Regulatory Toxicology and Pharmacology 138 (2023) 105337

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

An evaluation of reproductive toxicity studies and data interpretation of N-methylpyrrolidone for risk assessment: An expert panel review

C.R. Kirman [a,*], B.R. Sonawane [b], J.G. Seed [c], N.O. Azu [d], W.T. Barranco [e], W.R. Hamilton [f], T.J. Stedeford [g], S.M. Hays [a]

[a] SciPinion, Bozeman, MT, 59715, USA
[b] Independent Consultant, Gaithersburg, MD, 20878, USA
[c] Independent Consultant, Alexandria, VA, 22301, USA
[d] BASF Corporation, Houston, TX, 77079, USA
[e] Lyondell Chemical Company, Houston, TX, 77010, USA
[f] Ashland, Wilmington, DE, 19808, USA
[g] Bergeson & Campbell, P.C., Washington, DC, 20037, USA

ARTICLE INFO

Handling Editor: Dr. Martin Van den berg

Keywords:
Expert panel
Data quality scoring
Risk assessment
Reproductive toxicity

ABSTRACT

An expert panel was assembled to evaluate reproductive toxicology study data and their application to health risk assessment to provide input on the data quality, interpretation, and application of data from three multi-generation reproductive toxicity studies of N-methylpyrrolidone (NMP). Panelists were engaged using a double-blinded, modified Delphi format that consisted of three rounds. Key studies were scored using the U.S. Environmental Protection Agency's (EPA) questions and general considerations to guide the evaluation of experimental animal studies for systematic review. The primary conclusions of the panel are that one of the studies (Exxon, 1991) is not a high-quality study due to several design flaws that includes: (1) exceedance of the maximum tolerable dose in the high dose group; (2) failure to adjust feed concentrations of NMP during the lactation period, resulting in NMP doses that were 2- to 3-fold higher than nominal levels; and/or (3) underlying reproductive performance problems in the strain of rats used. For these reasons, the panel recommended that this study should not be considered for quantitative risk assessment of NMP. Exclusion of this study, and its corresponding data for male fertility and female fecundity, from the quantitative risk assessment results in a change in the identification of the most sensitive endpoint. Instead, changes in rat fetal/pup body weight, an endpoint previously selected by EPA, was identified as an appropriate basis for human health risk assessment based on a consideration of the best available science and weight of scientific evidence supported by the NMP toxicity database.

1. Introduction

Under the amended Toxic Substances Control Act (TSCA) Section 6, the U S EPA (EPA) requires screening and prioritizing of existing chemical substances as either high- or low-priority substances, and to perform risk evaluations on high-priority substances (EPA, 2016). If EPA determines that a chemical substance presents an unreasonable risk to human health or the environment under its conditions of use, EPA must initiate risk management rulemakings.

Prior to the TSCA amendments, EPA's authority to regulate existing chemical substances under TSCA Section 6 focused its efforts on existing chemical substances primarily through voluntary programs. One of these programs included the TSCA Work Plan Chemicals initiative, which began in March 2012. At that time, EPA developed a work plan of existing chemical substances for further assessment under TSCA, based on specific hazard and exposure criteria. EPA included N-methylpyrrolidone (NMP), a low volatility organic solvent, in the work plan. EPA based its decision to include NMP on hazard concerns identified for reproductive toxicity, exposure concerns from its use in consumer products (*e.g.*, pigments and cosmetics), and high reported releases to the environment.

In March 2015, EPA issued the final "*TSCA Work Plan Chemical Risk Assessment on N-Methylpyrrolidone: Paint Stripper Use*" (hereinafter the "NMP Work Plan Assessment"; USEPA, 2015). EPA identified fetal body

* Corresponding author.
E-mail address: ckirman@scipinion.com (C.R. Kirman).

https://doi.org/10.1016/j.yrtph.2023.105337
Received 8 August 2022; Received in revised form 28 December 2022; Accepted 10 January 2023
Available online 14 January 2023

weight changes in rats exposed to NMP on gestation days 6–20 (Saillenfait et al., 2003) as the critical endpoint for the NMP Work Plan Assessment. Physiologically based pharmacokinetic (PBPK) modeling used to quantify species differences in the toxicokinetics of NMP permitted an assessment for NMP on an internal dose basis (*e.g.*, using daily area under the curve [AUC] for NMP in maternal blood). Using benchmark dose (BMD) modeling, EPA derived a point of departure (POD) value of 411 mg*hr/L (internal dose basis), which corresponded to a human equivalent dose (HED) of 48 mg/kg-bw/day as estimated by EPA using PBPK modeling. This POD was used to assess human exposure scenarios for NMP paint stripper scenarios along with a benchmark margin of exposure (MOE) of 30 (*i.e.*, human exposures need to be at least a factor of 30 lower than the POD for EPA to make a no unreasonable risk to human health finding). EPA identified unreasonable risks to workers and consumers under specific acute and chronic exposure use scenarios.

On January 19, 2017, EPA published a proposed risk management rule to address the unreasonable risks it identified in the NMP Work Plan Assessment, based on its use in paints and coatings removal. However, one month earlier (*i.e.*, December 19, 2016), EPA identified NMP as one of the first ten existing chemical substances to undergo risk evaluation pursuant to the new requirements under Section 6(b)(2)(A) of the amended TSCA. Thereafter, EPA expanded the scope of the NMP Work Plan Assessment for the risk evaluation on NMP to include additional conditions of use and to meet its requirements for best available science and weight of scientific evidence under Section 26 of the amended TSCA.

In December 2020, EPA issued the final "Risk Evaluation for n-Methylpyrrolidone [*sic*]" (hereinafter the "NMP Risk Evaluation"). In contrast to the NMP Work Plan Assessment, EPA identified effects on fertility in male rats as the critical endpoint, based on the results of a multi-generation reproductive toxicity study (Exxon Biomedical Sciences, 1991), effects that were not observed in other multi-generation reproductive studies with the same study design (BASF Department of Toxicology, 1999; Huntingdon Life Sciences, 1999). PBPK modeling and BMD modeling were again used to support the NMP Risk Evaluation. EPA derived a POD value of 183 h*mg/L (internal dose basis), which corresponds to an HED of 28 mg/kg-bw/day. This POD, which is approximately a factor of two lower than the value EPA used in the NMP Work Plan Assessment, was used to assess human exposure scenarios, along with a benchmark MOE of 30 for quantifying risks. EPA identified 26 conditions of use that present an unreasonable risk to human health.

On January 15, 2021, EPA withdrew the 2017 proposed risk management rule on NMP, which was based on the NMP Work Plan Assessment. In doing so, EPA stated that it planned to initiate regulatory action on NMP to address the unreasonable risks to human health it identified in the NMP Risk Evaluation.

Under the amended TSCA, the U.S. Congress included Section 26, which requires EPA to use the best available science under Section 26(h) and weight of scientific evidence under Section 26(i) for decisions made under TSCA Sections 4, 5, and 6. The U.S. Congress did not define these terms in TSCA; however, through notice and comment rulemaking, EPA codified these terms in relevant part as follows (40 CFR 702.33):

> *"Best available science means science that is reliable and unbiased. Use of best available science involves the use of supporting studies conducted in accordance with sound and objective science practices, including, when available, peer reviewed science and supporting studies and data collected by accepted methods or best available methods (if the reliability of the method and the nature of the decision justifies use of the data).*
>
> *Weight of scientific evidence means a systematic review method, applied in a manner suited to the nature of the evidence or decision, which uses a pre-established protocol to comprehensively, objectively, transparently, and consistently, identify and evaluate each stream of evidence, including strengths, limitations, and relevance of each study and to integrate evidence as necessary and appropriate based upon strengths, limitations, and relevance (40 CFR 702.33)."*

EPA stated in the NMP Risk Evaluation that "[t]o meet these TSCA Section 26 science standards, EPA used the TSCA systematic review process described in the [2018] *Application of Systematic Review in TSCA Risk Evaluations* document [citation omitted; hereinafter the "2018 SR Guidance Document"]." However, in February 2021, the U.S. National Academies of Sciences, Engineering, and Medicine (NASEM) released its final report on EPA's 2018 SR Guidance Document (NASEM, 2021), finding that "the process outlined in the 2018 [SR] [G]uidance [D]ocument, and as elaborated and applied in the example evaluations [*i.e.*, TCE and 1-bromopropane], does not meet the criteria of "comprehensive, workable, objective, and transparent." The NASEM found that "the systematic reviews within the draft risk evaluations considered did not meet the standards of systematic review methodology."

The NASEM review of the 2018 SR Guidance Document, along with EPA's previous evaluations on the available reproductive toxicity data, including developmental toxicity studies on NMP, raised questions about quality and reliability of the Exxon Biomedical Sciences (1991) study and its suitability for use in the NMP Risk Evaluation. In 2007, EPA served as the sponsoring agency for evaluating NMP through the Organization for Economic Cooperation and Development's (OECD)'s Screening Information Data Set (SIDS) Initial Assessment Report. In that document, EPA assigned Klimisch reliability scores of 2 for the Exxon Biomedical Sciences (1991) study and 1 for the Saillenfait et al. (2003) study. In 2015, EPA discounted the effects on reduced fertility because of the lack of consistency for this finding between studies and used developmental toxicity endpoints for evaluating potential unreasonable risks in the NMP Work Plan Assessment. However, in 2020, EPA re-evaluated the quality and reliability of the Exxon Biomedical Sciences (1991) study and assigned it a data quality score of "High", according to the 2018 SR Guidance Document. Although study scorers noted that NMP doses administered during the lactation period were 2–3 times higher than target levels (resulting in a medium rating of "2" for the metric of reporting of doses/concentrations) (USEPA, 2020d), no comments or score modifications were noted regarding reproductive performance issues in control animals or maximum tolerated dose exceedance in this study.

The available data sets on NMP all pre-date the evaluations performed by EPA for the OECD SIDS Assessment, the NMP Work Plan Assessment, and the NMP Risk Evaluation. However, the interpretation of these studies has changed, depending on the criteria used for assessing data quality and reliability. Therefore, the purpose of the present investigation was to evaluate the available multi-generation reproductive and developmental toxicity studies on NMP using a systematic review approach for assessing data quality and reliability that meets the TSCA Section 26 scientific standards. The EPA's Integrated Risk Information System (IRIS) draft report titled "ORD Handbook for Developing IRIS Assessments" (hereinafter the "IRIS Handbook") is selected for this purpose, based on NASEM's recommendation for EPA to adopt the IRIS approach or a comparable system in its TSCA risk evaluations. To accomplish this work, an expert panel was engaged to evaluate the key underlying studies and to assess the data quality and reliability of these studies using the IRIS Handbook. Input from this panel is used to guide decisions regarding endpoint and data set selection to be used in EPA's Benchmark Dose Software (BMDS) to derive POD values for NMP that reflect the best available science. The goal of this work was to determine whether the hazard POD used in the NMP Risk Evaluation meets the scientific standards under TSCA Section 26.

2. Background

NMP is a toxicity data-rich chemical, for which three multi-generation, reproductive toxicity studies were conducted (Exxon Biomedical Sciences, 1991; Huntingdon Life Sciences, 1999; BASF

Department of Toxicology, 1999). Exxon Biomedical Sciences (1991) was conducted prior to the availability of test guidelines for reproduction and fertility effects (USEPA OPPTS 870.3800; OECD 416). The two later studies (Huntingdon Life Sciences, 1999; BASF Department of Toxicology, 1999) incorporated these guidelines into their study designs, and were specifically implemented to evaluate the reproducibility of the study results from Exxon Biomedical Sciences (1991). NMP was administered to rats by the dietary route in all three studies. In two studies Sprague Dawley rats were used (Exxon Biomedical Sciences, 1991; Huntingdon Life Sciences, 1999), and in one study Wistar rats were used (BASF Department of Toxicology, 1999). All three studies relied on the same basic study design, including the same number of test groups (*i.e.*, three plus one control group), the same number of test animals per group (25–30 per sex per group), and multigenerational exposure design. The P (F0) generation was treated for approximately ten weeks followed by mating to produce the F1a litters. Following weaning, the F0 animals were mated again to produce the F1b litters. The F1 generation that was selected to continue treatment was taken from the F1b litters. Following an additional ten weeks of treatment, the mating protocol was again followed to produce the F2a and F2b litters. The data from these matings were used to calculate two reproductive indices: (1) a male fertility index, calculated as the number of males impregnating females divided by the number of males used for mating; and (2) a female fecundity index, calculated as number of females pregnant divided by the number of females confirmed mated. Both indices were derived for each of the mating trials (F1a, F1b, F2a, F2b). In addition, these studies include measurements for organ weights and body weights in parental animals, pup body weights for all mating trials, and other metrics for reproductive status (*e.g.*, fetal/pup survival).

2.1. Key study design differences

Although the three multigeneration, reproductive toxicity studies share similar study designs and the two more recent studies (*i.e.*, Huntingdon Life Sciences, 1999 and BASF Department of Toxicology, 1999) were conducted according to validated test guidelines issued by EPA and OECD and under good laboratory practice standards, there are at least three important differences, as summarized in Table 1 and discussed briefly below.

Table 1
Summary of key design differences for the multigenerational reproductive toxicity studies in rats for NMP.

Design Parameter	Study		
	Exxon Biomedical Sciences (1991)	Huntingdon Life Sciences (1999)	BASF Department of Toxicology (1999)
1. NMP Dose groups	All litters: 0, 50, 160, 500 mg/kg-bw/day	F1a: 0, 50, 160, 500 mg/kg-bw/day All other litters: high dose reduced to 350 mg/kg-bw/day for all other litters due to MTD exceedance	F1a: 0, 50, 160, 500 mg/kg-bw/day All other litters: high dose reduced to 350 mg/kg-bw/day for all other litters due to MTD exceedance
2. Rat Strain: Time-to-Mating in Control Animals (range for all litters)	Sprague-Dawley 1st week: 28–47% 2 nd week: 36–52% 3rd week: 4–10% 4th week: 0–4%	Sprague-Dawley (International Genetic Standardization Program) Days 1–4: 89–100%	Wistar Days 1–4: 92–100%
3. Dose range of NMP administered during lactation period	48.2–151.0 mg/kg-bw/day 146.4–465.8 mg/kg-bw/day 335.8–1078.1 mg/kg-bw/day	30.2–70.3 mg/kg-bw/day 103.4–233.6 mg/kg-bw/day 217.6–573.3 mg/kg-bw/day	48.3–51.9 mg/kg-bw/day 157–173.3 mg/kg-bw/day 275.7–337.5 mg/kg-bw/day

2.1.1. NMP dose groups

The dose levels used initially in all three multigeneration studies in rats (*i.e.*, dietary NMP exposure) were 0, 50, 160, and 500 mg/kg-bw/day. In Exxon Biomedical Sciences (1991), these nominal dose levels were used throughout the entire study, despite significant postnatal toxicity and loss of pups in the 500 mg/kg-bw/day group. In Huntingdon Life Sciences (1999) and BASF Department of Toxicology (1999), an increased postnatal toxicity and loss of pups in the F1a litters at the 500 mg/kg-bw/day dose level replicated the results of Exxon Biomedical Sciences (1991) for this endpoint. However, in both follow-up studies it was concluded that the 500 mg/kg-bw/day dose exceeded the maximum tolerable dose (MTD, defined as the highest not producing overt toxicity) for NMP. Toxicity testing above the MTD has been criticized since it may initiate processes and events leading to effects that may not be expected at lower doses. For this reason, the highest treatment dose was adjusted from 500 to 350 mg/kg-bw/day for all subsequent matings (*i.e.*, F1b, F2a, and F2b).

2.1.2. Species and strain of rat used

Although BASF Department of Toxicology (1999) and Huntingdon Life Sciences (1999) studies were designed to determine the reproducibility of Exxon Biomedical Sciences (1991), different rat strains were used. Huntingdon Life Sciences (1999) used a newer strain of Sprague-Dawley rats from the International Genetic Standardization (IGS) Program, developed in 1992 to minimize inbreeding and genetic drift (Charles River, 2021). BASF Department of Toxicology (1999) used Wistar rats, a strain commonly used for toxicity testing in Europe. The reproductive performance of these rat strains differed from the unexposed control animals in Exxon Biomedical Sciences (1991) (Table 1). A substantial number of control animals in the strain used in Exxon Biomedical Sciences (1991) exhibited difficulty mating (>50% required more than one week), compared to the other two studies in which matings were complete within four days. Considering the mean length of estrus cycle in Sprague Dawley rats is between four and five days (Exxon Biomedical Sciences, 1991), a breeding rate of less than 50% within the first week of mating is indicative of an underlying reproductive problem.

2.1.3. NMP doses during lactation period

During lactation, feed intake normally increases to match the increased caloric demand of the rat dams with nursing pups. In BASF Department of Toxicology (1999) and Huntingdon Life Sciences (1999) studies, feed concentrations of NMP were adjusted during lactation to offset this increase in feed intake, so that NMP doses remain near nominal dose values (Table 1). As a result, maximum achieved dose levels during lactation in these two studies were as much as 1.6- and 1.2-fold higher than nominal values, respectively. In contrast, achieved doses levels in Exxon Biomedical Sciences (1991) were up to 3.1-fold higher than nominal dose levels due to the lack of any adjustments to NMP feed concentrations. The lack of adjusted NMP concentrations in Exxon Biomedical Sciences (1991) compounds the problem of exceeding the MTD in the highest test group.

2.2. Key study result differences

The results of the reproductive toxicity testing for NMP showed notable differences across the three studies. Some toxicity findings were unique to Exxon Biomedical Sciences (1991), particularly those related to the effects of NMP on fertility in male rats and fecundity in female rats (Table 2; See **Appendix A** for detailed data summaries).

2.2.1. Male fertility/female fecundity

Male fertility and female fecundity data in F2a and F2b litters in Exxon Biomedical Sciences (1991) indicate significant dose-dependent decreases in both endpoints (**Appendix A**), with greater than 50%

Table 2
Summary of key reproductive toxicity results from the multigenerational studies in rats conducted for NMP.

Endpoint	Study		
	Exxon Biomedical Sciences (1991)	Huntingdon Life Sciences (1999)	BASF Department of Toxicology (1999)
Male Fertility	No significant changes in F1a and F1b litters; Dose-dependent decreases in F2a and F2b litters (up to 59% decrease)	No significant change	No significant change
Female Fecundity	No significant changes in F1a and F1b litters; Dose-dependent decreases in F2a and F2b litters (up to 51% decrease)	No significant change	No significant change
Testes Weight	Not measured	Small, non-monotonic increases in absolute and relative testes weights (up to 6.1% increase in relative weight)	Small, non-monotonic increases in absolute and relative testes weights (up to 10.6% increase in relative weight)
Pup Body Weights	Dose-dependent decreases (up to 22% decrease)	Dose-dependent decreases (up to 20% decrease in F1a litter; up to 10% decrease in all other litters)	Dose-dependent decreases (up to 18% decrease in F1a litter; up to 7% decrease in all other litters)

decreases reported at the highest test dose. These effects were not observed in F1a and F1b litters from this study. Similarly, these effects were not noted in any litters at any dose level in the studies of Huntingdon Life Sciences (1999) and BASF Department of Toxicology (1999).

2.2.2. Testes weights

Testes weights were not measured in Exxon Biomedical Sciences (1991) but were included in the study designs for BASF Department of Toxicology (1999) and Huntingdon Life Sciences (1999). In both studies, small increases in absolute testes weights (~5% or less) were noted in treatment groups, and in combination with small decreases in body weights (again, ~5% or less). Study authors for both studies did not identify these organ weight changes as a substance-related adverse effect. As a result, treatment groups showed increases in relative testes weight (up to 6.1% in Huntingdon Life Sciences, 1999; up to 10.6% in BASF Department of Toxicology, 1999) (details in **Appendix A**). However, these changes in weight were generally within the range for historical controls, and were not accompanied by gross pathology findings, histopathology findings, or changes in sperm parameters. Furthermore, dose-response trends for this endpoint were non-monotonic (*i.e.*, response was larger in the mid-dose group than in the high-dose group), which complicates their potential application to risk assessment (*e.g.*, poor fits in benchmark dose modeling).

2.2.3. Pup body weight

Treatment-related effects of NMP on pup body weights have consistently been reported in all three studies (Exxon Biomedical Sciences, 1991; Huntingdon Life Sciences, 1999; BASF Department of Toxicology, 1999). Pup body weights for animals on PND 1 were not significantly different from controls for the 50 and 160 mg/kg-bw/day treatment groups. For animals in the 500 mg/kg-bw/day dose groups, pup body weights on PND 1 were significantly lower (up to 18–22% decrease). Pup body weights on PND 1 were also reduced in the 350 mg/kg-bw/day treatment groups, but with less severity (up to 10% decrease) (Huntingdon Life Sciences, 1999; BASF Department of Toxicology, 1999).

3. Methods

A white paper was prepared by the review sponsor on the key data sets and issues for the reproductive effects of NMP (**Appendix A**). The purpose of the white paper was to provide a summary of the key data sets and a road map (*i.e.*, page and table numbers) to where specific data are located, since the individual study reports were 1,800–2,200 pages each.

3.1. Expert panel on reproductive toxicity

SciPinion assembled an expert panel to address key questions and issues associated with the interpretation of multigenerational reproductive toxicity studies using the methods summarized in Kirman et al. (2019). The goal of this review process is to engage independent scientists to support decision-making, while minimizing sources of bias and conflicts of interest, and improving transparency. Because this approach is robust and compares favorably to standard practices for regulatory agency peer reviews, it has recently been used to support a proposed corrective action by the USEPA's Office of Pesticides (EPA, 2022b).

To support a review of the quality of the three multi-generational reproductive toxicity studies for NMP, an expert panel of three reproductive/regulatory toxicologists was identified using the following steps: (1) Panel Recruitment; (2) Panel Expertise Verification Audit; (3) Panel Selection; (4) and Panel Engagement. Each of these steps is summarized below.

3.1.1. Panel recruitment

The goal of the panel recruitment was to cast as wide a net to reach out to as many potential candidates as is feasible. Potential candidates were identified as having relevant experience in reproductive toxicology and risk assessment using a variety of sources, including: (1) SciPinion's internal database; (2) searches for authors of recent publications on the topic of interest in online databases (*e.g.*, PubMed; Google Scholar); (3) searches of profiles on social media databases (*e.g.*, LinkedIn); (4) general internet searches; and (5) referrals. Email addresses were obtained for potential candidates. An email invitation was sent to potential candidates, requesting interested candidates to volunteer on the website (https://app.scipinion.com/scipis) and upload a copy of their curriculum vitae (CV).

3.1.2. Panel Selection

Based on the recruitment effort, 91 candidates applied for this opportunity. Six applicants were excluded for failing to upload their CV. For the remaining 85 candidates, CVs were collected and reviewed. Three panel members were selected from the pool of available candidates based upon a consideration of objective expertise metrics (*e.g.*, number of publications, years of experience, key word counts) for reproductive toxicity testing and its application to human health risk assessment. One panelist had to drop out due to an unforeseen scheduling conflict, and an alternate expert was selected to fill the open spot. The general expertise of the selected panelists was characterized as follows:

- Advance degrees: PhD (2), MD (1)
- Combined years of post-degree experience: >130 years
- Combined publications: >350

Two panel members were formerly employed with USEPA, with strong experience in the interpretation of reproductive toxicology data and application in human health risk assessment of chemicals, and one panelist was formerly an editor-in-chief of a peer-reviewed journal

focused on the topic area (*i.e.*, *Reproductive Toxicology*).

3.1.3. Panel engagement

The peer review consisted of three rounds of participation, all conducted online *via* a web app (app.scipinion.com) using a blinded, modified Delphi format:

- *Round 1 (Review white paper and answer charge questions; December of 2021)* –The panelists worked independently to review the white paper, along with access to the individual study reports, and answer the charge questions. All three panel members participated in this round as scheduled.
- *Round 2 (Online comment and anonymous debate; January of 2022)* – Panelists participated in the online comment and debate anonymously (each panel member was assigned a display name, *e.g.*, "Expert 1"). All 3 panelists had access to the responses obtained in Round 1 and participated in the online debate on schedule. A total of 20 comments were submitted by the panelists during this round.
- *Round 3 (Revise answers to initial charge questions, answer additional charge questions; January of 2022)* – The panelists were permitted to revise their responses to Round 1 charge questions, post comments, and debate round (in case interaction with fellow panel members resulted in a change in their position) and were tasked with answering additional/follow-up charge questions.

3.1.4. Study quality scoring and charge questions

A series of charge questions was developed to probe the panelists opinions on study quality and interpretation. All charge questions and panelist's responses are included in Appendix B. Study quality was assessed using EPA's questions and general considerations to guide the evaluation of animal studies for systematic review (USEPA, 2020, Tables 6–9), which included 9 domains: 1) Reporting quality; 2) Allocation; 3) Observational bias/blinding; 4) Confounding; 5) Selective reporting and attrition; 6) Chemical administration and characterization; 7) Exposure timing, frequency, and duration; 8) Endpoint sensitivity and specificity; 9) Results presentation. For each domain, the panelists were tasked with rating the studies as either "good" (score =

Table 3
Dose-response data for fetal/pup (PND 1) body weight changes following oral exposures to NMP.

Reference	Exposure Regimen (Gestation period coverage)	Administered Dose (mg/kg-bw/day)	Average Daily AUC, GD13-20 (mg*hr/L)*	Sex	Fetal/Pup Weight							
					Litter 1		Litter 2		Litter 3		Litter 4	
					n	Mean±SD (g)	n	Mean±SD (g)	n	Mean±SD (g)	n	Mean±SD (g)
Saillenfait et al. (2002)	Gavage (GD6-20)	0	0	M	21	5.79±0.42	NA					
		125	1290		21	5.74±0.25						
		250	2820		24	5.32±0.45						
		500	6430		25	4.18±0.22						
		750	10500		8	3.03±0.40						
		0	0	F	21	5.62±0.50	NA					
		125	1290		21	5.47±0.20						
		250	2820		24	5.02±0.29						
		500	6430		25	3.88±0.28						
		750	10500		8	3.09±0.47						
Sitarek et al. (2012)	Gavage (5 d/week; pre-mating through lactation)	0	0	M	22	6.03±0.78	NA					
		150	1140		24	5.73±0.49						
		450	3790		20	5.13±0.39						
		1000	9890		15	--						
		0	0	F	22	5.55±0.51	NA					
		150	1140		24	5.47±0.48						
		450	3790		20	5.13±0.39						
		1000	9890		15	--						
Huntingdon/ Thornton (1999)	Feed (GD1-20)	0	0	M	25	6.9±0.64	24	6.7±0.78	29	7.3±0.66	25	7.3±0.66
		50	591		25	7.1±0.42	24	7.2±0.47	27	7±0.57	26	6.9±1.04
		160	1960		24	7±0.8	23	6.8±0.41	29	7±0.58	24	6.9±0.71
		500/350	7920/5010		22	5.5±0.84	27	6.7±0.64	27	6.8±0.69	23	6.6±0.93
		0	0	F	25	6.5±0.71	24	6.5±0.66	29	6.8±0.66	25	6.9±0.66
		50	591		25	6.7±0.4	24	6.8±0.46	27	6.7±0.59	26	6.5±1.01
		160	1960		24	6.6±0.75	23	6.5±0.47	28	6.6±0.57	23	6.5±0.76
		500/350	7920/5010		22	5.2±0.77	27	6.4±0.56	27	6.3±0.65	23	6.2±0.82
BASF (1999)	Feed (GD1-20)	0	0	M	25	6.6±0.41	24	6.4±0.59	24	6.4±0.57	25	6.5±0.59
		50	591		25	6.3±0.67	25	6.3±0.42	24	6.2±0.61	25	6.4±0.61
		160	1960		24	6.3±0.47	24	6.4±0.5	24	6.4±0.7	25	6.5±0.5
		500/350	7920/5010		16	5.5±0.95	25	6.1±0.59	21	6±0.64	23	6.2±0.55
		0	0	F	25	6.2±0.46	24	6.1±0.57	24	6.1±0.51	25	6.1±0.57
		50	591		25	6±0.55	25	5.9±0.44	24	5.8±0.52	25	6.1±0.59
		160	1960		24	5.9±0.5	24	6±0.58	24	6±0.64	25	6.1±0.48
		500/350	7920/5010		16	5.1±0.85	25	5.9±0.67	21	5.7±0.77	23	5.8±0.63
Exxon (1991)	Feed (GD1-20)	0	0	M	25	6.84±0.85	16	6.97±0.64	26	6.35±0.6	24	6.46±0.87
		50	591		25	6.49±0.75	16	6.7±0.78	22	6.35±0.94	19	6.57±0.71
		160	1960		22	6.55±0.75	17	6.51±0.84	20	6.34±0.74	18	6.37±0.67
		500	7920		21	5.44±0.8	16	5.67±0.78	13	5.31±0.52	7	5.05±0.55
		0	0	F	25	6.51±0.89	16	6.55±0.79	26	6.09±0.59	24	5.96±0.78
		50	591		25	6.12±0.73	16	6.17±0.72	22	6.05±0.76	19	6.21±0.62
		160	1960		22	6.23±0.71	17	6.23±0.71	20	5.88±0.57	18	6.05±0.66
		500	7920		21	5.14±0.78	16	5.24±0.69	13	5.04±0.63	7	4.82±0.46

3), "adequate" (score = 2), "deficient or not reported" (score = 1), or "critically deficient". In this way, the maximum score a study could receive is 27 (9 domains × maximum score of 3). In addition, the panelists were asked separately to rate their overall confidence in the study on a scale of 1–10 (1 = lowest; 10 = highest), which may also include considerations outside of the nine domains listed above. Additional charge questions were included to identify appropriate NOAEL and LOAEL values from each study, and guide decisions in the quantitative risk assessment (see Appendix B).

3.2. Benchmark dose modeling

Dose-response data for fetal body weight changes at the end of gestation or in pups on postnatal day (PND) 1 were assembled for oral (gavage or diet) exposures (Table 3) and for other routes of exposure to NMP (Table 4). Internal doses for NMP (area under the curve [AUC] for parent chemical in maternal blood during the window of susceptibility, GD 13–20) are provided in these tables, as determined in Poet et al. (2016) for these routes of exposure. In total, these tables summarize 34 data sets that describe the dose-response relationship for fetal/pup body weight changes, and include more than 3,000 litters, and 24 unique internal dose groups across more than two orders of magnitude of NMP exposure.

Benchmark dose (BMD) modeling was performed using EPA's BMD Software package (version 3.2), in a manner generally consistent with EPA guidance (USEPA, 2012). The best fitting model was selected based on a consideration of Akaike information criterion (AIC; lower value indicates a better fit), chi-square goodness of fit p-value (higher value indicates a better fit with $p > 0.1$ as EPA's recommended cutoff), ratio of the BMD:BMDL (lower value indicates less model uncertainty), and visual inspection. The internal dose corresponding to a decrease in the mean response equivalent to one standard deviation (SD) in untreated animals (BMC_{SD}) and its 95% lower confidence limit ($BMCL_{SD}$) and upper confidence limit ($BMCU_{SD}$) were calculated. Control SD values for pup/fetal body weight were on average 8.6% of the control mean values (Tables 3 and 4), which falls between values of 5% and 10% that are also sometimes used for the POD (i.e., $BMDL_{05}$ alternative POD values are provided in Appendix C). Based on this consideration, the $BMDL_{SD}$ is identified to be an appropriate POD for weight changes reported in developmental toxicity studies based on precedent assessments for other chemical (*e.g.,* noncancer assessments for tetrahydrofuran, cyclohexane, and phenol) (IRIS, 2022).

Table 4
Dose-response data for fetal/pup (PND 1) body weight changes following inhalation and dermal exposures to NMP.

Reference	Exposure Regimen (Gestation period coverage)	Administered Dose (mg/kg-bw/day) or Concentration (ppm)	Average Daily AUC, GD13–20 (mg*hr/L)*	Sex	Fetal/Pup Weight	
					Litter 1	
					n	Mean±SD (g)
Inhalation Studies						
Saillenfait et al. (2003)	Inhalation, 6 h/d (GD6-20)	0	0	M	24	5.81 ± 0.39
		30	158		20	5.74 ± 0.32
		60	322		19	5.64 ± 0.21
		120	668		25	5.52 ± 0.44
		0	0	F	24	5.54 ± 0.37
		30	158		20	5.42 ± 0.47
		60	322		19	5.32 ± 0.30
		120	668		25	5.21 ± 0.44
Solomon et al. (1995)	Inhalation, 6 h/d (GD1-20)	0	0	MF	39	7.48 ± 0.71
		10	52		16	7.03 ± 0.70
		51	272		15	7.14 ± 0.69
		116	643		22	6.66 ± 0.62
Lee et al. (1987)	Inhalation, 6 h/d (GF6-15)	0	0	MF	22	4.0 ± 0.4
		10	19.4		20	4.4 ± 0.4
		100	199		23	4.0 ± 0.3
Hass et al. (1995)	Inhalation, 6 h/d (GD4-20)	0	0	M	20	4.7 ± 0.1
		165	943		23	4.5 ± 0.1
		0	0	F	20	4.5 ± 0.1
		165	943		23	4.3 ± 0.1
Dermal Studies						
Becci et al. (1982)	Dermal, 8 h/d (GD6-15)	0	0	MF	24	3.45 ± 0.20
		75	340		22	3.49 ± 0.24
		237	1230		23	3.54 ± 0.29
		750	3910		22	2.83 ± 0.39

*Internal doses based on PBPK modeling for rats exposed to NMP via inhalation or dermal routes as reported in Poet et al. (2016).

4. Results

4.1. Input from expert panel

4.1.1. Study quality and use in risk assessment

With respect to data quality scoring, both the BASF Department of Toxicology (1999) and the Huntingdon Life Sciences (1999) studies scored high with mean scores of 26 (out of a maximum score of 27), and the study of Exxon Biomedical Sciences (1991) scored lower with a mean of 21.7 (Fig. 1A). The Exxon Biomedical Sciences (1991) study ratings were lowest for the data quality domains for observational bias, selection bias, and confounding variables. A similar pattern was observed for overall confidence in the studies with both the BASF Department of Toxicology (1999) and Huntingdon Life Sciences (1999) studies scoring high (mean confidence score of 9 out of 10), and the Exxon Biomedical Sciences (1991) study scoring considerably lower (mean confidence score of 5 out of 10) (Fig. 1B). The study results of BASF Department of Toxicology (1999) and Huntingdon Life Sciences (1999) for the endpoints of male fertility, female fecundity, and testes weights were recommended by a majority or all the panelists to be

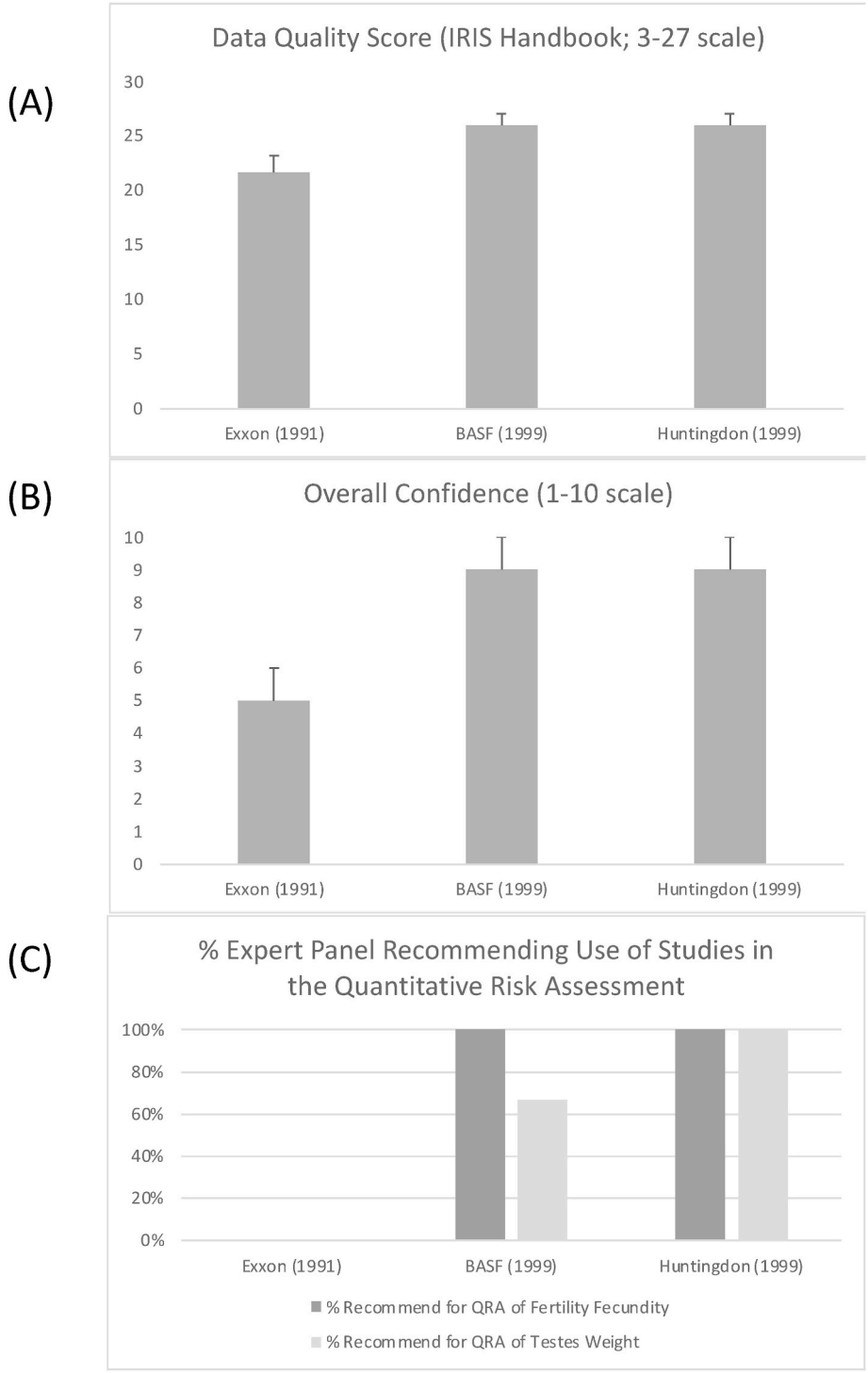

Fig. 1. Study quality, overall confidence, and recommendations for use in quantitative risk assessment.

included in the quantitative risk assessment (Fig. 1C). None of the panelists recommended including Exxon Biomedical Sciences (1991) in the quantitative risk assessment. Primary concerns with Exxon Biomedical Sciences (1991) are discussed below (see **Section 4.3**). Additionally, the expert panelists noted that the Exxon study was conducted in 1989, and therefore did not include measurements for some important endpoints (*e.g.*, organ weights, sperm analyses, oocyte counts, estrous cyclicity), which were included in the other two studies.

4.1.2. NOAEL designations

The expert panelists consistently identified NOAEL values for male fertility and female fecundity as corresponding to the mid-dose (160 mg/kg-bw/day) for Exxon Biomedical Sciences (1991), and to the high-dose (350/500 mg/kg-bw/day) for the other two studies (BASF Department of Toxicology, 1999; Huntingdon Life Sciences, 1999) (Table 5). Differences are noted for NOAEL designation for testes effects with the majority selecting the high-dose group, and one expert panelist selecting the mid-dose based on BASF Department of Toxicology (1999) and Huntingdon Life Sciences (1999). The expert panelists noted that these organ weight changes were not accompanied by any gross or histopathological changes. The expert panelists defined the mid-dose as a NOAEL for pup body weight changes in all three studies, with one expert panelist selecting the high dose for the BASF Department of Toxicology (1999) and Huntingdon Life Sciences (1999) studies.

4.1.3. Explanatory variables

As noted in Table 1, three important design differences were noted across the three studies, which may contribute to the differences in toxicity observed (Table 2). The expert panelists identified differences in adjusting the high dose [maintained at 500 mg/kg-bw/day in Exxon Biomedical Sciences (1991) but reduced to 350 mg/kg-bw/day in the other two studies] as having the greatest explanatory potential (mean rating of 8.3 out of 10). The expert panelists noted that the high-dose group in Exxon Biomedical Sciences (1991) should have been terminated or the dose adjusted to avoid exceeding the MTD. Differences in adjusting feed concentrations of NMP during lactation to offset feed consumption changes, which were not adjusted in Exxon Biomedical Sciences (1991) resulting in up to 3-fold higher than nominal levels of NMP during lactation, was also identified as a potential explanatory factor (mean rating of 5.3 out of 10) (Fig. 2). Underlying reproductive issues in the rat strain used was as explanatory as stochasticity (both variables with mean ratings of 4 out of 10). It is noted that these variables are not necessarily independent. The latter two variables may well exacerbate that first variable in that: (1) increased lactational exposures in the high-dose group of Exxon Biomedical Sciences (1991) likely resulted in lactational doses of NMP that far exceeded the MTD; and (2) due to increased litter mortality in the high-dose group for F1 litters from Exxon Biomedical Sciences (1991) fewer animals were left for F2 litter evaluations, resulting in an increased probability of sibling mating pairs.

Table 5
Expert panel determinations of NOAEL values for NMP based on reproductive endpoints.

Endpoint	Study		
	Exxon Biomedical Sciences (1991)	BASF Department of Toxicology (1999)	Huntingdon Life Sciences (1999)
Male Rat Fertility	160 mg/kg-bw/day (3)[a]	500/350 mg/kg-bw/day (3)	500/350 mg/kg-bw/day (3)
Female Rat Fecundity	160 mg/kg-bw/day (3)	500/350 mg/kg-bw/day (3)	500/350 mg/kg-bw/day (3)
Testes Weight	NM	160 mg/kg-bw/day (1) 500/350 mg/kg-bw/day (2)	160 mg/kg-bw/day (1) 500/350 mg/kg-bw/day (2)
Pup Body Weight Changes	160 mg/kg-bw/day (3)	160 mg/kg-bw/day (2) 500/350 mg/kg-bw/day (1)	160 mg/kg-bw/day (2) 500/350 mg/kg-bw/day (1)

[a] Value in parentheses indicates the number of expert panel members identifying each NOAEL value; NM = not measured.

4.2. Weight of scientific evidence and endpoint selection for risk assessment

Noncancer endpoints considered for NMP risk assessment are summarized and discussed below.

4.2.1. Male fertility/female fecundity

NMP-related effects on male fertility and female fecundity are limited to a single study (Exxon Biomedical Sciences, 1991) and not observed in higher quality studies conducted subsequently (Huntingdon Life Sciences, 1999; BASF Department of Toxicology, 1999). The expert panel concluded that the Exxon Biomedical Sciences (1991) study should not be used in risk assessment due to concerns over exceeding the MTD in the high-dose group, failure to adjust NMP feed concentrations during lactation resulting in NMP doses that were 2- to 3-fold higher than nominal levels, and/or underlying reproductive performance concerns in the rat strain used (Fig. 2). Based upon the NOAEL values of 350/500 mg/kg-bw/day from the two high-quality studies (Huntingdon Life Sciences, 1999; BASF, 1999), the weight of scientific evidence does not support identifying male fertility or female fecundity as a reliable endpoint for NMP risk assessment.

4.2.2. Testes weight

The potential effects of NMP on testes weight is supported by the two studies that assessed this endpoint (Huntingdon Life Sciences, 1999; BASF, 1999). Dose-response trends for this endpoint were weak and non-monotonic. Additionally, absolute, and relative testes weights fell within the range of historical controls. Statistically significant increases noted in both studies may have been affected by unusually low values in concurrent controls (as compared to historical controls). The increased testes weights in Huntingdon Life Sciences (1999) were not considered to be treatment related by the study authors. The lack of any gross pathology findings, histopathological findings, or changes in sperm parameters also reinforce the conclusion that these slight changes in organ weights were due to normal variation in this endpoint and not due to NMP exposure. The majority of the expert panel members supported identifying the highest dose group (500/350 mg/kg-bw/day) as a NOAEL, with one member identifying the mid-dose group (160 mg/kg-bw/day) as a NOAEL. Accordingly, the weight of scientific evidence does not support identifying testes weight changes as biologically significant, and as such are not a reliable and sensitive endpoint for NMP risk assessment.

4.2.3. Pup body weights

All three multigenerational reproductive toxicity studies for NMP reported decreases in pup body weights (Exxon Biomedical Sciences, 1991; Huntingdon Life Sciences, 1999; BASF Department of Toxicology, 1999). In the 500 mg/kg-bw/day dose groups, the effects of NMP treatment on pup PND1 body weights were large (approximately 20%) and consistent across all three studies. More moderate reductions in pup body weights were noted in animals exposed to NMP doses of 350 mg/kg-bw/day or lower, exhibiting non-monotonicity in some cases (see detailed description in **Appendix A**). The results of these three studies are also consistent with results observed in developmental toxicity studies and single generation reproductive toxicity studies (Tables 3 and 4), which exhibit a consistent trend as a function of NMP internal dose across studies and routes of exposure. The mid-dose (160 mg/kg-bw/day) was identified by most or all expert panel members as a NOAEL for all three studies. Accordingly, the weight of scientific evidence supports identifying pup body weight changes as a relevant, dependable, and sensitive endpoint for use in an NMP human health risk

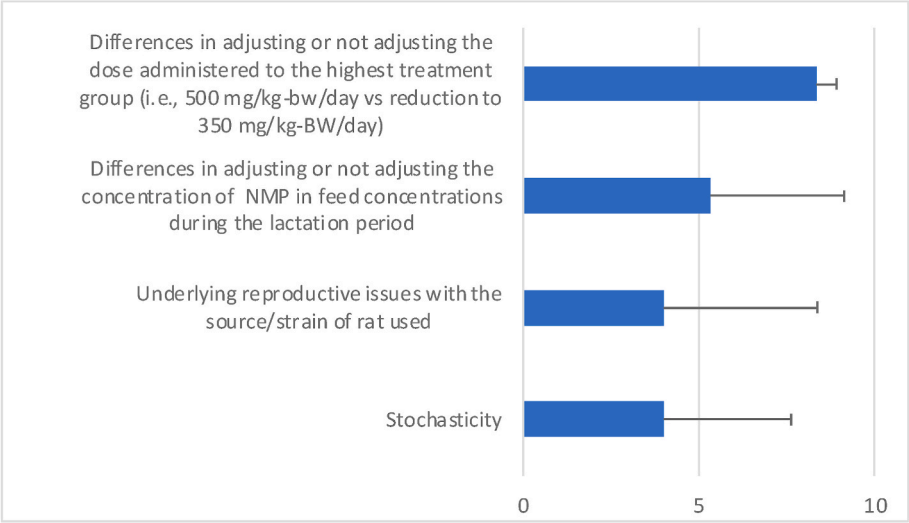

Fig. 2. Degree that Study Design Elements are Explanatory for Different Study Results (1 = not explanatory; 10 = completely explanatory).

assessment. This conclusion is consistent with EPA's NMP Work Plan Assessment (USEPA, 2015) in which the fetal body weight effects of NMP from Saillenfait et al. (2003) were selected as the quantitative basis for risk assessment (Table 4).

4.2.4. Other endpoints

All three studies reported an increase in the number of stillbirths at the highest test dose (Exxon Biomedical Sciences, 1991; Huntingdon Life Sciences, 1999; BASF Department of Toxicology, 1999). The points of departure for this endpoint derived by EPA (Table 4) are higher than those for other endpoints (*e.g.*, fetal/pup body weight changes). Accordingly, the weight of scientific evidence consistently identifies stillbirths as an effect of NMP but does not support identifying this endpoint as the most relevant and sensitive for use in an NMP human health risk assessment.

4.3. @Benchmark dose modeling

BMD modeling results for the fetal/pup body weight effects of NMP are summarized in Table 6 and depicted in Fig. 3. For the pooled data analysis, a POD value ($BMDL_{SD}$, rounded to two significant figures) of 3,300 mg*hr/L was derived. The goodness-of-fit p-value for the pooled data was low (<0.0001), which is not unexpected given the application of a simple polynomial regression model that does not account for the effects of additional variables (*e.g.*, test laboratory, rat strain, route of exposure) that may also contribute to the observed response. Visual

Table 6
Benchmark dose results for the fetal/pup body weight effects of NMP.

Reference	Litter	Sex	Model	BMR	Benchmark Dose			Model Performance Critiera		
					BMD	BMDL	BMDU	BMD:BMDL	Test 4 P-Value	AIC
Saillenfait et al. (2003)	1	M	Exponential(3)	1SD	2.1E+03	1.7E+03	2.6E+03	1.3E+00	3.9E-01	−2.7E+02
		F	Hill	1SD	2.0E+03	1.5E+03	2.5E+03	1.3E+00	9.2E-01	−2.7E+02
Sitarek et al. (2012)	1	M	Exponential(2)	1SD	2.3E+03	1.7E+03	3.7E+03	1.4E+00	9.3E-01	−1.2E+02
		F	Polynomial(2)	1SD	4.0E+03	2.6E+03	8.5E+03	1.5E+00	9.9E-01	−1.4E+02
Huntingdon/Thornton (1999)	1	M	Polynomial(3)	1SD	6.1E+03	4.2E+03	6.6E+03	1.4E+00	7.7E-01	−1.7E+02
		F	Polynomial(3)	1SD	6.2E+03	4.3E+03	6.7E+03	1.4E+00	7.6E-01	−1.6E+02
	2	M	Polynomial(2)	1SD	8.3E+03	5.8E+03	Infinity	1.4E+00	2.8E-02	−1.9E+02
		F	Polynomial(3)	1SD	7.8E+03	5.6E+03	Infinity	1.4E+00	2.1E-01	−2.0E+02
	3	M	Exponential(2)	1SD	7.8E+03	4.7E+03	2.2E+04	1.7E+00	3.1E-01	−2.3E+02
		F	Exponential(2)	1SD	6.4E+03	4.1E+03	1.4E+04	1.6E+00	9.6E-01	−2.2E+02
	4	M	Exponential(2)	1SD	7.6E+03	4.4E+03	2.3E+04	1.7E+00	3.6E-01	−1.4E+02
		F	Exponential(2)	1SD	7.4E+03	4.3E+03	2.1E+04	1.7E+00	3.4E-01	−1.3E+02
BASF Department of Toxicology (1999)	1	M	Exponential(2)	1SD	4.7E+03	3.4E+03	7.1E+03	1.4E+00	4.0E-01	−1.7E+02
		F	Exponential(2)	1SD	4.1E+03	3.1E+03	5.9E+03	1.3E+00	7.5E-01	−1.7E+02
	2	M	Polynomial(3)	1SD	6.3E+03	5.2E+03	3.3E+04	1.2E+00	8.7E-01	−2.1E+02
		F	Exponential(2)	1SD	2.4E+04	7.9E+03	Infinity	3.0E+00	4.9E-01	−1.8E+02
	3	M	Polynomial(3)	1SD	6.2E+03	5.1E+03	3.5E+04	1.2E+00	6.1E-01	−1.7E+02
		F	Polynomial(3)	1SD	6.8E+03	5.2E+03	Infinity	1.3E+00	2.1E-01	−1.6E+02
	4	M	Polynomial(3)	1SD	6.4E+03	5.2E+03	5.0E+04	1.2E+00	8.9E-01	−2.0E+02
		F	Polynomial(3)	1SD	6.1E+03	5.1E+03	2.9E+04	1.2E+00	1.0E+00	−1.9E+02
Saillenfait et al. (2003)	1	M	Exponential(2)	1SD	8.1E+02	5.1E+02	1.9E+03	1.6E+00	9.5E-01	−2.4E+02
		F	Exponential(2)	1SD	8.2E+02	5.2E+02	1.9E+03	1.6E+00	8.4E-01	−2.1E+02
Solomon et al. (1995)	1	MF	Exponential(2)	1SD	6.1E+02	4.2E+02	1.1E+03	1.4E+00	1.4E-01	−1.7E+02
Lee et al. (1987)	1	MF	NA (lack of dose-response trend)							
Hass et al. (1995)	1	M	NA (single dose group)							
		F	NA (single dose group)							
Becci et al. (1982)	1	MF	Polynomial(3)	1SD	2.9E+03	2.4E+03	3.2E+03	1.2E+00	6.0E-01	−1.9E+02
Geometric Mean				*1SD*	*4.3E+03*	*2.9E+03*	*8.0E+03*	1.5E+00		
Pooled Data	all	all	Polynomial(3)	0.5SD	3.7E+03	3.3E+03	4.0E+03	1.1E+00	<0.0001	−4.4E+03

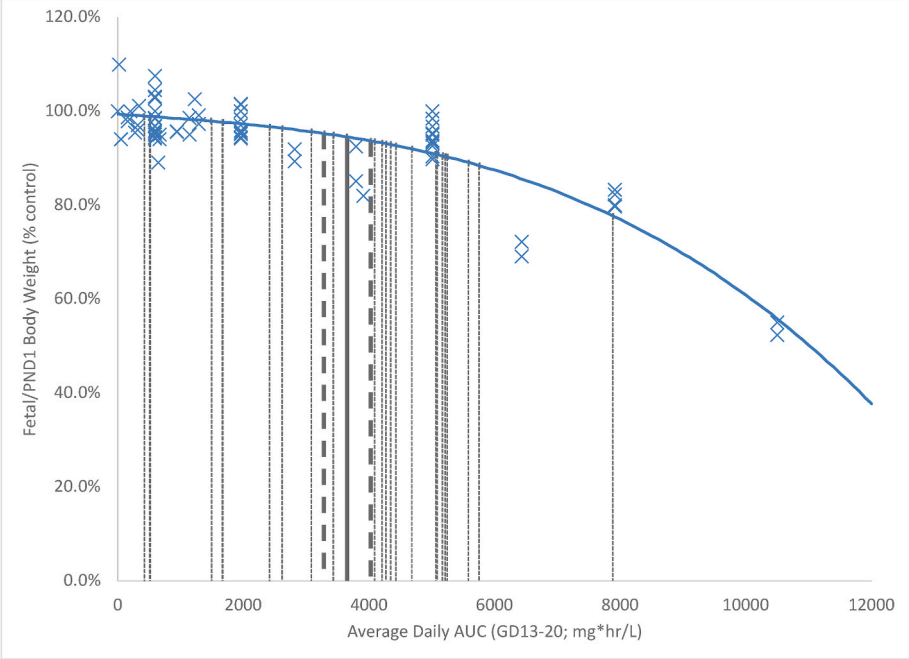

Fig. 3. Benchmark Dose Results for Fetal/Pup Body Weight Changes of NMP in Rats. Solid vertical line = $BMD_{0.5SD}$ for pooled data set; heavy dashed vertical lines = 90% confidence interval for pooled data set ($BMDL_{0.5SD}$, $BMDU_{0.5SD}$); light dashed vertical lines = $BMDL_{1SD}$ values based on individual data sets.

inspection of the polynomial model (Fig. 3) suggests that the fit appears reasonable. The pooled POD value does not, however, satisfy the predetermined model fit criteria, as discussed under Section 3.2. Notwithstanding this limitation, it is noteworthy that the pooled POD value is within the range of values EPA (2015) calculated using the pooled oral and inhalation datasets from Saillenfait et al. (2003, 2003) (*i.e.*, $BMDL_{0.5RD}$ = 1,424 mg*hr/L) and the dermal POD value from Becci et al. (1982) (*i.e.*, $BMDL_{0.5RD}$ = 4,018 mg*hr/L). The POD values calculated herein on individual data sets ranged from 420 to 7,900 mg*hr/L, with an arithmetic mean of 3,700 mg*hr/L. Further, goodness-of-fit p-values for individual data sets were much higher with most satisfying the predetermined model fit criteria, including EPA's recommendation of a goodness-of-fit p-value greater than 0.1.

5. Discussion & conclusions

An expert panel on the evaluation of reproductive toxicology studies and the application of their data to human health risk assessment was assembled to provide input on the quality, interpretation, and application of data from three multi-generation reproductive toxicity studies for NMP. This panel engagement utilized a multi-round Delphi format. Charge question responses saw only small changes between Round 1 (collection of independent input) and the final responses (collection of deliberative input) (*i.e.*, no changes to Fig. 1A and C, or Fig. 2; small changes to Fig. 1B). The primary conclusions of this expert panel are that the study of Exxon Biomedical Sciences (1991) is not a high-quality study, and due to several design flaws (Table 1) should not be used for human health risk assessment of NMP. This conclusion stands out in contrast to the data quality review performed by EPA, which concluded that all three studies were of high quality (USEPA, 2020c). Based on the expert panel's recommendations, EPA's selection of male fertility data from Exxon Biomedical Sciences (1991) as the basis for the quantitative human health risk assessment is not supported.

Consideration of the weight of scientific evidence for reproductive toxicity, fetal/pup body weight changes are identified as the most sensitive based on a consideration of the POD values calculated (Table 4). Although the lowest POD value was calculated for the one-generation reproductive toxicity study of DuPont (1990)/Solomon et al. (1995) (human equivalent dose of 27 mg/kg-bw/day, Table 4), EPA (2015) concluded that "… *the dose-response relationship in the DuPont study was not as robust as the Saillenfait study. Lower variability in body weights was observed in the Saillenfait study than in the DuPont study. In the DuPont study, statistically significant differences only occurred in the lowest and highest dose groups, not the middle dose group*" (USEPA, 2015). For this reason, the POD for fetal body weight changes based on Saillenfait et al. (2003) (human equivalent dose of 48 mg/kg-bw/day, Table 4) serves as an appropriate basis for human health risk assessment and is consistent with EPA's conclusion in their 2015 assessment (USEPA, 2015).

A comparison of recently derived POD values for NMP is provided in

Table 7
Comparison of POD values for NMP.

Parameter	NMP Workplan Assessment (USEPA, 2015)	Poet et al. (2016)	NMP Risk Evaluation (USEPA, 2022b)	This Assessment
Endpoint	Fetal Body Weight Changes	Fetal Body Weight Changes	Male Rat Fertility	Fetal/Pup Body Weight Changes
Key Study	Saillenfait et al. (2003)	Saillenfait et al. (2003); Solomon et al. (1995)	Exxon Biomedical Sciences (1991)	Pooled data set, excluding Exxon Biomedical Sciences (1991) (see Tables 3 and 4)
POD Type[a]	$BMDL_{05RD}$	Geometric mean of two $BMDL_{1SD}$ values	$BMDL_{10ER}$	$BMDL_{0.5SD}$
POD Value (AUC, mg*hr/L; rounded to 2 significant figures)	410	470	180	3,300

[a] RD = relative deviation; SD = standard deviation; ER = extra risk.

Table 7. The POD value derived here is more than an order of magnitude higher than derived previously by EPA (2015, 2020) or Poet et al. (2016). The primary reason for this difference is the reliance on a pooled data set for body weight changes, which reflects our current state of knowledge for the exposure-response relationship for this endpoint, rather than selecting a single conservative study (or two studies). Pooling data of separate data sets is consistent with USEPA BMD guidelines (USEPA, 2012), and was evaluated by EPA for chronic effects in the NMP Work Plan Assessment and selected by EPA in their assessment of acute effects in the NMP Work Plan Assessment and the NMP Risk Evaluation (USEPA, 2015, 2020). However, given the unacceptable model fits obtained using the pooled data for fetal/pup body weight changes, the POD values derived by EPA (2015) and Poet et al. (2016) for fetal body weight change represent the most robust and health protective POD values for use in an NMP human health risk assessment.

From the POD values summarized in Table 6, the potency of NMP in reducing fetal/pup body weights appears greater following inhalation exposures (Saillenfait et al., 2003; Solomon et al., 1995) when compared to other routes of exposure. As discussed by Poet et al. (2016), the route-dependency of NMP's potency for this endpoint may be an artifact of the physiologically based pharmacokinetic (PBPK) model used to estimate internal dose. Specifically, the PBPK model's reliance upon a nose-only exposure study in rats for model parameters may result in underestimating internal doses of NMP following whole-body exposures to rats by not considering contributions from other exposure pathways (*e.g.*, dermal absorption of NMP vapor; ingestion of NMP adsorbed to fur when grooming). Poet et al. (2016) identified additional research on the relative importance of these pathways as a data need for NMP dosimetry. Despite these potential limitations in the inhalation studies, these studies were included in the pooled POD analysis, which leads to a more conservative number, and as noted, a value that is within the range of values derived by EPA (2015) for pooled oral and inhalation studies and the available dermal study.

The foregoing discussion focused on the evaluation of studies, endpoints, and POD values using methods that meet the scientific standards under TSCA Section 26, thereby representing the best available science and weight of scientific evidence. It is important to note, however, that there were other indications in the NMP Risk Evaluation, beyond EPA's evaluation of the three studies discussed herein, that suggest EPA did not perform an adequate review of the available data in a manner that would satisfy the TSCA Section 26 requirements. For example, in the NMP Work Plan Assessment, EPA concluded that the reproduction and developmental toxicity study performed by Sitarek and Stetkiewicz (2008) was "unreliable" due to inconsistencies in the published data (USEPA, 2015). In comparison, EPA assigned a data quality rating of "High" to this study in the NMP Risk Evaluation, which corresponded to the descriptor "No notable deficiencies or concerns are identified in the domain metric that are likely to influence results [score of 1]" (USEPA, 2018, 2020). The NMP Risk Evaluation did not, however, discuss the inconsistencies EPA had previously identified in the published data (USEPA, 2022b). Despite this, on July 1, 2022, EPA revised and re-issued the unreasonable risk determination on NMP in preparation for drafting a risk management regulation, as discussed below (USEPA, 2022a).

In the revised unreasonable risk determination document, EPA stated that it "*views the peer reviewed hazard and exposure assessments and associated risk characterization [in the NMP Risk Evaluation that used the 2018 SR Guidance Document] as robust and upholding the standards of best available science and weight of the scientific evidence,* per *TSCA sections 26 (h) and (i)*" (USEPA, 2022a). Therefore, it is unclear whether EPA will consider additional available information when it finalizes the unreasonable risk determination document.

In conclusion, the study design of Exxon Biomedical Sciences (1991) was replicated and improved (*i.e.*, eliminating design flaws, as summarized in Table 1) in Huntingdon Life Sciences (1999) and BASF Department of Toxicology (1999). Some of the critical effects observed in Exxon Biomedical Sciences (1991) are not confirmed in the other two studies. Specifically, Huntingdon (1999) and BASF Department of Toxicology (1999) failed to confirm any test article-related effects on male fertility and female fecundity in any treatment group. Exclusion of the Exxon Biomedical Sciences (1991) study from the quantitative risk assessment per the expert panel's recommendations has a ripple effect in the risk assessment with respect to identifying the most sensitive endpoint. Due to the overall consistency of observations noted for fetal/pup body weight changes following oral, inhalation, and dermal exposures to NMP (Fig. 3), and the relative sensitivity of this endpoint compared to other endpoints (testis weight changes, stillbirths), fetal/pup body weight changes are identified as the most appropriate/sensitive endpoint for NMP risk assessment. Based on the large database of studies on this endpoint for NMP, the POD values of 410 or 470 mg*hr/L, as derived in the respective evaluations by EPA (2015) and Poet et al. (2016), can be used to support human health risk assessment decisions for this chemical.

Funding body information

This work was funded by the NMP Producers Group (signed proposal 081721); 2200 Pennsylvania Avenue, NW, Suite 100W, Washington, D. C. 20037

Credit author statement

Christopher Kirman: Conceptualization, Methodology, Writing Original Draft & Revisions, Project Administration, Funding Acquisition; **Babasaheb Sonawane:** Expert Panelist**,** Writing – Review and Editing; **Jennifer Seed:** Expert Panelist, Writing – Review and Editing; **Nkem Azu:** Writing – Review and Editing; **Wade Barranco:** Writing – Review and Editing; **Ryan Hamilton:** Writing – Review and Editing; **Todd Stedeford:** Writing Original Draft & Revisions; **Sean Hays:** Conceptualization, Writing – Review and Editing, Funding Acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: This work was funded by the NMP Producers Group (signed proposal 081721), and member company representatives (NOA, WTB, WRH) and group management representatives (TJS) serve as coauthors. Authors CRK and SMH are owners of Summit Toxicology, and therefore have a financial interest in this work.

Data availability

No data was used for the research described in the article.

Acknowledgements

This work was funded by the NMP Producers Group. Source of funding was not disclosed to the experts during recruitment, selection, and engagement, and disclosed after the review was complete in preparing this manuscript. Similarly, the Sponsor of the review was blinded to the identities of the experts during recruitment, selection, and engagement, and their identities were revealed after the review in preparing this manuscript. All expert panelists were provided an opportunity to review and edit a draft of this manuscript, of which two accepted (co-authors JS and BS). We would like to thank Drs. Anthony Scialli (Scialli Consulting) for his valuable insights while participating in the expert panel. We would also like to thank Dr. Wil Faber (independent consultant) for his assistance in preparing sections of the white paper. The authors, CRK and SMH, are owners of SciPinion and thus have a financial interest in the content of this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.yrtph.2023.105337.

References

BASF Department of Toxicology, 1999. Two Generation Reproduction Toxicity Study with N-Methylpyrrolidone (NMP) in Wistar Rats - Administration in the Diet. Department of Toxicology of BASF Aktiengesellschaft, Ludwigshafen, Germany (Project No. 70R0056/97008).

Becci, P.J., Knickerbocker, M.J., Reagan, E.L., Parent, R.A., Burnette, L.W., 1982. Teratogenicity study of N-methylpyrrolidone after dermal application to Sprague-Dawley rats. Mar-Apr Fund. Appl. Toxicol. 2 (2), 73–76. https://doi.org/10.1016/s0272-0590(82)80117-6. PMID: 7185604.

Dupont, 1990. 1-Methyl-2-Pyrrolidinone (NMP): reproductive and developmental toxicity in the rat. Dupont HLR 294, 90.

Exxon Biomedical Sciences, 1991. Multigeneration Rat Reproduction Study with N-Methylpyrrolidone. Project No. 236535, Nov. 26. GAF Corp, Wayne, NJ.

Hass, U., Jakobsen, B.M., Lund, S.P., 1995. Developmental toxicity of inhaled N-methylpyrrolidone in the rat. Jun Pharmacol. Toxicol. 76 (6), 406–409. https://doi.org/10.1111/j.1600-0773.1995.tb00170.x.PMID:7479584.

Huntingdon Life Sciences, 1999. Two Generation Reproduction Toxicity Study with N-Methylpyrrolidone (NMP) in Sprague Dawley Rats: Administration in the Diet. (Project No. 97-4106). Huntingdon Life Sciences, Millstone, NJ.

IRIS, 2022. Integrated Risk Information System. U.S. Environmental Protection Agency, Washington, DC. Online database records accessed March 2022.

Kirman, C.R., Simon, T.W., Hays, S.M., 2019. Science peer review for the 21st century: assessing scientific consensus for decision-making while managing conflict of interests, reviewer, and process bias. Apr Regul. Toxicol. Pharmacol. 103, 73–85. https://doi.org/10.1016/j.yrtph.2019.01.003. Epub 2019 Jan 8. PMID: 30634024.

Lee, K.P., Chromey, N.C., Culik, R., Barnes, J.R., Schneider, P.W., 1987. Toxicity of N-methyl-2-pyrrolidone (NMP): teratogenic, subchronic, and two-year inhalation studies. Aug Fund. Appl. Toxicol. 9 (2), 222–235. https://doi.org/10.1016/0272-0590(87)90045-5. PMID: 3653566.

NASEM, 2021. National Academies of Sciences, Engineering, and Medicine. 2021. The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations. The National Academies Press, Washington, DC. https://doi.org/10.17226/25952.

Poet, T.S., Schlosser, P.M., Rodriguez, C.E., Parod, R.J., Rodwell, D.E., Kirman, C.R., 2016. Using physiologically based pharmacokinetic modeling and benchmark dose methods to derive an occupational exposure limit for N-methylpyrrolidone. Apr Regul. Toxicol. Pharmacol. 76, 102–112. https://doi.org/10.1016/j.yrtph.2015.12.020. Epub 2016 Jan 8. PMID: 26776754.

Saillenfait, A.M., Gallissot, F., Morel, G., 2003. Developmental toxicity of N-methyl-2-pyrrolidone in rats following inhalation exposure. Apr Food Chem. Toxicol. 41 (4), 583–588. https://doi.org/10.1016/s0278-6915(02)00300-9. PMID: 12615131.

Sitarek, K., Stetkiewicz, J., 2008. Assessment of reproductive toxicity and gonadotoxic potential of N-Methyl-2-pyrrolidone in male rats. Int. J. Occup. Med. Environ. Health 21 (1), 73–80.

Sitarek, K., Stetkiewicz, J., Wąsowicz, W., 2012. Evaluation of reproductive disorders in female rats exposed to N-methyl-2-pyrrolidone. Jun Birth Defects Res B Dev Reprod Toxicol 95 (3), 195–201. https://doi.org/10.1002/bdrb.21001. Epub 2012 Feb 6. PMID: 22311695.

Solomon, H.M., Burgess, B.A., Kennedy Jr., G.L., Staples, R.E., 1995. 1-Methyl-2-pyrrolidone (NMP): reproductive and developmental toxicity study by inhalation in the rat. Nov Drug Chem. Toxicol. 18 (4), 271–293. https://doi.org/10.3109/01480549509014324.PMID:8586021.

USEPA, 2012. Benchmark Dose Technical Guidance. Risk Assessment Forum. U.S. Environmental Protection Agency (USEPA), Washington, DC. EPA/100/R-12/001, 99 pp., available at: https://www.epa.gov/sites/default/files/2015-01/documents/benchmark_dose_guidance.pdf.

USEPA, 2015. TSCA Work Plan Chemical Risk Assessment, N-Methylpyrrolidone: Paint Stripper Use, CASRN: 872-50-4. Office of Chemical Safety and Pollution Prevention, Washington, DC. https://www.epa.gov/sites/default/files/2015-08/documents/nmp_ra_3_23_15_final.pdf.

USEPA, 2018. Application of Systematic Review in TSCA Risk Evaluations. Office of Chemical Safety and Pollution Prevention, U.S. Environmental Protection Agency (USEPA), Washington, DC. EPA 740-P1-8001, 248 pp., available at: https://www.epa.gov/sites/default/files/2018-06/documents/final_application_of_sr_in_tsca_05-31-18.pdf.

USEPA, 2020a. Additional Studies in Support of the Risk Evaluation for N-Methylpyrrolidone, Systematic Review Data Extraction: Data Extraction Tables for Reproductive and Developmental Effects Reported in Two-Generation Reproduction Studies Following Dietary Exposure. Office of Chemical Safety and Pollution Prevention. June 2020.

USEPA, 2020b. n-Methylpyrrolidone (2-Pyrrolidinone, 1-Methyl-) (NMP) Final Risk Evaluation. Non-technical Summary. December 2020.

USEPA, 2020c. Risk Evaluation for N-Methylpyrrolidone (2-Pyrrolidinone, 1-Methyl) (NMP). EPA-740-R1-8009, December 2020.

USEPA, 2020d. Final risk evaluation for n-methylpyrrolidone systematic review supplemental file: data quality evaluation of human health hazard studies– animal and in vitro studies. CASRN 872–50–4. December 2020.

USEPA, 2022a. n-Methylpyrrolidone [sic] (NMP); draft revision to toxic substances control Act (TSCA) risk determination; notice of availability and request for comment, agency: environmental protection agency (EPA). Action: Notice, Federal Register 87, 39511–39517 available at: https://www.govinfo.gov/content/pkg/FR-2022-07-01/pdf/2022-14108.pdf.

USEPA, 2022b. The EPA Needs to Improve the Transparency of its Cancer-Assessment Process for Pesticides. Office of Inspector General, U.S. Environmental Protection Agency, Washington, DC. Report No. 22-E-0053. July 20, 2022.