Regulatory Toxicology and Pharmacology 71 (2015) 114–124

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

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

Regulatory acceptance and use of the Extended One Generation Reproductive Toxicity Study within Europe

Marie-Jeanne W.A. Schiffelers [a,*], Bas J. Blaauboer [b], Wieger E. Bakker [a], Coenraad F.M. Hendriksen [c,d], Cyrille Krul [e,f]

[a] Utrecht University School of Governance, Bijlhouwerstraat 6, 3511 ZC Utrecht, The Netherlands
[b] Utrecht University, Institute for Risk Assessment Sciences (IRAS), P.O. Box 80.178, 3508 TD Utrecht, The Netherlands
[c] Utrecht University, Faculty of Veterinary Medicine, Department Animals in Science and Society, P.O. Box 80.163, 3508 TD Utrecht, The Netherlands
[d] Institute for Translational Vaccinology (Intravacc), P.O. Box 450, 3720 AL Bilthoven, The Netherlands
[e] TNO, P.O. Box 360, 3700 AJ Zeist, The Netherlands
[f] University of Applied Sciences Utrecht, Life Sciences & Chemistry, F.C. Dondersstraat 65, 3572 JE Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 1 July 2014
Available online 28 October 2014

Keywords:
Reproductive toxicity testing
EOGRTS
Regulatory acceptance and use
Barriers and drivers
REACH

ABSTRACT

The two-generation study (OECD TG 416) is the standard requirement within REACH to test reproductive toxicity effects of chemicals with production volumes >100 tonnes. This test is criticized in terms of scientific relevance and animal welfare. The Extended One Generation Reproductive Toxicity Study (EOGRTS), incorporated into the OECD test guidelines in 2011 (OECD TG 443) has the potential to replace TG 416, while using only one generation of rats and being more informative. However, its regulatory acceptance proved challenging. This article reconstructs the process of regulatory acceptance and use of the EOGRTS and describes drivers and barriers influencing the process. The findings derive from literature research and expert interviews. A distinction is made between three sub-stages; The stage of Formal Incorporation of the EOGRTS into OECD test guidelines was stimulated by retrospective analyses on the value of the second generation (F2), strong EOGRTS advocates, animal welfare concern and changing US and EU chemicals legislation; the stage of Actual Regulatory Acceptance within REACH was challenged by legal factors and ongoing scientific disputes, while the stage of Use by Industry is influenced by uncertainty of registrants about regulatory acceptance, high costs, the risk of false positives and the manageability of the EOGRTS.

1. Introduction

Chemicals are subjected to a broad range of requirements to guarantee safety for humans, animals and the environment. The requirements describe the endpoints for which chemical substances have to be assessed and generally also the test procedures that need to be performed for a particular endpoint. Reproductive and developmental toxicity are two of the main endpoints in the assessment of industrial- and agrochemicals. These endpoints include the toxic effects of a substance on an organism's reproduction and development of its offspring. The reproduction cycle of mammals, being a highly complex process, is very difficult to investigate *in vitro*. For this reason regulatory reproductive and developmental toxicity tests are still conducted in laboratory animals with a prenatal developmental study in rodents and a non-rodent species and a one- or two generation reproduction toxicity study in rats (Janer et al., 2007b).

Since the 1980s the OECD 416 two-generation study has been the most comprehensive reproductive toxicity study (OECD, 2001). Up to 30% of the reproductive toxicity tests conducted are

Abbreviations: ACSA, Agricultural Chemical Safety Assessment of the ILSI HESI; ARA, Actual Regulatory Acceptance by regulatory authorities; CARACAL, Competent Authorities for REACH and GLP; CLP, classification and labelling process; CMR, Carcinogenic, Mutagenic and Reprotoxic; DIT, developmental immunotoxicity; DNT, developmental neurotoxicity; EC, European Commission; ECETOC, European Centre for Ecotoxicity and Toxicology of Chemicals; EURL-ECVAM, European Centre for the Validation of Alternative Methods; ECHA, European Chemicals Agency; EOGRTS, Extended One Generation Reproductive Toxicity Study; EOGRTS EG, Expert group on the EOGRTS established within CARACAL; F2, second generation of offspring; FI, Formal Incorporation into regulatory requirements; HESI, Health and Environmental Sciences Institute; ICAPO, International Council on Animal Protection in OECD programmes; ILSI, International Life Sciences Institute; MSC, Member States Committee of ECHA; MSCA, Member State Competent Authority of ECHA; OECD, Organisation for Economic Co-operation and Development; OPTS, EPA Office of Pesticides and Toxic Substances; RAC, Risk Assessment Committee of ECHA; REACH, Registration, Evaluation, Authorization and restriction of CHemicals; RIVM, Dutch National Institute for Public Health and the Environment; TG, test guideline; TG 416, test guideline two-generation reproductive toxicity study; TG 443, test guideline EOGRTS; TG 426, test guideline neuro developmental study; UI, Use by Industry for regulatory purposes; US EPA, United States Environment Protection Agency.

* Corresponding author. Fax: +31 30 2537200.
E-mail address: m.j.w.a.schiffelers@uu.nl (M.-J.W.A. Schiffelers).

http://dx.doi.org/10.1016/j.yrtph.2014.10.012
0273-2300/

two-generation studies (Spielmann and Vogel, 2006), requiring around 2600 animals per study (Lilienblum et al., 2008). The two generation test is estimated to use nearly 40% of the laboratory animals under REACH (Janer et al., 2007a) and thereby is one of the major users of rodents in safety test programs.

In anticipation of the introduction of the European Directive for the Registration, Evaluation, Authorization and restriction of CHemicals – REACH (EU, 2006) concern was expressed that reproductive toxicity testing would lead to a significant increase in numbers of animals needed. Reproductive and developmental toxicity were even estimated to become the largest animal user for safety testing within REACH (Pedersen et al., 2003; Van der Jagt et al., 2004) since approximately 10,000 chemicals with an annual volume of >100 tonnes would have to be tested on reproductive toxicity. The estimates ranged from 40% to 90% of the total number of animals to comply with REACH that would be needed for reproductive toxicity testing purposes (Van der Jagt et al., 2004; Spielmann and Vogel, 2006; Hartung and Rovida, 2009; Martin et al., 2011). At about the same time, several studies became available that questioned the added value of the second generation (Cooper et al., 2006; Janer et al., 2007a,b; Martin et al., 2009; Piersma et al., 2011) and criticized the limited predictive value of the OECD TG 416 for developmental immunotoxic and neurotoxic parameters (See Section 2.1.).

In 2006 the Agricultural Chemical Safety Assessment (ACSA) Technical Committee of the ILSI Health and Environment Sciences Institute (HESI) proposed a whole new testing paradigm, which constituted a tiered approach of toxicity testing. Part of this paradigm was a proposal for an alternative protocol for OECD TG 416 which required only one generation of animals while being more informative in data obtained (Cooper et al., 2006). This protocol became the basis for the Extended One Generation Reproductive Toxicity Study – EOGRTS – with a reduction of up to 40% in animal use – i.e. a total of 1200 animals per study – compared to the two-generation study. In addition the EOGRTS protocol includes parameters for developmental neurotoxicity-DNT – and developmental immunotoxicity-DIT –. The Cooper protocol was proposed to the OECD secretariat for incorporation into the OECD guidelines in 2007 and accepted in 2011 as OECD TG 443 after a process in which many amendments were made, as will be described in Section 2.1. of this manuscript.[1]

The EOGRTS matches with the ambition of the European Commission to diminish the use of laboratory animals and to stimulate the acceptance and use of models to replace, reduce and refine (3Rs) existing animal models (Russel and Burch, 1959). This ambition is laid down in Directive 2010/63/EU on the protection of animals used for scientific purposes and in REACH. Directive 2010/63/EU states in article 13.2 that in choosing between procedures, those which use the minimum number of animals shall be selected (EU, 2010). Furthermore, REACH states in article 25 (1) that in order "*…to avoid unnecessary animal testing, testing on vertebrate animals for the purpose of this Regulation shall be undertaken only as a last resort*". (EU, 2006) (See also Section 2.2.). Despite these legislative stimulants and the incorporation of the EOGRTS into the OECD test guidelines, the regulatory acceptance and use of the EOGRTS within Europe has been a point of strong disparity. This raises the following key questions which will be addressed in this paper:

- *Which factors influence the regulatory acceptance and use of the EOGRTS within Europe?*[2]
- *What is needed to augment the current process?*
- *Which lessons can be drawn from the case of the EOGRTS for future processes?*

To improve the use of the 3Rs-congruent with the EC's ambition – an exhaustive comprehension of the process of regulatory acceptance and use and its drivers and barriers is needed. In order to understand and examine the regulatory process, we made a distinction between the following three successive stages:

Sub stages of regulatory acceptance and use of a new test method
FI: Formal Incorporation into the OECD test guidelines
ARA: Actual Regulatory Acceptance by regulatory authorities
UI: Use for regulatory purposes by Industry

Full regulatory acceptance and use means that a 3R model has passed all three stages.

This manuscript builds on earlier work of the authors (Schiffelers et al., 2012, 2014) which examined the process of regulatory acceptance and use of 3R models from a technology acceptance perspective (see also Section 3). The reconstruction of the EOGRTS case offers additional in depth knowledge of this process.

The EOGRTS is currently in a critical phase. Although there is agreement on the inclusion of the EOGRTS in the fifth adaptation of the REACH test methods regulation, the discussion on the actual regulatory acceptance (ARA) and the use of the EOGRTS by industry (UI) for the release of chemicals is still taking place within Europe. Disentangling the process from a more general perspective of technology acceptance can offer relevant input for this discussion and lessons for future processes.

2. Results

This section reconstructs the process of the acceptance and use of the EOGRTS and gives an overview of the barriers and drivers on this process throughout the three sub stages of Formal Incorporation (FI) of the EOGRTS in the OECD Test Guidelines – Section 2.1.; the Actual Regulatory Acceptance (ARA) by European regulatory authorities for chemical registration and authorization purposes under REACH – Section 2.2.; and the Use by Industry (UI) for chemical registration and authorization purposes under REACH – Section 2.3. The findings derive from examination of available documents connected to the acceptance process (e.g. meeting- and workshop reports) and a series of interviews with experts involved in this process (see Appendix A for a description of the methodology). To elucidate the results, several quotes from respondents are inserted in the description of drivers and barriers.

[1] "*This Test Guideline is designed to provide an evaluation of reproductive and developmental effects that may occur as a result of pre- and postnatal chemical exposure as well as an evaluation of systemic toxicity in pregnant and lactating females and young and adult offspring. In the assay, sexually-mature males and females rodents (parental (P) generation) are exposed to graduated doses of the test substance starting 2 weeks before mating and continuously through mating, gestation and weaning of their pups (F1 generation). At weaning, pups are selected and assigned to cohorts of animals for reproductive/developmental toxicity testing (cohort 1), developmental neurotoxicity testing (cohort 2) and developmental immunotoxicity testing (cohort 3). The F1 offspring receive further treatment with the test substance from weaning to adulthood. Clinical observations and pathology examinations are performed on all animals for signs of toxicity, with special emphasis on the integrity and performance of the male and female reproductive systems and the health, growth, development and function of the offspring. Part of cohort 1 (cohort 1B) may be extended to include an F2 generation; in this case, procedures for F1 animals will be similar to those for the P animals*".; http://www.oecd-ilibrary.org/environment/test-no-443-extended-one-generation-reproductive-toxicity-study_9789264122550-en.

[2] Although this paper focusses on the European situation, major parts of the discussion in the US are also covered in this manuscript.

2.1. The Formal Incorporation (FI) of the EOGRTS in the OECD test guidelines

From 2006 onwards several developments have led to the development of the EOGRTS and its incorporation into the OECD test guidelines, with the ILSI HESI strategy for agricultural chemical safety assessment (ACSA) (Cooper et al., 2006) as the starting point. Shortly after the publication of the Cooper report, several groups of experts examined the possibilities of leaving out the second generation for other products such as industrial chemicals.[3] This discussion was especially important in the light of the new REACH regulation that was estimated to lead to a significant increase in reproductive toxicity studies (see Section 1). The EOGRTS was seen as a possible answer to some of the future questions in terms of the risk assessment of chemical substances, such as meeting the REACH deadlines for CMR (Carcinogenic, Mutagenic and Reprotoxic) substances, the animal welfare concern related to reproductive toxicity testing under REACH and mounting questions regarding endocrine disruptors. The conclusion was that the EOGRTS, as proposed by the ACSA project, was applicable for industrial chemicals, if handled in a flexible way and modified to the existing requirements.[4]

Next to this European sense of urgency there was pressure from the side of the US EPA due to the aspiration to incorporate the EOGRTS into the OCSPP (Office of Chemical Safety and Pollution Prevention) guidelines that needed to be revised. The attention for the topic on both sides of the Atlantic led to a shared initiative in 2007 by the US, Germany and the Netherlands to submit a proposal to the OECD secretariat to draft an OECD test guideline (TG) based on the EOGRTS. An expert group was formed which drafted the guideline through a series of teleconferences and meetings chaired by the US and the Netherlands. In 2009, the OECD member states, except for Sweden, agreed on a draft of the TG (Gilbert, 2011).

To scientifically booster the process, four retrospective reviews were conducted (Janer et al., 2007a; Martin et al., 2009; Piersma et al., 2011; Rorije et al., 2011).[5] The retrospective analyses of Janer et al., Martin et al. and Piersma et al. concluded that the second generation has very limited added value and is not essential to establish the lowest effect level (LEL)/lowest observed adverse effect level (LOAEL) of the substances under examination. The study of Rorije et al. concluded that the second generation did not play a crucial role in the classification decision of 50 classified reproductive toxicants in Europe. Moreover, these studies underlined that the EOGRTS includes a more comprehensive evaluation, offering more information than the current two generation study, while using far less animals. The studies therefore concluded that the existing testing strategies for reproductive toxicity testing needed revision in favour of the EOGRTS.

The national authorities that submitted the proposal to the OECD, were strong advocates of the new protocol and proved successful in involving frontrunners within the American/European agrochemical industry, ICAPO-International Council on Animal Protection- and other OECD Member States. The extensive animal use, the time consumed by the two generation study and the limited added value of the second generation, were the dominant drivers for these stakeholders to support the FI of the EOGRTS in the OECD TG's. Subsequently some countries took the lead (US and NL) in keeping the EOGRTS high on the OECD agenda and guiding the protocol through the OECD process. Five years after the proposal was sent to the OECD secretariat, the EOGRTS was formally incorporated into the OECD TG's (OECD, 2011a).

The process of FI of the EOGRTS is often viewed upon as a success in terms of formal regulatory acceptance, however this does not mean that the process went effortless. Or as one of the respondents said: "*This dream became reality in 4 years' time, but it was a very rough ride through the OECD expert committee*".

Four of the points of discussion throughout this process of FI are elaborated on below.

Firstly, the issue of the added value of the second generation; this issue gave rise to the dominant discussion throughout the process of FI. The studies of Janer et al. (2007a,b), Martin et al. (2009). Piersma et al. (2011) and Rorije et al. (2011) did not convince all parties involved. Some stakeholders (e.g. the European Chemicals Agency (ECHA), Sweden and France) questioned whether the retrospective analyses covered all categories of compounds and/or whether the involved studies did meet the right criteria to reflect the reality of risk assessment and classification and labelling (e.g. Ruden and Hansson, 2008). To meet the concerns of those stakeholders who did not feel at ease with leaving out the second generation the OECD Joint Meeting (JM) agreed not to specify the EOGRTS as a replacement for the two-generation test (OECD TG 416). The guideline leaves it to the competent authority when and how the EOGRTS can be accepted as a replacement for TG 416.

Secondly, a discussion took place on the incorporation of clear triggers for a second generation. This discussion was held in the context of the US legal requirement to perform a two-generation reproduction and fertility study for food-use pesticides (Cooper et al., 2006). The triggers (i.e. effects that can be found in the first generation of animals) were to be interpreted as signals that a second generation of animals was needed to monitor the full effect of a substance. The US developed a broad range of triggers for this purpose, but the evaluation of these triggers revealed that they would lead to a second generation study in about 48% of the examined studies, while in only 1% of the cases there was a scientific justification for doing so. This way of operating would lead to an unjustifiable increase in the use of laboratory animals and therefore resulted in the decision to address this issue in a guidance document-GD[6] 117-supporting OECD TG 443 (OECD, 2011b).

A third discussion point concerned the reduction of the premating exposure period from 10 weeks in the two generation study to

[3] For instance, both the European Centre for the Validation of Alternative Methods (ECVAM) and the European Partnership for Alternative Approaches to Animal Testing (EPAA) organized workshops in 2006 with experts from industry and regulatory authorities to evaluate the applicability of this approach for industrial chemicals.

[4] The EOGRTS, as proposed by the ACSA project was developed for agrochemicals. It therefore addresses much more endpoints as requested under REACH for industrial chemicals. This means that triggering and/or waiving criteria were needed for these additional endpoints. (ECETOC, 2008).

[5] The Janer study evaluated 176 multi-generation studies on 148 substances to assess the potential differences between the first and the second generation. The study concluded that the F2 did not affect the overall NOAEL. No critical effect was observed in de second generation and the F2 had no impact on ensuing risk assessment nor on classification and labelling (Janer et al., 2007a). In 2009 the US EPA provided the OECD EOGRTS expert panel with a report on a retrospective analysis of 350 multi-generation reproductive toxicity rat studies mostly conducted for pesticides (Martin et al., 2009), concluding that . . .*"There is a great deal of redundancy in the second generation tests*." (Gilbert, 2011). The OESO expert meeting concluded in 2008 that all available two generation studies should be combined into one database to be able to fully judge the value of the second generation. This analysis was executed in the Netherlands based on a US database structure (US EPA ToxRefDB). In 2011 Piersma et al. published a retrospective analysis of 498 multi-generation reproductive toxicity study reports performed since the 1980s to review the added value of the second generation (Piersma et al., 2011). The manuscript was officially published in 2011. However, the authors have made their intermediate results continuously available to the delegates in the OECD expert group, through a secured website, monthly teleconferences and two conferences at the Dutch airport Schiphol. The retrospective analysis was finalized in October 2010 and got accepted for publication in November 2010, just a week before the OECD Joint Meeting accepted the new test guideline. Additionally, in 2011 a study of Rorije et al. became available in which the impact of the 2nd gen on classification & labelling was studied in all 50 reproductive toxicants classified at the time (Rorije et al., 2011).

[6] A guidance document serves to specify the guideline and is far easier to adapt compared to a TG.

2 weeks in the EOGRTS.[7] This approach would, according to the proponents of the EOGRTS, have sufficed in all known cases to affect fertility, but this was heavily debated by several stakeholders.

A fourth point of discussion, initiated in 2009 by a coalition of animal welfare organisations and industry, concerned the necessity of including immunotoxic parameters into the TG. The concern of these stakeholders was that the inclusion of immunotoxic parameters would lead to an unjustified increase of animal use and additional costs. The discussion was part of a broader debate between OECD experts on how to incorporate the extra cohorts DNT and DIT in the guideline. In 2008 the DNT and DIT cohorts were still optional but at the second Joint Meeting in 2009 Canada and the US insisted to make both cohorts obligatory. According to several respondents this resulted in the demand of other OECD member states to make additional – optional-parameters (e.g. clinical chemistry, haematology, pathology and necropsy in extra organs) obligatory. By doing so the discretionary space, that was originally offered by the Cooper protocol, gradually disappeared. The decision to require these parameters was, according to several respondents, not based on scientific arguments. It was part of the negotiation process that was needed to get the guideline adopted.

These discussion points illustrate that the process of FI was a very delicate process that took a lot of sensitivity and diplomacy. For the sake of the process numerous concessions had to be made. One of the respondents described the process as both encouraging and frustrating: "*On the one hand this guideline was established in a short time span and had an unprecedented pack of data to back the proposed change. On the other hand a small minority of OECD members remained withholding and several political decisions were made that were superfluous from a scientific perspective*."

2.2. The Actual Regulatory Acceptance (ARA) of the EOGRTS within the context of REACH

Ever since the establishment of TG 443 there has been discussion about its implementation in the context of REACH. The regulatory acceptance of the EOGRTS for risk assessment purposes proved to be highly challenging. One of the respondents remarked; "*writing the guideline was quite easy; it is the implementation which is the difficult part*". This is the result of a combination of factors that are described below.

First of all, OECD test guidelines are non-binding guidance documents leaving the implementation of the test up to the relevant competent authority. Furthermore, TG 443 was not defined as a direct replacement for the two-generation test and is flexible in the way it can be operationalized. This offers regulatory authorities and registrants the possibility to adapt the protocol to the needs of a specific situation, but at the same time leads to an ongoing discussion regarding the use of the EOGRTS within the context of REACH.

In August 2011 an Expert group-EOGRTS EG – was established within CARACAL (the Competent Authorities for REACH and CLP) on how to operationalize TG 443 within REACH. The preliminary conclusion of this expert group was to use the EOGRTS as the preferred method under REACH. This conclusion was supported by CARACAL. Subsequently the European Commission (EC) was urged to initiate the inclusion of OECD TG 443 in the Test Method Regulation (TMR) and to modify the relevant REACH Annexes accordingly. However, the legal, procedural and financial analysis on the application of TG 443 was not covered by the mandate of the EOGRTS EG and needed to be resolved by the Commission. During the November meeting of 2011 of the Member States Committee of ECHA (MSC), the secretariat of the European Chemicals Agency (ECHA) gave a presentation on the legal considerations and procedural aspects of the EOGRTS. The MSC was informed that in accordance with Article 13(3) of REACH Regulation, ECHA can in principle recognize OECD TG 443 as an appropriate study guideline, but to meet the requirements in Annex IX/X, 8.7.3 of REACH there are legal considerations that have to be taken into account. This means that as long as the EOGRTS is not part of the REACH TMR its legal position remains under discussion.

During the ECHA MSC meeting of September 2011 it became apparent that the MSC for the first time could not come to a unanimous agreement on a draft testing proposal for reproductive toxicity for scientific and technical reasons. In this case a few members of the MSC were in favor of the EOGRTS, while others preferred to leave the choice for the EOGRTS or the two-generation study to the registrant. A third group supported the latter argumentation but wanted to include the second generation in the EOGRTS. To deal with this disagreement, the procedure foreseen in Article 51(7) of the REACH Regulation was used for the first time. This procedure offers the possibility to split the draft decision on the registration of chemicals in those cases where the MSC fails to reach unanimous agreement. The part on which no agreement is reached, is sent to the EC to take the final decision. ECHA received around 230 testing proposals for reproductive toxicity (Annex IX/X, 8.7.3.) for phase-in substances registered by the December 2010 deadline. These should have been examined and decided upon by December 2012. However, due to the fact that in 2012 the MSC was not able to find unanimous agreement on the use of the EOGRTS and the need for a second generation, no decisions have been taken on these dossiers. Instead the parts of the dossier containing the reproductive toxicity endpoint were sent to the EC for decision making in the REACH Committee.

In November 2012 the Commission outlined their proposed approach to introduce the EOGRTS into the REACH regulatory framework. The approach involved inclusion of TG 443 in the Test Method Regulation (TMR) via the 5th Adaptation to Technical Progress (ATP) to the CLP Regulation, meaning "*. . .a modification of the REACH Annexes IX and X to include a "core" EOGRTS as the standard information requirement under point 8.7.3, and a 5 year review phase*".[8] (also see at the end of Section 2.2)

The discussion on the acceptance of the EOGRTS within Europe is dominated by two lines of argumentation i.e. the line of precaution versus the line of innovation.

2.2.1. The precaution frame

The precaution frame is the result of a strict interpretation of the Precautionary Principle (as laid down in EU Commission Communication 2000)[9] and is driven by a perceived risk of chemicals as being involuntary and dreadful. This category of risks requires a high level of protection from the government. In response regulators in the area of chemicals are stimulated to be very conservative in the decisions they take (Schiffelers et al., 2012). Any change in the existing way of testing has an additional risk to it, often leading to a preference for "the devil we know" (Storer et al., 2010). In this case the risk avoidance is increased by the complexity of the reproductive cycle and the fear of calamities such as the thalidomide disaster in the 1960s. The precautionary line of argumentation is in this case mainly represented by ECHA, Sweden, Finland and to a lesser extent France. The main objection of these stakeholders when it comes to the EOGRTS is connected to skipping the second generation and

[7] The 10 weeks originate from the duration of one complete spermatogenic cycle. However the EOGRTS prescribes 2 weeks premating exposure, plus extended exposure of males after mating up to a total of 10 weeks exposure followed by testis histopathology to check for testis effects.

[8] http://ec.europa.eu/enterprise/sectors/chemicals/files/caracal/minutes-121128-29_en.pdf; consulted February 2014.

[9] http://ec.europa.eu/dgs/health_consumer/library/pub/pub07_en.pdf; consulted February 2014.

thereby running the risk of missing out relevant information. According to an officer from the European Commission these stakeholders "*were very firm that the EOGRTS is not offering the information sought*" to safeguard the high level of protection on Human Health and Environment that REACH aims at in article 1.1 of REACH (EU, 2006). This was affirmed by ECHA in a statement to Nature: "*The two-generation study is the only study that covers effects on reproduction after exposure during all life-stages*" (Gilbert, 2011). An additional concern of these stakeholders is that the EOGRTS was originally designed for pesticides and not for industrial chemicals, while the retrospective analyses (see Section 2.1) are believed to cover insufficient industrial chemicals to offer a clear frame of reference for industrial chemicals. ECHA c.s. therefore emphasized that the two-generation study is the formal requirement in the REACH TMR (EU, 2008a). This was formulated as follows in the meeting minutes of April 2011 of ECHA's MSC: "*The current legal requirement under REACH is not the EOGRTS but the two generation reproductive toxicity study. ECHA needs to ensure with its decisions on testing proposals that this requirement is covered and the information expected from a two-generation reproductive toxicity study is available and adequate for the purposes of risk assessment and classification and labelling. Therefore, ECHA currently can accept the EOGRTS as a testing proposal of a registrant only if it is modified/specifically designed to cover the key parameters of a two-generation reproductive toxicity study (EU Test Method B.35).*"[10] The US and Canada, having to deal with a comparable legal situation for agrochemicals in which the second generation is legally required, solved this through guidance document 117 (the waiving guidance) (OECD, 2011b). It sets out internal triggers for those cases where a second generation is required by the Canadian Pest Management Regulatory Agency and by the United States Environmental Protection Agency, Office of Chemical Safety and Pollution Prevention. If these triggers are not found, the second generation can be waived. This transition regulation will be used for the intermediate period until new regulation including the EOGRTS is at force. The European OECD member states did not support guidance 117, according to one of the respondents, "*out of fear by some European member states that the goal of animal reduction would not be achieved.*"

Additionally, the European regulation on classification, labelling and packaging of substances and mixtures requires that substances are classified as reprotoxic when proven that they interfere with the sexual function and fertility (EU, 2008b). Fertility is defined as the ability to mate and create offspring. Rorije et al., argue that fertility can also be examined through parameters such as the sperm/follicle quality of the first generation (Rorije et al., 2011), but according to ECHA examining fertility requires an F2.[11]

Another point of concern of the representatives of the precautionary frame relates to the interpretation of the additional parameters that are generated by the EOGRTS. The additional information is according to ECHA only valuable for risk assessment purposes if it can be interpreted in a useful way. However, for many of these parameters no frame of reference exists, making it highly challenging from a regulators perspective to interpret the information. In addition, TG 443 is a flexible protocol which can be tailor made. This is on the one hand an advantage but on the other hand it also offers additional uncertainties in terms of how to conduct the test and how to deal with the additional data obtained. Holding on to the existing way of testing is the safest option for the parties doubting the usability of the EOGRTS. This is intensified by the fact that two generation test (TG 416) is a one size fits all protocol with a long history of application and extensive experience in interpreting the results.

2.2.2. The innovation frame

The innovation frame is represented within Europe by a group of EU MS (NL, UK, DE, DK and AU) together with several animal welfare organisations and industries. These stakeholders are advocates of the EOGRTS for a combination of scientific and animal welfare reasons which have been elaborated on in the Sections 1 and 2.1.

The EOGRTS advocates point to the fact that REACH offers discretion to regulatory authorities and manufacturers to choose the testing method best suitable for the job. Moreover, REACH favors alternative methods to conventional *in vivo* testing and article 25 (1) states that: *In order to avoid unnecessary animal testing, testing on vertebrate animals for the purpose of this Regulation shall be undertaken only as a last resort* (EU, 2006). Some of the member states have therefore already rejected draft registration dossiers through the MSC that still required a second generation. REACH in addition states that: *the methods shall be regularly reviewed and improved with a view to reducing testing on vertebrate animals and the number of animals involved. The Commission, following consultation with relevant stakeholders, shall, as soon as possible, make a proposal, if appropriate, to amend the Commission Regulation on test methods…. so as to replace, reduce or refine animal testing*-art. 13.2 REACH – (EU, 2006).

It must be noted though that the group of advocates is not as homogenous as it might seem. Several MS for example are very much in favor of the EOGRTS procedure with the DIT/DNT cohorts (e.g. DE, DK and NL) while others like the UK are not. This additional disagreement further complicates the issue.

When it comes to the ARA of the EOGRTS, the representatives of the two lines of argumentation each have their own legal and scientific arguments which are brought forward repetitively. The trench between precaution and innovation lies at the centre of the existing controversy and has led to fundamental disagreement in which the various institutions and particularly their experts have taken firm positions on the issue. This resulted in rigid opinions on all aspects of the discussion and moving away from the existing entrenched positions, held the risk of lose face. All in all, the two lines of argumentation led to an impasse in the discussion in which "*They agreed to disagree*".

The ball is now in the European Commission's court. The way the EC deals with the impasse is by generating more information to reassure those stakeholders who do not feel at ease with skipping the second generation. This is done through an additional 5 year review on the value of the second generation for a selection of substances and through a period of parallel testing. "*Once enough EOGRTS information has been collected, it will be necessary to re-assess the test design.*"[12] There may not be a clear scientific base for the additional review phase but science alone was unable to build the bridge to escape from the existing deadlock. Gaining trust in and experience with the procedure is crucial at this stage.

Early 2014 the Commission decided that the basic EOGRTS, without an F2 study, is the preferred test method to achieve the standard information requirement under REACH. However, it recognizes the "*scientific uncertainty*" over the "*added value*" of the F2 generation and states that EOGRTS should include F2 tests for a "*certain number of substances for which significant exposure of relevant populations (consumers, professional users) occurs*".[13]

[10] echa.europa.eu/.../meet_minutes_msc_17_en.pdf; consulted February 2014.

[11] To attest their point of view ECHA presented several substances that were perceived to cause unique effects in the second generation and therefore would only be classified and labeled as reprotoxic after examination of the F2. However, the RAC discussed some of these examples and concluded that there was no need for the C&L of these substances. http://echa.europa.eu/meetings-of-the-rac;consulted February 2014.

[12] http://chemicalwatch.com/18058/caracal-discusses-extended-one-generation-study-in-reach; consulted June 2014.

[13] http://chemicalwatch.com/18058/caracal-discusses-extended-one-generation-study-in-reach; consulted June 2014.

With regards to the DNT and DIT cohorts the Commission recently proposed that the DNT/DIT tests should only be carried out in "certain cases", due to "technical, economic and practicality reasons."[14] The decision to leave out the DNT and DIT cohorts initially led to opposition from several stakeholders such as the Netherlands. It was seen as a missed opportunity in terms of protecting the population from potential harmful effect of chemicals.[15] The Commission has asked member states to suggest conditions (triggers) for inclusion of DNT/DIT cohorts. Once decided on, such conditions will be part of the amendment to the REACH annexes.[16]

2.3. The Use by Industry (UI) of the EOGRTS to comply with REACH

The adoption of the OECD Test Guideline 443 has given registrants the possibility to choose between the two-generation reproductive toxicity study (test method: EU TM B.35/OECD TG 416) and the extended one-generation reproductive toxicity study (OECD TG 443) to meet the REACH requirements in Annex IX or X 8.7.3. ECHA's current position is that for the EOGRTS to meet the REACH information requirements (EU TM B.35), it will need to include an extension of Cohort 1B to mate the F1 animals to produce the F2 generation, which are kept until weaning. Nonetheless, ECHA also states that the EOGRTS can, under certain conditions, be suitable for a higher-tier study on a registered substance to fulfil the current information requirement in Annexes IX and X 8.7.3. There may be cases where registrants have specific information on properties of a substance explaining that the F2 in the EOGRTS is not necessary. The underlying scientific arguments can be used in a weight of evidence approach according to Annex XI, 1.2. of REACH to legitimize the adaptation of the standard information requirement (EU, 2006). It is the responsibility of the registrant to present these arguments in their testing proposal. They can modify the registration dossier to include the test method they prefer to use for reproductive toxicity before they receive a draft decision from ECHA. These arguments will be deliberated on in the analysis of the testing proposal by ECHA and the successive decision making. In the next stage the registrants can comment on the draft decision for the testing proposal. In this case ECHA expects that registrants express their preference on the method they want to use, so that their preference can be considered during the decision making procedure.

Some registrants actually already have proposed TG 443 in their registration dossier to meet the REACH information requirements on reproductive toxicity. However due to the fact that the MSC could not come to an agreement, most of the test proposals from registrants for reprotoxic endpoints have been stacked up for an undefined period of time.

The Use by Industry of the EOGRTS in the context of REACH is influenced by the following factors:

First of all, the EOGRTS protocol was originally developed for pesticides/agrochemicals. Agrochemicals need to comply to more test parameters than industrial chemicals, since they are designed to disrupt biological processes in the target species they intend to control. To meet these requirements the Cooper protocol included several endpoints (e.g. DNT and DIT) to the endpoints of reproductive toxicity to make it an all-inclusive test protocol for agrochemical compounds. As a result the EOGRTS produces more- and more accurate-information when compared to the TG 416. This is on the one hand a scientific advantage but may at the same time be an economic drawback due to the fact that many new models are oversensitive and prone to false positives (Storer et al., 2010). Consequently, more compounds will not pass the safety criteria, which is a serious entrepreneurial risk in terms of the development/use of compounds. Consequently, a group of companies is not supportive of an EOGRTS which includes the DNT and DIT parameters, since it has a higher risk of hindering products from getting onto the market.

Secondly, the costs for the EOGRTS are estimated, according to the European Commission, to be 2.5 times higher than the traditional two generation study. This estimation is based on information provided by industry and CRO's who refer to the costs of EOGRTS including the 2nd generation and the DIT/DNT cohorts (Cehtra, 2012). A survey conducted by Cehtra in request of ECHA speaks about an increase in costs of 41% if the EOGRTS is conducted with the second generation (Cehtra, 2012). Despite the difference in estimations it is clear that the EOGRTS leads to a substantial increase of costs considering the fact that a TG 416 costs about 500,000 Euro.[17] It should be notified however that the most expensive part of the EOGRTS testing concerns the performance of DIT and DNT cohorts.[18, 19] And the recent proposal of the EC excludes the DNT and DIT cohorts from the standard EOGRTS protocol.[20] For agrochemicals the EOGRTS normally is cost effective since both TG 416 and the Developmental Neurotoxicity study (TG 426) are often required for these substances. For industrial chemicals, however, it will results in increased test costs if all the additional parameters are required (Cehtra, 2012).

Thirdly, the EOGRTS is quite a complex and labour intensive testing procedure. Several labs have run trial studies and the results show that the procedure is feasible but complex (Schneider et al., 2010; Fegert et al., 2012). Moreover, the laboratory capacity to undertake the EOGRTS is still limited (Cehtra, 2012).

According to one of the respondents these practical/economical drawbacks were insufficiently taken into account during the stage of FI where "*for the sake of a swift adoption of the protocol, parameters were stacked on each other*".

From the side of agrochemicals the situation is more straightforward. The additional parameters targeted by the EOGRTS are required anyhow and the costs are comparable to or even lower than the required combination of TG 416 and TG 426. Several agrochemical companies have been involved in setting up the Cooper test protocol from the start and have thereby gained abundant experience with the EOGRTS. This combined with the higher standard of data and the use of substantially less animals has led to the fact that several agrochemical companies have already taken the step of executing the EOGRTS to meet with US/EU test requirements. The US EPA and Health Canada in the meanwhile have already accepted two studies on pesticides submitted by Dow conducted with the EOGRTS. Despite this progress the situation for industrial chemicals within Europe under REACH remains uncertain as has been described in Section 2.2. And as long as there is no clarity on how TG 443 needs to be conducted to meet the REACH information requirements, registrants will be cautious in using the EOGRTS protocol for the registration and authorization of their compounds, especially in those cases where higher costs are involved and the risk of false positives increases.

[14] http://chemicalwatch.com/20219/eu-commission-notifies-wto-of-reach-amendment-for-eogrts; consulted June 2014.

[15] http://chemicalwatch.com/20219/eu-commission-notifies-wto-of-reach-amendment-for-eogrts; consulted June 2014.

[16] http://chemicalwatch.com/18534/eu-test-method-regulation-update-disappoints-animal-groups; consulted June 2014.

[17] http://www.oecd.org/chemicalsafety/testing/46436593.pdf; page 487; consulted June 2014.

[18] http://echa.europa.eu/documents/10162/13578/meet_minutes_msc_23_en.pdf; consulted June 2014.

[19] Especially the DNT is very expensive due to the embedding and histology of tissues and behavioral assessments.

[20] http://chemicalwatch.com/20219/eu-commission-notifies-wto-of-reach-amendment-for-eogrts; consulted June 2014.

3. Analyses

Section 2 reveals that the sub stage of ARA has been the most challenging part in the process of regulatory acceptance and use of the EOGRTS. The controversy between the MS has become deeply entrenched. To comprehend this situation it is important to disclose the interrelatedness of the drivers and barriers described above. For this purpose two analytic steps are taken.

First of all, the influencing factors on regulatory acceptance and use are placed in the multilevel perspective on technology transitions which covers three main levels of influence on technology transitions (see text box for an explanation of this perspective).

Secondly the connection between the sub-stages of FI, ARA, and UI is analyzed to better understand the recent impasse.

3.1. The drivers and barriers from the multilevel perspective on technology transitions

THE MULTILEVEL PERSPECTIVE ON TECHNOLOGY TRANSITIONS

Regulatory acceptance and use of 3R models, such as the EOGRTS, is influenced by a combination of scientific, political, institutional, economic, ethical and social factors (Schiffelers et al., 2012). Such system innovations are hardly ever the effect of a single cause but the ''result of the interplay between many factors and actors''. Therefore an integrative approach is needed to understand such processes (Geels, 2006; Schiffelers et al., 2012). The multilevel perspective on technology transitions, developed by Schot and Rip (1996) offers such an integrative approach.

This perspective addresses three levels of influences which play a role in technology transitions (Kemp, 2010; Schiffelers et al., 2012):

- The micro- or niche level where innovations are developed and validated;
- The meso- or sociotechnical regime level covering the existing rules and regulations, expertise, dominant practices and the standing institutions;
- And the macro- or sociotechnical landscape level covering the material infrastructure, existing political culture and coalitions, social values, the macro-economy, demography and the natural environment.

For a new technology to become accepted and used, developments at these three levels need to reinforce each other before a shift in favor of the new technology can occur. In other words, an alignment or conjunction of the three levels can only take place if an innovation (e.g. a 3R model) fulfils the needs of the meso- and the macro level (Schiffelers et al., 2012). As long as the innovation does not fulfil these needs it will be unable to escape from the niche where it was developed (Kemp, 2010) as has been the case for several 3R models. Furthermore, the distinction between the levels can assist in defining the factors that are more suitable to start working on in order to improve the acceptance of the 3R model, since factors at the micro- and partly the meso level usually offer more possibilities for change than the broader societal developments at the macro level.

As long as an innovation's value is disputed, it will face difficulties in getting accepted at the meso level to become part of the existing regulatory regime (see text box on multilevel perspective of technology transitions). Fig. 1 illustrates this dispute with regards to the EOGRTS by displaying the contradicting pressures as described in Section 2, using the multilevel perspective on technology transitions.

The drivers at the micro level proved strong enough for the EOGRTS to become accepted at the OECD level. Yet the actual regulatory acceptance needs to be effectuated at the European level and at the level of individual countries such as the US and Canada. In the context of REACH, the drivers at the micro level were not convincing enough for all parties. The underlying cause is the difference in the frames of reflection embraced by the two opposing groups. The advocates of the EOGRTS focus on the advantages of the innovation while the advocates of TG 416 focus on the uncertainties connected to the switch to new way of testing. And since consensus is needed within ECHA's MSC the EOGRTS repeatedly bounces back to the micro level where it's suitability is disputed and put to the test. This for example is illustrated by the decision of the Commission to conduct an additional review phase performing the 2nd generation for a limited number of substances.

The colliding factors at the micro level have led to a policy controversy. "*Such disputes are resistant to resolution by appeal to facts or reasoned argumentation because the parties' conflicting frames – i.e. the innovation versus the precautionary frame – determine what counts as a fact and what arguments are taken to be relevant and compelling*" (Schön and Rein, 1994). While in policy disagreements, "*the parties to contention are able to resolve the questions at the heart of their disputes by examining the facts of the situation*" the parties are no longer able to do so when it comes to policy controversies (Schön and Rein, 1994).

The formal discussions are primarily of a scientific and/or legal nature. However, the scientific and legal arguments used by both parties tend to disguise the underlying interests, uncertainties, sympathies and antipathies. From the interviews it became clear that the controversy is fed by dissimilar interests and values of the stakeholders involved. Moreover, the different individual experts represent unalike levels of influence, expertise and thereby trust in the EOGRTS and different levels of responsibility in case anything goes wrong. Off the record, respondents were sometimes willing to refer to aspects such as perceived lack of expertise and experience, reputational issues, the fear of lose face and clashing personalities. Such psychological and institutional aspects strongly influence the existing impasse but are normally only addressed in a private manner.

3.2. The connectedness of the sub stages FI, ARA and UI

Although the sub stages of FI, ARA and UI have their own specific drivers and barriers, the stages are strongly connected and barriers and drivers at one stage also influence the other stages. If we want to understand why consensus was hard to reach at the stage of ARA it is essential to take the drivers and barriers at the stage of FI into account. Several respondents specified that the swiftness of the process of FI has slowed down progress at the stage of ARA. The doubts that played a role at the stage of FI resurfaced at the stage of ARA. Several issues were unsolved at the stage of FI and were transferred to the stage of ARA to be decided upon. Some respondents suggested that parties that had doubts about the EOGRTS only accepted TG 443 because they knew that the actual acceptance needed to be effectuated at the European level where they would have chance to reopen the discussion. The discussion during the OECD process was predominantly a scientific one. The influence of other aspects such as feelings of discomfort and lack of experience with the proposed protocol and impracticalities of the model in terms of costs and operationalization, may have been partially underestimated. The heritage of the stage of FI is perceived to have had a big influence on the stage of ARA while the sub stage of ARA on its turn influences the sub stage of UI. As long as there is uncertainty about the ARA and the

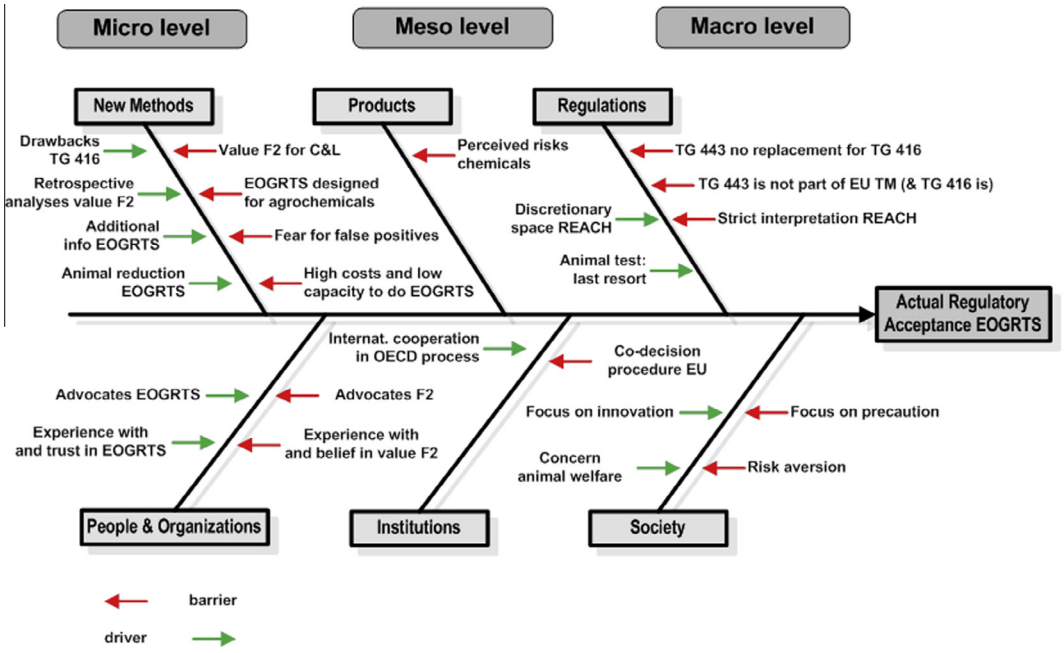

Fig. 1. Multilevel perspective on drivers and barriers influencing regulatory acceptance and use of the EOGRTS within the context of REACH (based on Schiffelers et al., 2012).

way the EOGRTS needs to be conducted, UI will be delayed. However frontrunners within industry that proposed the EOGRTS despite these uncertainties, have on their turn also influenced the ARA by pressurizing the system to come to a decision.

In other words the sub stages of regulatory acceptance and use are largely connected. For a 3R model to become accepted and used for regulatory purposes it is therefore very important that the involved stakeholders anticipate on the criteria/obstacles a 3R model might face in the other sub stages.

4. Discussion

Despite the fact that the OECD adopted TG 443 in July 2011: "*A quick implementation of the EOGRTS has been hampered by disagreement among experts on the value of information obtained from the 2nd generation (F2) and the extended debate whether the DNT/DIT cohorts should be included in the default study design in REACH and its relevance for the assessment when the F2 generation is not performed*,"[21]

The challenging implementation process within the European context is thought-provoking since directive 2010/63/EU states in article 13.2 that: "*...in choosing between procedures, those which use the minimum number of animals shall be selected*". In addition recital 11 of the directive states that "*Where no alternative method is recognized by the legislation of the Union, the numbers of animals used may be reduced by resorting to other methods and by implementing testing strategies, such as the use of in vitro and other methods that would reduce and refine the use of animals.*" (EU, 2010). Furthermore, ECHA's board of appeal declared in a recent decision to annul an animal study that; "*...Directive 2010/63/EU on the protection of animals used for scientific purposes cannot be treated as directly imposing any obligations on ECHA, but the latter's actions should not run counter to the principles – 3R – laid down therein....*".[22]

In this light the challenging acceptance of 3R models such as TG 443 begs the following final question; *'What is needed to augment the current process and what can be learned from this case study for future processes?'*

For 3R models such as the EOGRTS to really enter the area of risk assessment it is important that the conflicting parties reflect on the existing frames and the connected disputes. Reconsideration is needed whether these are really opposing or to a certain extent also appending arguments. Such a reflection on the existing frames (Schön and Rein, 1994) may well reveal that the parties share many mutual interests. All of the involved stakeholders for example adhere to the general principle that science should be at the basis of the decision making process and they all aim for risk minimization with regards to chemical substances. The precautionary frame and the innovation frame may therefore be less worlds apart as they at first might appear. New protocols such as the EOGRTS offer possibilities to simultaneously improve the level of innovation and of precaution. This however requires that all parties get the chance to build experience with the new protocol and investigate its – dis-advantages for a certain period. This way the seemingly conflicting frames can slowly merge into one shared frame. Reframing though requires a neutral policy entrepreneur (Bryson and Crosby, 1992) with sufficient mandate to bring the conflicting parties together to reconsider the present situation and formulate a strategy to blend the diverging interests. When it comes to the EOGRTS this role best fits the EC.

Furthermore, it requires shared *Commitment* of all parties involved to a joint policy goal, intense *Communication* between the parties on diverging and mutual interests – which took place during the OECD process but some interests were parked for the sake of speed of the process, enduring *Cooperation* between all stakeholders and a strong process *Coordination*, which was well arranged during the OECD process but was taken up quite late within the European process. These 4C's (Schiffelers et al., 2014) are pivotal to transform slow acceptance processes, in which non-decision making lurks, into a proactive process in which all parties work towards a clear defined policy goal.

[21] http://chemicalwatch.com/18058/caracal-discusses-extended-one-generation-study-in-reach; consulted June 2014.
[22] http://echa.europa.eu/documents/10162/13575/a_005_2011_boa_decision_en.pdf; consulted June 2014.

Buying some additional time through an additional 5 year review (see Section 2.2) might in our opinion in this case well be a wise move from the EC to let all parties get used to the new procedure and reframe the discussion. In this process it can become clear how and when the EOGRTS should be conducted. In the meantime the Commission can evaluate the existing positions and work towards a stepwise strategy to overcome the remaining controversies.

In short the following recommendations are made to enhance the process of regulatory acceptance and use of the EOGRTS and similar future processes:

- The EOGRTS case reveals that the combination of profuse scientific information and strong advocates are important ingredients for the regulatory acceptance of a 3R model. However, it also reveals that regulatory acceptance is often a highly politicized process in which science can become part of the existing disagreement and in which other arguments e.g. the lack of experience with and trust in the new model, institutional agenda's and political realities are seen to regularly outweigh the scientific 'facts'.
- These 'other' arguments therefore continuously have to be taken into consideration. This means that a permanent anticipation is needed on the legal, practical and psychological requirements for the 3R model to proceed to the following sub stages of ARA and UI (e.g. legal issues regarding the existing regulatory frame, costs connected to the new protocol, capacity needed to perform it, trust in the innovation and knowledge required to work with it). A regular meta-communication is important to discuss the remaining doubts and existing interests and to identify the criteria needed to merge the remaining diverging interests.
- All stakeholders need to be aware of the fact that the three sub stages of FI, ARA and UI are largely connected. The heritage of a previous stage is likely to remain of influence at the subsequent stage.
- The current implementation process needs reflection on the existing frames and on the mutual and conflicting aspects within these frames.
- To guide a 3R model through the sub stages of regulatory acceptance and use, a neutral process manager is needed to set out a clear strategy, steer the process, keep all parties involved and be sensitive to the different interests at stake. Clearly, there is a role for the EC in bringing together the views of the different European stakeholders on such issues.
- Commitment needs to be asked by the EC from all European stakeholders in light of Directive 2010/63/EU in terms of stimulating the use of 3R models.
- In short mutual trust, shared commitment, a strong process coordination and close cooperation and communication are needed to enhance the process of regulatory acceptance and use of 3R models in general and the EOGRTS in particular (Schiffelers et al., 2014).

Finally, when it comes to reproductive toxicity testing, the EOGRTS should not be considered as a stand-alone procedure. The EOGRTS is an important step in terms of animal welfare and scientific progress. However, since many chemicals need to be tested for reproductive toxicity the step from the two generation study to the EOGRTS will not solve the testing bottleneck for the many chemical substances in commerce. For this, a paradigm shift is needed from extensive animal testing to efficient and focussed animal and *in vitro* testing. This other way of thinking combines clever ways of chemical prioritization with intelligent testing strategies in which different reproductive toxicity testing methods are combined, as for example proposed by Schaafsma et al. (2009), Spielmann (2009) and Martin et al. (2011).[23] Such intelligent testing strategies, offer the best potential to minimize *in vivo* reproductive toxicity testing.

Conflict of interest

The authors declare that there are no conflicts of interest.

Acknowledgment

We thank the Doerenkamp-Zbinden Foundation and the Dutch Ministry of Economic Affairs for funding this project.

We thank the respondents for sharing their expertise and perspectives with us.

Appendix A

This appendix describes the research approach, the research methods, the criteria for the case selection and the process of respondent selection that are used for this manuscript.

A.1. Case study approach: causal process tracing

Regulatory acceptance and use is a process which is influenced by a broad variety of drivers and barriers (Schiffelers et al., 2012, 2014). With this manuscript the authors aim at creating a profound clarification of the underlying mechanism of regulatory – non-acceptance of available 3R models through the examination of the variables influencing this complex problem. This means creating an in depth picture to unravel causal mechanisms, by reconstructing events and situations that have unfolded over time. For this purpose the case study approach of causal process-tracing is used (George and Bennett, 2005; Blatter and Haverland, 2012). Through this qualitative analysis technique the intervening causal process between an independent variable (i.e. regulatory acceptance and use of 3R models) and various dependent variables (e.g. scientific information, level of risk aversion, concern about animal welfare, regulatory frame, etc.) is scrutinized.

A.2. Research methods

Case study research relies on multiple sources of evidence. Different research methods are combined in order to get to a comprehensive representation of the situation examined (Yin, 2003). The variables influencing the process of regulatory acceptance and use of the EOGRTS were identified through a combination of literature review and expert interviews. The literature research provided an overview of the regulatory framework, stakeholders involved, existing testing practices and variables influencing the regulatory acceptance and use. The examined sources consisted of scientific publications, meeting reports –e.g. minutes between 2010 and 2013 of the Member States Committee of ECHA (MSC), the Risk Assessment Committee of ECHA (RAC) and the Competent Authorities for REACH and GLP (CARACAL)-, websites of involved stakeholders –e.g. OECD, EC, ECHA-, press releases and correspondence between stakeholders. In addition, a series of 18 in depth interviews was conducted between 2012 and 2013 to collect the respondents' perspectives on the process of acceptance and use of the EOGRTS (see also Section A.4.). The interviews were semi-structured, asking open-ended questions designed to reconstruct the process and identify the drivers and barriers per subsequent sub stage, i.e. FI, ARA and UI. The interviews began with the

[23] Another approach could be the NIEHS/NTP approach for reproductive toxicity testing (see Dr. Paul Foster for references).

question of the involvement of the respondent in the process, a short chronology of this involvement and the position of his/her organization regarding the EOGRTS. Next, a series of questions were asked regarding the barriers and drivers per sub stage of the process. Lastly, interviewees were asked to give their views on optimizing the current process of regulatory acceptance and use of the EOGRTS within Europe. The main questions were the same for every respondent but the focus differed depending on the respondents involvement in the process.

Most interviews were audio recorded and subsequently transcribed. In those cases where interviews were not recorded ($N = 4$) the interviews were transcribed and made available to the respondent for validation of the findings. Next the transcripts were analyzed to make an inventory of drivers and barriers per stub stage and of the optimization possibilities.

A.3. Case selection

In order to fully depict the issue of regulatory acceptance and use, the case study had to meet the following criteria:

1. The existing regulatory test is an animal model which is under discussion;
2. There is a model available to reduce, replace or refine the existing animal model (3R model);
3. And this 3R model is in the process of becoming regulatory accepted/used.

The EOGRTS case meets all three criteria. The fact that the EOGRTS is already quite far in the process of becoming accepted/used for regulatory purposes, offers the possibility to depict the full process of different causes (independent variables) influencing the outcome of regulatory acceptance and use (dependent variable).

A.4. Respondent selection

The respondents selection was done through a combination of criterion and snowball sampling (Patton, 2001). Through criterion sampling[24] a small group of relevant respondents was selected beforehand using the selection criteria of being a scientific, legal and/or political expert with experience in/or knowledge of the EOGRTS case study and with former or current involvement in this case study. Involvement means having been able to closely follow or take part in – parts of – the process of acceptance and use of the EOGRTS. The first sample existed of a group of 5 experts involved in the process of regulatory acceptance and use of the EOGRTS. Next, the population was broadened through snowball sampling asking each respondent for other suitable candidates. Suitability was defined as direct or indirect involvement in one or more of the sub stages of FI, ARA or UI. We have explicitly looked for respondents with different perspectives on the case both in terms of stakeholder groups as in terms of opinion regarding the EOGRTS. The respondents came from the following stakeholder groups: European legislators (5) European and national regulatory authorities (4), industry/contract research organizations (5), academia (2) and animal welfare organizations (2).

References

Blatter, J., Haverland, M., 2012. Designing Case Studies: Explanatory Approaches in Small-N Research. Palgrave Macmillan, Houndmills Basingstoke.

Bryson, J.M., Crosby, B.C., 1992. Leadership for the Common Good. Jossey-Bass Publishers, San Francisco, CA.

Cehtra, June 2012. Report on Survey of Worldwide CROs: Costs and Practicalities of Two New OECD Guidelines for Testing Chemical Substances OECD 443, Extended One-Generation Reproductive Toxicity Study, and OECD 488, Transgenic Rodent Somatic and Germ Cell Mutation Assay.

Cooper, R.L., Lamb, J.C., Barlow, S.M., Bentley, K., Brady, A.M., Doerrer, N.G., et al., 2006. A tiered approach to life stages testing for agricultural chemical safety assessment. Crit. Rev. Toxicol. 36 (1), 69–98.

ECETOC, 2008. Workshop on Triggering and Waiving Criteria for the Extended One-generation Reproduction Toxicity Study: 14–15 April 2008, Barza d'Ispra. Workshop Report No. 12 pp. 33 Brussels, Belgium: European Centre for Ecotoxicology & Toxicology of Chemicals. Available at: <http://www.ecetoc.org/index.php?mact=MCSoap,cntnt01,details,0&cntnt01by_category=6&cntnt01template=display_list_v2&cntnt01order_by=Reference%20Desc&cntnt01display_template=display_details_v2&cntnt01document_id=321&cntnt01returnid=93/> (accessed April 2014).

EU, 2006. Regulation (EC) no 1907/2006 of the European parliament and of the council of 18 December 2006 concerning the registration, evaluation, authorization and restriction of chemicals (REACH), establishing a European chemicals agency. OJ Eur. Union L396, 1–849. Available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:396:0001:0849:EN:PDF/> (accessed April 2014).

EU, 2008a. Council Regulation (EC) No 440/2008 of 30 May 2008 Laying Down Test Methods Pursuant to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). Available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:142:0001:0001:EN:PDF/> (accessed April 2014).

EU, 2008b. Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 december 2008 on Classification, Labelling and Packaging of Substances and Mixtures, Amending and Repealing Directives 67/548/EEC and 1999/45/EC, and Amending Regulation (EC) No 1907/2006 Available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:353:0001:1355:en:PDF/> (accessed April 2014).

EU, 2010. Directive of the European parliament and of the council of 22 September 2010 on the protection of animals used for scientific purposes. OJ Eur. Union L276, 33–79. Available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:276:0033:0079:en:PDF/> (accessed April 2014).

Fegert, I., Billington, R., Botham, P., Carney, E., FitzGerald, R.E., Hanley, T., Lewis, R., Marty, M.S., Schneider, S., Sheets, L.P., Stahl, B., van Ravenzwaay, B., 2012. Feasibility of the extended one-generation reproductive toxicity study (OECD 443). Reprod. Toxicol. 34 (3), 331–339.

Geels, F.W., 2006. Multi-level perspective on system innovation relevance for industrial transformation. In: Olshoorn, X., Wieczorek, A.J. (Eds.), Understanding Industrial Transformation Views from Different Disciplines. Springer, pp. 163–186.

George, A.L., Bennett, A.B., 2005. Case Studies and Theory Development in the Social Sciences. The Mitt Press, Cambridge, Massachusetts.

Gilbert, N., 2011. Data gaps threaten chemical safety law: European companies are not providing robust information to regulators or alternatives to animal experiments. Nature 475, 150–151. Available at: <http://www.nature.com/news/2011/110712/full/475150a.html/> (accessed April 2014).

Hartung, T., Rovida, C., 2009. Opinion: chemical regulators have overreached. Nature 460, 1080–1081.

Janer, G., Hakkert, B.C., Slob, W., Vermeire, T., Piersma, A.H., 2007a. A retrospective analysis of the two-generation study: what is the added value of the second generation? Reprod. Toxicol. 24, 97–102.

Janer, G., Hakkert, B.C., Piersma, A.H., Vermeire, T., Slob, W., 2007b. A retrospective analysis of the added value of the rat two-generation reproductive toxicity study versus the rat subchronic toxicity study. Reprod. Toxicol. 24, 103–113.

Kemp, R., 2010. The Dutch energy transition approach. Int. Econ. Econ. Policy 7, 291–316.

Lilienblum, W., Dekant, W., Foth, H., Gebel, T., Hengstler, J.G., Kahl, R., Kramer, P.J., Schweinfurth, H., Wollin, K.M., 2008. Alternative methods to safety studies in experimental animals: role in the risk assessment of chemicals under the new European chemicals legislation (REACH). Reprod. Toxicol. 82, 211–236.

Martin, M.T., Mendez, E., Corum, D.G., Judson, R.S., Kavlock, R.J., Rotroff, D.M., et al., 2009. Profiling the reproductive toxicity if chemicals from multi-generation studies in the toxicity reference database. Toxicol. Sci. 110 (1), 181–190.

Martin, M.T., Knudsen, T.B., Reif, D.M., Houck, K.A., Judson, R.S., Kavlock, R.J., Dix, D.J., 2011. Predictive model of rat reproductive toxicity from ToxCast high throughput screening. Biol. Reprod. 85, 327–339.

OECD, 2001. TG 416. OECD Guideline for Testing of Chemicals. Two-Generation Reproduction Toxicity Study. Organisation for Economic Co-operation and Development, Paris, France.

OECD, 2011a. TG 443. OECD Guideline for Testing of Chemicals. Extended One-Generation Reproductive Toxicity Study. Organisation for Economic Co-operation and Development, Paris, France.

OECD, 2011b. Guidance Document 117 on the Current Implementation of Internal Triggers in Test Guideline 443 for an Extended one Generation Reproductive Toxicity Study, in the United States and Canada. Organisation for Economic Co-operation and Development, Paris, France.

Patton, M.Q., 2001. Qualitative Research and Evaluation Methods, 2nd ed. Sage Publications, Thousand Oaks, CA.

Pedersen, F., De Bruijn, J., Munn, S., Van Leeuwen, K., 2003. Assessment of Additional Testing Needs Under REACH. Effects of (Q)SARS, Risk Based Testing and Voluntary Industry Initiatives. European Commission, Available at: <http://

[24] Criterion sampling involves selecting respondents through some predetermined criteria of importance (Patton, 2001).

home.kpn.nl/REACH/downloads/reachtestingneedsfinal.pdf/> (accessed April 2014).

Piersma, A.H., Rorije, E., Beekhuijzen, M.E.W., Cooper, R., Dix, D.J., Heinrich Hirsch, B., Martin, M.T., Mendez, E., Muller, A., Paparella, M., Ramsingh, D., Reaves, E., Ridgway, P., Schenk, E., Stachiw, L., Ulbrich, B., Hakkert, B.C., 2011. Combined retrospective analyses of 498 rat multi-generation reproductive toxicity studies: on the impact of parameters related to F1 mating and F2 offspring. Reprod. Toxicol. 31, 392–401.

Rorije, E., Muller, A., Beekhuijzen, M.E.W., Hass, U., Heinrich-Hirsch, B., Paparella, M., Schenk, E., Ulbrich, B., Hakkert, B.C., Piersma, A.H., 2011. On the impact of second generation mating and offspring in multi-generation reproductive toxicity studies on classification and labelling of substances in Europe. Regul. Toxicol. Pharmacol. 61 (2), 251–260.

Ruden, C., Hansson, S.O. 2008. Letter to the editor. Comment on: "A retrospective analysis of the two-generation study: what is the added value of the second generation?" by G. Janer, B.C. Hakkert, W. Slob, T. Vermeire, A.H. Piersma (Reprod. Toxicol. 2007. 24: 97-102.) Reprod. Toxicol. 25, 397–405.

Russel, W., Burch, R., 1959. The Principles of Humane Experimental Technique. Methuen, London.

Schaafsma, G., Kroese, E.D., Tielemans, E.L.J.P., Van de Sandt, J.J.M., Van Leeuwen, C.J., 2009. REACH, non-testing approaches and the urgent need for a change in mind set. Regul. Toxicol. Pharmacol. 53, 70–80.

Schiffelers, M.J.W.A., Blaauboer, B.J., Hendriksen, C.F.M., Bakker, W.E., 2012. Regulatory acceptance and use of 3R models: a multilevel perspective. ALTEX 29 (3), 287–300. Available at: <http://www.altex.ch/All-issues/Issue.50.html?iid=133&aid=4/> (accessed June 2014).

Schiffelers, M.J.W.A., Blaauboer, B.J., Bakker, W.E., Beken, S., Hendriksen, C.F.M., Koëter, H.B.W.M., Krul, C., 2014. Regulatory acceptance and use of 3R models for pharmaceuticals and chemicals: expert opinions on the state of affairs and the way forward. Regul. Toxicol. Pharmacol. 69, 41–48.

Schneider, S., Kaufmann, W., Strauss, V., van Ravenzwaay, B., 2010. Vincloszolin: a feasibility and sensitivity study of the ILSI–HESI F1-extended one-generation rat reproduction protocol. Regul. Toxicol. Pharmacol. 59, 91–100.

Schön, D., Rein, M. 1994. Frame Reflection Toward the Resolution of Intractable Policy Controversies. New York.

Schot, J., Rip, A., 1996. The past and future of constructive technology assessment. Technol. Forecast. Soc. Chang. 54, 251–268.

Spielmann, H., 2009. The way forward in reproductive/developmental toxicity testing. ATLA 37, 1–16.

Spielmann, H., Vogel, R., 2006. Editorial: REACH testing requirements must not be driven by reproductive toxicity testing in animals. ATLA 34, 365–366.

Storer, R.D., Sistare, F.D., Reddy, M.V., DeGeorge, J.J., 2010. Toxicological pathology an industry perspective on the utility of short-term carcinogenicity testing in transgenic mice in pharmaceutical development. Toxicol. Pathol. 38, 51–61. Available at: <http://intl-tpx.sagepub.com/content/38/1/51.full.pdf+html/> (accessed April 2014).

Van der Jagt, K., Munn, S., Tørsløv, J., de Bruijn, J., 2004. Alternative Approaches can Reduce the Use of Animals Under REACH. Available at: <http://publications.jrc.ec.europa.eu/repository/bitstream/111111111/8790/1/EUR%2021405%20EN.pdf/> (accessed April 2014)

Yin, R.K., 2003. Case Study Research: Design and Methods. Applied Social Research Methods Series, vol. 5. Sage Publications, Thousand Oaks Inc.