Regulatory Toxicology and Pharmacology 66 (2013) 241–247

Contents lists available at SciVerse ScienceDirect

Regulatory Toxicology and Pharmacology

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

A systematic approach for evaluating and scoring human data

Chris D. Money [a,*], John A. Tomenson [b], Michael G. Penman [c], Peter J. Boogaard [d], R. Jeffrey Lewis [e]

[a] ExxonMobil Petroleum Chemical, Hermeslaan 2, B-1831 Diegem, Belgium
[b] Causation Ltd, 2 Field View Drive, Macclesfield, Cheshire SK11 7JN, United Kingdom
[c] Penman Consulting BVBA, Rue Royale 157, Bte 13, B-1210 Brussels, Belgium
[d] Shell International B.V., P.O. Box 162, 2501 AN The Hague, Netherlands
[e] ExxonMobil Biomedical Sciences Inc., 1545 US Highway 22 East, Annandale, NJ, USA

ARTICLE INFO

Article history:
Received 9 October 2012
Available online 8 April 2013

Keywords:
Human data
REACH
IUCLID
Data quality
Hazard assessment
Data reliability

ABSTRACT

An approach is described for how the quality of human data can be systematically assessed and categorised. The approach mirrors the animal data quality considerations set out by Klimisch et al., in order that human data quality can be addressed in a complementary manner and to help facilitate transparent (and repeatable) weight of evidence comparisons. Definitions are proposed for the quality and adequacy of data. Quality is differentiated into four categories. A description of how the scheme can be used for evaluating data reliability, especially for use when contributing entries to the IUCLID database, is shown. A discussion of how the criteria might also be used when determining overall data relevance is included. The approach is intended to help harmonise human data evaluation processes worldwide.

1. Introduction

The basis of the human health hazard assessment which must be undertaken for a registered substance under the EU REACH Regulation (EU, 2007) is an evaluation of all available animal and human data. To further clarify these expectations, Technical Guidance has been developed (ECHA, 2011a,b) that describes the process which must be adopted by registrants i.e. relevant data (including those that may be available from analogous substances and from suitable structure activity models) are compiled for each endpoint, evaluated with regard to their quality; and then used to determine the nature of the hazard for the endpoint. The Guidance recommends the use of the criteria proposed by Klimisch et al. (1997) as the basis for the quality determination of available animal studies.

The 'Klimisch criteria' were originally developed as a response to the obligations placed upon industry by the precursor to REACH, the EU Existing Substances Regulation (EC, 1994). Specifically, as part of the reporting requirements for that Regulation, relevant hazard data for the substance was expected to be entered into its IUCLID record, together with an assessment of the relevance and reliability of the data (European Commission, 2003). Klimisch et al. describe a series of considerations that can be applied to both animal and environmental experimental studies in order that they can be allocated into one of four categories (Table 1). For example, a high quality study might be identified as one which was carried out according to the OECD Guidelines for a particular endpoint. This approach to quality categorisation has enabled substance IUCLID records to be structured in a manner that allows primary data, together with supporting and other studies, to be contained and displayed in the record in a manner that is scientifically valid, repeatable and is consistent across substances.

With the exception of data obtained from human volunteer studies, human data (HD), particularly those derived from epidemiological studies, do not readily fit into the scheme proposed by Klimisch et al. Most obviously, HD are typically observational rather than experimental in nature and epidemiological studies are invariably opportunistic and are limited by the exposures that occur in the general population or occupationally. As a consequence, they are far less likely to give repeatable results. There will also be other risk factors that cannot be controlled which need to be identified, measured and adjusted for in the interpretation of HD findings. In particular, it is much more difficult to assess what exposure has occurred. Although occupational exposures can often more reliably be assessed than exposures in the general population, most exposure assessments will fall far short of the animal experiment where the daily intake and route of intake is known. The notable exception may be biological monitoring, provided sufficient toxicokinetics data are available (Boogaard and Money, 2008). In short, human epidemiological data are seldom targeted to answer specific questions 'per endpoint' and the protocols under

* Corresponding author. Fax: +32 2 722 3710.
E-mail addresses: chris.money@exxonmobil.com (C.D. Money), john_tomenson@causation.co.uk (J.A. Tomenson), Mie@PenmanConsulting.com (M.G. Penman), peter.boogaard@shell.com (P.J. Boogaard), r.jeffrey. lewis@exxonmobil.com (R. Jeffrey Lewis).

0273-2300/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.yrtph.2013.03.011

Table 1
Description of Klimisch categories for animal studies.

Code	Category	Category definitions
1	Reliable without restriction	This includes studies or data from the literature or reports which were carried out or generated according to generally valid and/or internationally accepted testing guidelines (preferably performed according to GLP) or in which the test parameters documented are based on a specific (national) testing guideline (preferably performed according to GLP) or in which all parameters described are closely related/comparable to a guideline method
2	Reliable with restriction	This includes studies or data from the literature, reports (mostly not performed according to GLP), in which the test parameters documented do not totally comply with the specific testing guideline, but are sufficient to accept the data or in which investigations are described which cannot be subsumed under a testing guideline, but which are nevertheless well documented and scientifically acceptable
3	Not reliable	This includes studies or data from the literature/reports in which there are interferences between the measuring system and the test substance or in which organisms/test systems were used which are not relevant in relation to the exposure (e.g., unphysiologic pathways of application) or which were carried out or generated according to a method which is not acceptable, the documentation of which is not sufficient for an assessment and which is not convincing for an expert judgment
4	Not assignable	This includes studies or data from the literature, which do not give sufficient experimental details and which are only listed in short abstracts or secondary literature (books, reviews, etc.)

which such studies are executed are not generally formalised in the manner encountered with animal studies. Yet the REACH Technical Guidance (ECHA, 2011a) not only states that "all available data must be compiled and evaluated per endpoint" but that "where available, human data are preferred".

These demands present a challenge; if human studies are to be transparently and equitably incorporated into REACH IUCLID records, then how should available information be evaluated and categorised in order that this can be treated in a manner consistent with that applied to animal studies? Moreover, how might any human data evaluation scheme be structured such that it can apply across the range of human study types (e.g. case control studies; cohort studies; human volunteer studies; case reports; molecular epidemiology investigations; etc.) and also be applied in the 'weight of evidence' comparisons encouraged by the Guidance and also being advocated elsewhere (Meek et al., 2003; Boobis et al., 2006, 2008; NRC, 2011; SCENIHR, 2012)?

2. Considerations behind a systematic approach

The question of identifying how available human data should be best accounted for within the process of regulatory risk assessment is not new. The fact that incorporating the human experience into risk assessments serves to strengthen the integrity of the findings has been regularly highlighted as an issue and a need (Meijers et al., 1992; Dourson et al., 2001; IPCS, 2005). However, data that describe the human experience are also often viewed as being inherently compromised. For instance, epidemiological data and data from case reports (for example, those reported by poisons centres following consumer over-exposures or those recorded at occupational health facilities following industrial incidents) usually relate to uncontrolled exposure conditions and are therefore prone, amongst other uncertainties, to inherent exposure misclassification. Paradoxically, the very fact that human data are reflective of real world experiences is both their weakness and strength.

Although codes exist for how epidemiology should be conducted (IEA, 2007), no 'standard protocols' are available that might be applied when assessing the integrity of study design and the interpretation of findings. This contrasts with animal findings where protocols according to the OECD guidelines provide such a reference point. But, as previously noted, the human experience extends beyond formalised epidemiological investigations. Therefore, if it is to be systematically incorporated into the process of risk assessment, then further guidance would appear to be required (ECETOC, 2004; Money, 2007). The 'guidance gap' was intended to be accounted for during the process, carried out in 2006–8, that was employed to develop the Technical Guidance for REACH, but even though the resulting Guidance is very comprehensive, there are still areas where the advice is either insufficient or absent (Gade et al., 2008; Kreider and Williams, 2010). For example, the fact that little mention had been made of how human data should inform the choice of Assessment Factors that are applied in the process for the development of DNEL/DMELs was the topic of an ECETOC workshop in 2007 (ECETOC, 2009). The workshop resulted in the creation by the European Chemicals Bureau (which subsequently transferred its responsibilities to the European Chemicals Agency, ECHA) of a partner expert group (PEG) activity to address this deficiency. The recommendations of the PEG have now been published as a supplement to the original Technical Guidance (ECHA, 2010).

While various schemes have been suggested for how human epidemiological study information might be evaluated when it is applied during the risk assessment process (Hertz-Picciotto, 1995; ECETOC, 2002; Vlaanderen et al., 2008), within the context of routine application for chemicals regulation, these schemes generally suffer from three major drawbacks; first, their focus is too narrowly defined for them to be applied to the range of considerations expected of REACH (and which are expected to extend beyond risk assessment to include hazard/effect identification for a range of endpoints; identification of LOELs/NOELs; characterisation of null findings; characterisation of dose response relationships; and contributing to weight of evidence evaluations of the data). Second, the schemes are too technically sophisticated to be routinely applied by the range of industrial organisations affected by legislation such as REACH (and that will range from large corporations to small and medium sized companies). Lastly, the schemes predominantly focus on the potential contribution of 'classical' epidemiological studies, whereas relevant human data can be expected to range widely, including information derived from case report and volunteer studies.

3. Proposed categorisation scheme

REACH places emphasis on the application of an information collection and analysis strategy that helps ensure that hazard assessment is not only carried out using the best available data, but also avoids the unnecessary use of *in vivo* tests to acquire such information. These strategies embrace and encourage the use of grouping of similar substances into categories; reading across hazard information within and between categories in order to help fill 'data gaps'; the application of relevant (Q)SAR models and judgements, as well as the appropriate use of *in vitro* information.

Responding to the need to have available processes that enable data evaluations to be consistent across substances, ECETOC initiated a task force to examine how animal findings and human data