vaccine technologies have used platform data to support their vaccine development. Aiming at reduction of animal testing, we investigated the regulatory emphasis on the supportive data in the EMA assessment (Figure 3). Concerning the RDT studies, supportive studies were mentioned especially during the assessment of effects of the product-specific studies. This was specifically the case for one mRNA vaccine. Additionally, supportive studies were supplied in the context of study design, missing information or when signifying relevance of the data. Logically, comparison with product-specific studies was applied mostly for justification of study design. Comparison to other SARS-CoV-2 vaccines the assessment was engaged primarily in the discussion of effects and relevance of data in the assessment of RDT studies.

Concerning DART studies, supportive studies (usually studies from the same company with different antigens but with the same adjuvants) were provided mainly when missing information was being discussed. The supportive studies complemented the missing information on neurological developmental effects, potential of reprotoxic/teratogenic effects and the antigen:antigen ratio effect on immunogenicity. A smaller proportion of the delivered supportive studies were mentioned in the context of discussions of observations and study design. Opposite to the EMA assessment concerning

414

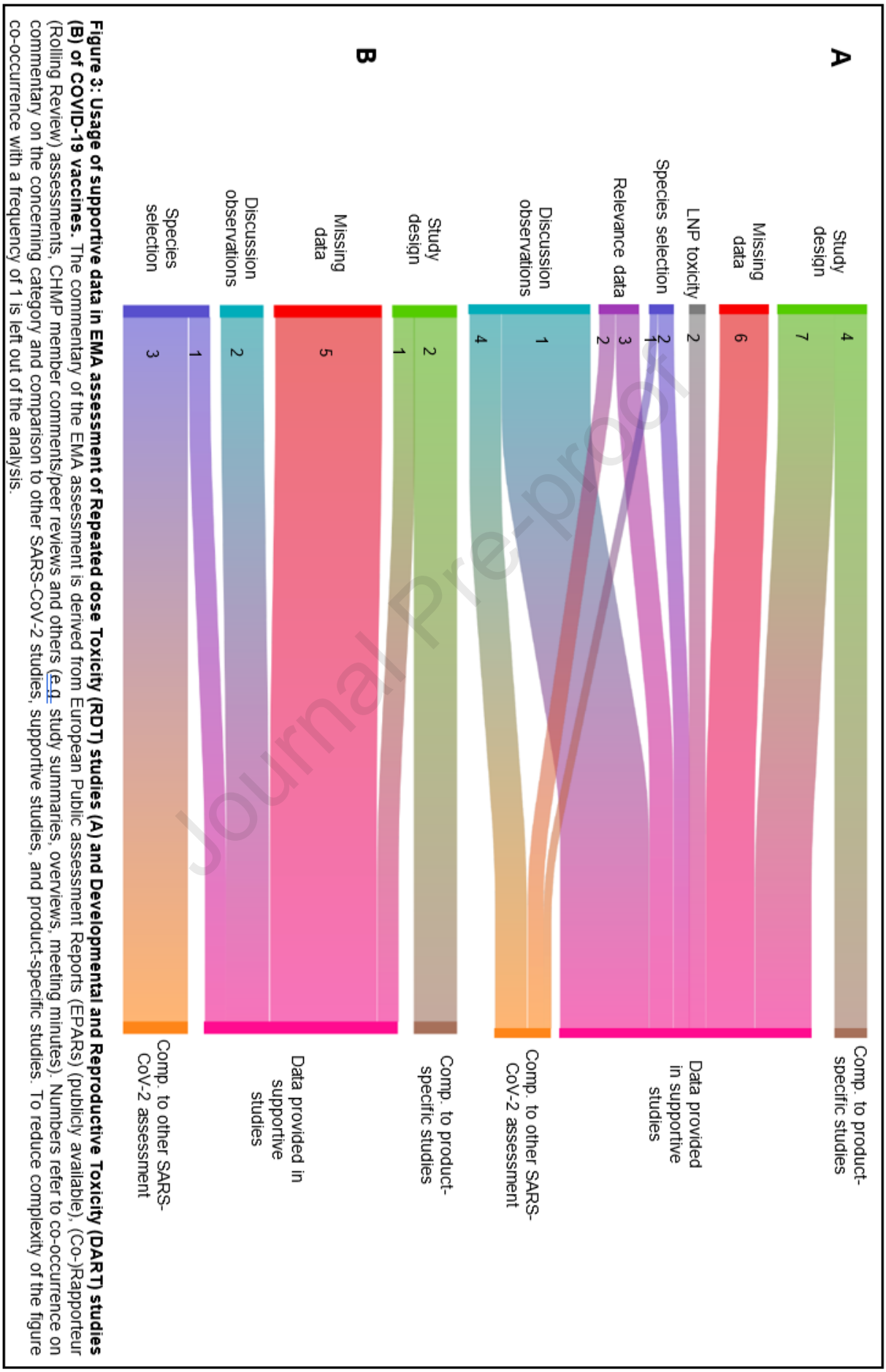


Figure 3: Usage of supportive data in EMA assessment of Repeated dose Toxicity (RDT) studies (A) and Developmental and Reproductive Toxicity (DART) studies (B) of COVID-19 vaccines. The commentary of the EMA assessment is derived from European Public assessment Reports (EPARs) (publicly available), (Co-)Rapporteur (Rolling Review) assessments, CHMP member comments/peer reviews and others (e.g. study summaries, overviews, meeting minutes). Numbers refer to co-occurrence on commentary on the concerning category and comparison to other SARS-CoV-2 studies, supportive studies, and product-specific studies. To reduce complexity of the figure co-occurrence with a frequency of 1 is left out of the analysis.

RDT studies, comparison to other SARS-CoV-2 assessment was most profoundly applied in the justification of the species selection. Comparison of the studies with product-specific studies (e.g. RDT or genotoxicity studies) was mostly carried out to resolve commentary concerning the provided study design and their shortcomings, e.g. incompliance with guidelines.

4. Discussion and Conclusions

By evaluating the non-clinical SARS-CoV-2 vaccine toxicity studies, we have found that there were no clear adverse effects. Effects are, not unexpectedly, local reactogenicity, immune response and macroscopic findings at the injection site. The majority of the observed findings are immunostimulatory effects, which, together with their reversible nature, can be considered non-adverse. Inflammatory reactions at the injection site are considered as intended reactions representing the immune response to the antigen[18]. This has been described earlier in a workshop report[19].

The outcome of the examination of 30 RDT studies of potential new vaccines across multiple species [20], was in the same line, showing no unexpected adverse findings. Evidence of an inflammatory response at the dose site was often accompanied by changes in draining lymph nodes, spleen and clinical pathology. Baldrick's work[11] recently confirmed these findings, both for general toxicity testing and developmental and reproductive toxicity (DART) for 5/6 COVID-19 vaccines, which are also included in our analysis. Even though vaccine dosing for DART testing took place on different days during (early) gestation, these differences did not affect the overall safety assessment. Taking into account the ICH S5 (R3)[6] Guideline on Reproductive toxicity and WHO Guidelines on non-clinical evaluation of vaccines and vaccine adjuvants and adjuvanted vaccines (2005, 2013)[4,5], it is important to keep in mind that in the design and evaluation of vaccination during early organogenesis a distinction should be made between the potential effects of the inflammatory response induced by vaccine components, and the exposure of the offspring to the vaccine-induced antibodies

It is clear that small scale non-clinical vaccine toxicity studies do not have sufficient power to predict the rare potentially clinically relevant adverse effects[21]. Theoretically this is a lasting obstacle in the general development of preventive vaccines, which applies also to the SARS-CoV-2 vaccine RDT and DART toxicity studies.

With regard to the reported effects on AST and ALT it is important to have a specific discussion. These effects also received attention in the comments of the EMA evaluation. Should this be taken as a specific toxicity phenomenon, or might it be related to the primary induction of an innate immune response? We have seen that minimal changes in AST and ALT have been reported in relation to various vaccine technologies, i.e. with mRNA vaccines, but also with an adenovector vaccine and an adjuvanted vaccine. Donahue et al[22] have described effects of mRNA vaccines on AST and ALT. In our dataset a relation could be found with the occurrence of the acute phase protein response, indicating a involvement of the liver in the innate immune response. This suggests that these effects are more related to the primary pharmacology of vaccines, and should not be qualified as toxic.

Having appreciated the previously emphasized importance of a relevant animal species for non-clinical vaccine safety testing, Principal Component Analysis (PCA) indicated that effects in both the RDT and DART toxicity studies are mostly determined by the choice of species, i.e. species-specific, and suggest that the choice of species impacts the outcome of the study more than the utilized vaccine technology. This confirms the knowledge on species specificity retrievable from adequate historical control data[23]. Although the PCA analysis provided a useful way on visualising the inter-individual variation of findings between the COVID-19 vaccines, the analysis on the contributing categories to variation indicated that the major contributors to the variation in findings observed in both RDT and DART studies demonstrate a low relevance of the differences between these studies. These contributors were not clearly adverse, as suspected co-occurring dam-pup body weight gain decrease, due to disruption of lactation, was not observed and other contributors were not found to be vaccine specific, due to their marginal and immune stimulatory nature.

This also indicates that a less optimal choice in species selection is unlikely to contribute to overlooking relevant adverse effects. Still, due the limited numbers of studies as well as qualitative nature of the scoring the interpretation of the PCA analysis has limitations. Future work should focus on validating the results obtained in this study in a quantitative way and giving a more accurate representation of variation between the different studies.

The combination of negligibly low frequency of findings with absence of adverse effects, together with a low inter-individual variability, show that product-specific DART studies provide little added value to the existing platform data existing on DART findings of vaccines. DART studies have often been delivered after the start of clinical studies for COVID vaccines[24], and obstetrician can rely only on DART studies in animals as data on pregnancy outcome in women vaccinated in clinical trials is limited[25], supportive DART studies using products with similar vaccine technology have a large potential in making the gap in the lack of evidence on pregnant women smaller. Surprisingly, none of the applicants have actively included supportive platform data on DART studies. Recently, research has shown that experience using mRNA and adenoviral vector based vaccine platform data have been useful in providing evidence of absence of specific reproductive safety concerns[26]. Additionally, most of the adverse effects of vaccines have been argued to be caused by the encapsulation/vehicle and not by vaccine-related immune response, i.e. the biological activity of the expressed antigens, as has been extensively discussed for LNP encapsulations[27]. This accumulative knowledge encourages the use of supportive platform data for vaccine safety assessment. A recent workshop of the Coalition of Epidemic Preparedness (CEPI)[28] supports the statement that new vaccine development might be accelerated by usage of identical platform technology for determination of non-clinical safety , making a case that additional (product-specific) DART studies might not be required, however still warranting a case-specific assessment.

Of note, in the current study we did not observe an increased incidence of birth defects, embryo fetal lethality or growth abnormalities, which is a major public concern of vaccination and drug treatment in general during pregnancy. A number of systematic reviews on possible adverse effects of COVID-19 mRNA vaccines on pregnancy in human have been published[29–31] indicating that COVID-19 (mRNA) vaccination in pregnancy appeared to be safe. The lack of adverse findings of mRNA vaccines on pregnancy in these systematic reviews corresponds with the lack of adverse findings in the EFD studies in animals taken into account in our analysis.

Innovatively, we found that assessment of the studies by the EMA consisted most frequently of comments related to study design, species selection and missing information regardless of the utilized vaccine technology and supportive platform studies often substantiated the comments on these main three categories. This was the case for both RDT and DART studies. It was observed that data from supportive studies were used to fill in knowledge gaps of product-specific studies. Usually the product-specific studies were less elaborative in comparison to the supportive studies. In particular, one of the applicants of a mRNA-based vaccine (mRNA2) seemed to include a product-specific RDT study with poor study design as part of the strategy to rely mostly on multiple provided supportive studies with highly similar mRNA constructs. Noteworthy, supportive information, such as available clinical data, were used to highlight shortcomings in the design or to argue the necessity of additonal measurements. In some cases, this even led to product-specific studies being regarded as unnecessary by the assessors, because of the availability of clinical studies.

Collectively, our study demonstrated the limited added and translational value of product-specific non-clinical studies for SARS-CoV-2 vaccines, due to their low frequency of observations outside of expected pharmacological inflammatory responses and their species-specificity. Thus, from a 3R perspective, both for RDT and for DART studies applicants are encouraged to use supportive platform technology data. However, the absence of clear adversity in the main contributors to the findings of the studies of the various vaccines indicate that suboptimal choices in species selection does not lead

to overlooking relevant adverse effects. This statement answers also the question raised in the Introduction, whether these nonclinical data were rather needed to decide about the safety of the vaccines. In fact, sponsors could have saved time and animals without compromising human safety when relying on supportive platform technology data.

Therefore, product-specific safety studies confined to minimal requirements, and with a justified package of supportive studies, these minimal studies are considered sufficient to support clinical development. This can form an intermediate step in the shift towards fully animal-free methods in non-clinical toxicity testing. Regulatory improvement on supportive platform technology data and further development of new animal-free approach methods (NAM) are promising tools in this shift.

5.1 Author contributions

NS analysed all raw data on studies submitted by the MAHs to the EMA and EMA assessment documents compiled by the EMA on all available COVID-19 vaccines and wrote the manuscript. JWvdL did the first conceptualization. BT, PT, KO and JWvdL contributed to the advice and discussion of content, reviewed the manuscript before submission. The Dutch Medicines Evaluation Board (MEB) provided funding.

5.2 Declaration of interests

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest with the subject matter discussed in the manuscript.

References

1. Krammer F. SARS-CoV-2 vaccines in development. *Nature*. 2020;586(7830):516-527. doi:10.1038/s41586-020-2798-3

2. Busquet F, Hartung T, Pallocca G, Rovida C, Leist M. Harnessing the power of novel animal-free test methods for the development of COVID-19 drugs and vaccines. *Arch Toxicol*. 2020;94(6):2263-2272. doi:10.1007/s00204-020-02787-2

3. Mykytyn AZ, Lamers MM, Okba NMA, et al. Susceptibility of rabbits to SARS-CoV-2. *Emerg Microbes Infect*. 2021;10(1):1. doi:10.1080/22221751.2020.1868951

4. World Health Organization. WHO guidelines on nonclinical evaluation of vaccines. *WHO Tech Rep Ser No 927, 2005 Annex*. 2005;(927).

5. World Health Organization. Guidelines on the nonclinical evaluation of vaccine adjuvants and adjuvanted vaccines. *WHO Press*. 2013;(October).

6. ICH S5. ICH S5 (R3) guideline on reproductive toxicology: Detection of Toxicity to Reproduction for Human Pharmaceuticals Step 5. *Eur Med Agency Comm Med Prod Hum Use*. 2020;5(EMA/CHMP/ICH/544278/1998).

7. Vandeputte J, Saville M, Cavaleri M, et al. IABS/CEPI platform technology webinar: Is it possible to reduce the vaccine development time? *Biologicals*. 2021;71(April):55-60. doi:10.1016/j.biologicals.2021.04.005

8. European Medicines Agency. Guideline on data requirements for vaccine platform technology master files. *Eur Med Agency*. Published online 2022.

9. Global regulatory workshop on COVID-19 vaccine development | International Coalition of Medicines Regulatory Authorities (ICMRA). Accessed March 18, 2022. https://www.icmra.info/drupal/en/news/March2020/summary

10. ICH M3 (R2) EMA. ICH M3 (R2) - Non-clinical Safety studies for the conduct of human clinical trials and marketing authorisation for pharmaceuticals. *ICH Guidel*. 2009;3(R2):31.

11. Baldrick P. Development of COVID-19 therapies: Nonclinical testing considerations. *Regul Toxicol Pharmacol*. 2022;132. doi:10.1016/J.YRTPH.2022.105189

12. Committee for Human Medicinal Products. Guideline on Repeated Dose Toxicity. *Eur Med Agency*. Published online 2010.

13. U.S. Department of Health and Human Services Food and Drug Administration. Considerations for Developmental Therapeutic Vaccines for Infectious Toxicity Studies for Preventive and Disease Indications, Guidance for Industry. 2006;(February). doi:10.1186/1477-7525-4-79

14. Wilson S. The MICE Algorithm. Published 2021. Accessed March 19, 2022. https://cran.r-project.org/web/packages/miceRanger/vignettes/miceAlgorithm.html

15. ATLAS.ti GmbH. ATLAS.ti: The Qualitative Data Analysis; Research Software. Web. Published 2022. Accessed May 22, 2022. https://atlasti.com/

16. How to read PCA biplots and scree plots - BioTuring's Blog. Accessed June 9, 2022. https://blog.bioturing.com/2018/06/18/how-to-read-pca-biplots-and-scree-plots/

17. Verbeke R, Lentacker I, De Smedt SC, Dewitte H. COVID-19 vaccines. *Drugs Lact Database*. 2022;28:100766. Accessed June 9, 2022. https://www.ncbi.nlm.nih.gov/books/NBK565969/

18. Ramot Y, Kronfeld N, Ophir Y, et al. Toxicity and Local Tolerance of a Novel Spike Protein RBD Vaccine Against SARS-CoV-2, Produced Using the C1 Thermothelomyces Heterothallica Protein Expression Platform. *Toxicol Pathol*. 2022;50(3):294-307. doi:10.1177/01926233221090518

19. Van Der Laan JW, Forster R, Ledwith B, et al. Nonclinical Testing of Vaccines: Report from a Workshop. *Ther Innov Regul Sci*. 2009;43(1):97-107. doi:10.1177/009286150904300115

20. Baldrick P. Dose site reactions and related findings after vaccine administration in safety studies. *J Appl Toxicol*. 2016;36(8):980-990. doi:10.1002/jat.3314

21. Berlin JA, Glasser SC, Ellenberg SS. Adverse Event Detection in Drug Development: Recommendations and Obligations Beyond Phase 3. *Am J Public Health*. 2008;98(8):1366. doi:10.2105/AJPH.2007.124537

22. Donahue DA, Ballesteros C, Maruggi G, et al. Nonclinical Safety Assessment of Lipid Nanoparticle-and Emulsion-Based Self-Amplifying mRNA Vaccines in Rats. *Int J Toxicol*. 2023;42(1):37-49. doi:10.1177/10915818221138781

23. Namdari R, Jones K, Chuang SS, et al. Species selection for nonclinical safety assessment of drug candidates: Examples of current industry practice. *Regul Toxicol Pharmacol*. 2021;126:105029. doi:10.1016/J.YRTPH.2021.105029

24. COVID-19 vaccine tracker. Accessed March 19, 2022. https://vac-lshtm.shinyapps.io/ncov_vaccine_landscape/

25. Rasmussen SA, Kelley CF, Horton JP, Jamieson DJ. Coronavirus Disease 2019 (COVID-19) Vaccines and Pregnancy: What Obstetricians Need to Know. *Obstet Gynecol*. 2021;137(3):408. doi:10.1097/AOG.0000000000004290

26. Shook LL, Fallah PN, Silberman JN, Edlow AG. COVID-19 Vaccination in Pregnancy and Lactation: Current Research and Gaps in Understanding. *Front Cell Infect Microbiol*. 2021;11:899. doi:10.3389/FCIMB.2021.735394/BIBTEX

27. Sedic M, Senn JJ, Lynn A, et al. Safety Evaluation of Lipid Nanoparticle-Formulated Modified mRNA in the Sprague-Dawley Rat and Cynomolgus Monkey. *Vet Pathol*. 2018;55(2):341-354. doi:10.1177/0300985817738095

28. Voss G, Jacquet JM, Tornieporth N, Kampmann B, Karron R, Meulen AS, Chen R, Gruber M, Lurie N, Weller C, Cramer JP, Saville M, Darko M. Meeting report: CEPI consultation on accelerating access to novel vaccines against emerging infectious diseases for pregnant and lactating women, London, 12-13 February 2020. Vaccine. 2021 Dec 17;39(51):7357-7362. doi: 10.1016/j.vaccine.2021.10.048.

29. Joubert E, Kekeh AC, Amin CN. COVID-19 and novel mRNA vaccines in pregnancy: an updated literature review. *BJOG An Int J Obstet Gynaecol*. 2022;129(1):21-28. doi:10.1111/1471-0528.16973

30. Kalafat E, Heath P, Prasad S, O`Brien P, Khalil A. COVID-19 vaccination in pregnancy. *Am J Obstet Gynecol*. Published online May 2022. doi:10.1016/J.AJOG.2022.05.020

31. Pratama NR, Wafa IA, Budi DS, Putra M, Wardhana MP, Wungu CDK. mRNA Covid-19 vaccines in pregnancy: A systematic review. *PLoS One*. 2022;17(2 February). doi:10.1371/JOURNAL.PONE.0261350

Highlights

- In vivo animal toxicity testing shows limited value in establishing safety of vaccines.
- Vaccine-induced effects are strongly species-specific.
- Platform studies support EMA comments on product-specific vaccine assessment.
- Confined non-clinical studies provided with supportive studies support clinical development.

Funding body information
This study has been funded by the Medicines Evaluation Board (MEB) (Utrecht, the Netherlands), the MEB being part of the European Network of Regulatory Authorities on Human Pharmaceuticals.

More information can be found on https://english.cbg-meb.nl/.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: