Priority setting strategies for regulatory testing of industrial chemicals

Anna Nordberg
STOCKHOLM 2007
This licentiate thesis consists of an introduction and two papers:

Paper I: Nordberg, A., Rudén, C., and Hansson, S.O. *Towards more efficient testing strategies - analyzing the efficiency of toxicity data requirements in relation to the criteria for classification and labelling*, submitted manuscript.


Paper II is reprinted with kind permission of Toxicology Letters, Elsevier.

Nordberg, A. *Priority setting strategies for regulatory testing of industrial chemicals*, Theses in Risk and Safety from the Division of Philosophy at the Royal Institute of Technology 1, Stockholm

ISSN 1654-627X


Printed by US AB, Stockholm, Sweden, 2007
Abstract
For the majority of the estimated 70,000 industrial chemical substances available on the European market today there is not enough information to enable a reasonably complete assessment of the risks that they might pose to man and the environment. Any strategy for the generation of additional data for these substances should aim at making testing as efficient as possible taking into account environmental and health protection, time, monetary cost and animal welfare. To achieve this, appropriate priority setting rules are needed.

The main criterion currently used for regulatory priority setting for testing of industrial chemicals is production volume; the higher the production volume, the more information is required. This was also the main criterion in the former legislation, preceding REACH (Registration, Evaluation and Authorisation of Chemicals). The aim of this thesis is to evaluate other priority setting criteria and their implications for risk management, in particular classification and labelling.

The first paper in this thesis includes a study of the efficiency ratio for some of the tests required for the notification of new substances, i.e. the ratio between the likelihood that the test will lead to a classification, and the monetary cost of performing the test. The efficiency ratio was determined for the standard tests for acute oral toxicity, irritation, sensitisation and subacute toxicity using data from 1409 new chemicals notified in Europe between 1994 and 2004. The results of this investigation suggest that, given limited resources for testing, it is more efficient to perform acute toxicity tests on a larger number of substances rather than to perform additional subacute toxicity studies on the substances already tested for acute toxicity.

The second paper included in this thesis, reports the results from a comparative study of the bioaccumulating properties of substances being (a) classified as carcinogenic, mutagenic and/or toxic to reproduction (CMR-substances), or (b) classified as acutely toxic or (c) unclassified. The purpose of this investigation was to evaluate potential consequences of prioritising bioaccumulating chemicals for evaluation and testing, as this is one of the strategies prescribed in REACH. The results of this study suggest that bioaccumulating substances are neither over- nor underrepresented among the CMR-substances. This result lends support to the use of the bioconcentration factor for priority setting.

The studies reported in this thesis utilize existing data on classification of substances as an indicator of the outcome of the risk assessment process, relating priority setting methods to the risk management measures that they give rise to. To the best of my knowledge there are still only very few studies published that address the issue of priority setting in chemicals control using this approach, and in my view there is need for more studies of priority setting methods and a further development of priority setting strategies that are science-based.

Keywords: risk assessment, risk management, regulatory toxicology, test strategies, prioritisation, REACH, industrial chemicals
Acknowledgement

First of all, I would like to thank my supervisor Christina Rudén for her support and encouragement, without which this licentiate thesis would not have come about. I would also like to thank my co-supervisor Sven Ove Hansson for help and support. Furthermore, I would like to thank Misse Wester Herber for support and discussions. I would also like to thank Katarina Victorin for careful reading and valuable comments. I am also grateful for the help I have received from the Swedish Chemicals Agency and the European Chemicals Bureau. I would also like to thank Betty Jurdell for the help with the first version of the database used in Paper II. Any and all faults in this thesis are of course my own.

Many thanks also to all my colleagues at the Division of Philosophy for stimulating discussions, delicious lunches and friendship. A special thanks to my roommate and friend Sara Svensson for all encouragement and support. Furthermore, I would like to give many thanks to my mother and father. My friends have also been of great support and for this I am grateful. A special thanks goes to Svante Eriksson for all help and support during my work with Paper II.

This work has been financially supported by The Animal Welfare Agency (Djurskyddsmyndigheten), StiFUD (Swedish Fund for Research Without Animal Experiments) and MISTRA (The Fund for Strategic Environmental Research). The support is gratefully acknowledged.

Stockholm, Anna Nordberg
October, 2007
## Contents

1. **Introduction** ................................................. 1

2. **Background** ................................................ 1
   2.1. Risk Assessment ........................................ 1
   2.2. Priority setting methods ............................ 2
   2.3. European Legislation regarding industrial chemicals .... 3
       2.3.1. Classification and labelling ....................... 3
       2.3.2. New and Existing Chemicals ....................... 4
       2.3.3. REACH – the new chemicals legislation .......... 6
   2.4. Animal welfare ........................................... 8

3. **Study of priority setting strategies** .................. 9
   3.1. Aim and methods of the thesis ....................... 9
   3.2. Preview of the papers ................................ 9
       3.2.1. Paper I .............................................. 9
       3.2.2. Paper II ............................................. 13
   3.3. Concluding remarks and future directions ........ 14

### Paper I

Nordberg, A., Rudén, C., and Hansson, S.O. *Towards more efficient testing strategies - analyzing the efficiency of toxicity data requirements in relation to the criteria for classification and labelling*, submitted manuscript.

### Paper II

1. Introduction

There is an estimated 70,000 industrial chemical substances commercially available on the European market today. For a majority of these substances the information regarding their toxicological and ecotoxicological properties is scarce. However, such information is needed to enable a robust risk assessment for the purpose of protecting human health and the environment. Thus, there is a need for the generation of more data for these chemicals. As the time and resources available for testing are limited, it is in practice impossible in the short run to generate data on all relevant endpoints for all these substances. Therefore there is a need for priority setting methods to help determine which substances to test first and what tests to start with. This thesis presents studies of two types of priority setting criteria and the evaluation of some of their consequences.

Section 2 gives a background outlining the process of risk assessment (2.1) and presenting a background on the priority setting in risk assessment of chemicals (2.2), the former and existing legislation regarding industrial chemicals in the European Union (2.3) and the work for reduction of the number of animals used in the generation of toxicity data (2.4). In section 3, the study and evaluation of priority setting methods conducted in this thesis is presented, with aim, methods and results (3.1-3.2), and finally some concluding remarks are made and future directions are discussed (3.3).

2. Background

2.1 Risk assessment


Hazard identification is the first step, aiming at determining the immanent ability of a substance to cause adverse effects in humans or in the environment in general by using an experimental (animal) model. The next step is the dose-response assessment, which is an assessment of the relationship between the dose and the response in the exposed subjects. The aim is to determine a point of departure for the risk assessment, such as a dose or concentration that does not produce any adverse effect in animals under certain experimental conditions, i.e. the No Observed Adverse Effect Level (NOAEL) for assessing the risks to human health and the corresponding Predicted No-Effect-Concentration (PNEC) for effects in environmental species. If it is not possible to obtain a NOAEL, then the lowest dose administered resulting in an adverse effect, the Lowest Observed Adverse Effect Level (LOAEL), may be determined instead. A benchmark dose is also sometimes used, i.e. the dose that corresponds to a specified response level, e.g. 10% in the exposed groups. Often the lower confidence limit of this dose is applied (WHO, 1999).
As a third step, an exposure assessment is performed, in order to estimate the size and nature of exposures to human populations or environmental species, including the relevant exposure pathways (ingestion, inhalation or through the skin) and the concentration, frequency and duration of the exposure. For human exposure three major exposure situations are usually considered: occupational exposure, consumer exposure and exposure via the environment. For environmental exposure a value for the Predicted Environmental Concentration (PEC) is estimated. (European Commission, 2003)

The last step is the risk characterisation. By combining the information from the first three steps the probability that adverse effects in man or the environment will occur is estimated. The risk to human health can be evaluated for each studied effect where a value of the N(L)OAEL is available, by comparing it with the measured or modelled level of human exposure. The ratio between the estimated exposure and the N(L)OAEL is called the Margin Of Safety (MOS). If the predicted exposures exceeds the N(L)OAEL, then the substance is considered to be of concern for this particular exposure and endpoint. If the exposure is found to be below the N(L)OAEL the assessor has to decide whether the substance is of concern or not, and on what further actions are warranted, if any, taking into consideration a number of factors such as intra- and interspecies variation, type of effect, differences in the exposure and which population that is likely to be exposed to the substance. If the general population, that includes elderly and children, is likely to be exposed the margin of safety should generally be larger than what is allowed if only workers are exposed. For the assessment of risk to the environment the PEC/PNEC ratio is determined; if this ratio is higher than 1 the substance is considered to present a risk to the environment. (European Commission, 2003)

2.2 Priority setting methods

The discussion on designing strategies for a more efficient testing of chemicals is not new. In 1984 the National Research Council published a report investigating the need for testing and how to construct systems for priority setting (NRC, 1984). For chemicals in commerce, they found a need for additional information and more testing, as (partial) health-hazard assessment was only possible for approximately 10% of the substances under study based on the information available. In the report it is concluded that the test systems and priority setting systems must be evaluated against their ability to fulfil their main goal: to give information that can guide decisions regarding the use and risk management of chemicals (NRC, 1984).

In a subsequent report by the Committee on Risk Assessment of Hazardous Air Pollutants in 1994, a tiered, iterative strategy for risk assessment was outlined (NRC, 1994). According to this strategy screening methods that “err on the side of caution” and relies on default assumptions should be used in lower tiers. When substances are found to present a risk according to the assessment based on the results in the first tiers, methods based on more precise information (instead of default assumptions) can be applied at higher tiers contributing to an improved basis for risk assessment.

Recently the National Research Council published a new report on the need for improved testing strategies (NRC, 2007). This report presents a strategy focusing on the increased use of in vitro tests and groups of such tests (e.g. tests utilising cells instead of living organisms,
Tiered test systems, where tests are performed in series, are common in the regulatory framework. Depending on the results of the test(s) at the first tier, the test(s) for the next tier are chosen (if any). A common way to construct tiered test systems with the purpose to make it possible to test many chemicals, is to apply tests with low costs at lower tiers and more expensive tests (for instance representing a higher complexity and duration) at higher tiers. Ideally tests should be combined on the basis of information regarding their predictive value, meaning that to design efficient test systems we need to know to what extent the initial, inexpensive tests predicts the outcome of the more resource demanding tests (Hansson & Rudén, 2007). This is an issue related to the ongoing, broader discussion on correlations between different test outcomes and what kinds of extrapolations (e.g. between different endpoints and between effects seen in different experimental models and species) that are scientifically acceptable (Rosenkranz and Cunningham, 2004; Pieters et al, 1997; Bokkers and Slob, 2005; Gaylor and Gold, 1998). Much of these discussions have however not been pursued with the purpose of designing efficient test systems.

2.3 European Legislation regarding industrial chemicals

A new legislation regarding industrial chemicals, REACH (Registration, Evaluation and Authorisation of Chemicals), has entered into force this year. However, parts of this thesis draw upon the rules in the former legislation. Therefore, the following subsections will describe these different legislations, both the ones no longer in force and the new legislation.

2.3.1 Classification and labelling

International work with harmonization of the classification and labelling of chemicals started in the 1950s and recommendations for the labelling of approximately 500 chemicals was published by the Council of Europe in 1962 (Lönngren, 1992). The Directive 67/548/EEC regarding the classification, packaging and labelling of dangerous substances was adopted in the European Communities in 1967, and implemented in 1970. One of the main purposes of this legislation is to provide hazard information to the users of chemicals thereby promoting the safe handling of chemicals in order to protect human health and the environment. The directive has so far been updated 29 times (Adaptation to Technical Progress), with the last update in 2004 (Commission Directive 2004/73/EC). The classification and labelling legislation specifies the criteria for classification and warning labelling of chemical substances and preparations within the European Union. Substances and preparations are classified based on their chemical and toxicological properties into the following classes: Explosive, Extremely flammable, Highly flammable, Flammable, Oxidizing, Very toxic, Toxic, Corrosive, Harmful, Irritant, Sensitisation, Carcinogenic, Mutagenic, Toxic to reproduction and Dangerous for the environment. Each of these classifications give rise to
corresponding warning labelling. The labelling of a substance is constituted of danger symbols, safety phrases and risk phrases (R-phrases).

The classification and labelling system also prescribes the appropriate test methods for identification of the different endpoints covered by the classification criteria (in Annex V), but it does not require any testing of chemicals. Test requirements have to be determined by separate rules such as REACH. Still, substances that have data fulfilling any of the criteria for classification shall be classified accordingly and be labelled with the corresponding warning label.

If data for a particular chemical and endpoint are lacking, then the criteria cannot be applied. Therefore, the chemical will remain unclassified for that particular endpoint. Hence, if a substance lacks a classification for a certain effect or property, this means that either (a) the substance has been tested and found not to fulfil the classification criteria, or (b) the chemical is insufficiently tested for that endpoint. This is a deficiency with the existing classification and labelling system, since it has the effect that the users of chemicals can not distinguish between a substance that has been tested and found not harmful and a substance that has not been (sufficiently) tested, by only studying the warning label of chemical products (see also Hansson & Rudén, 2003).

The classification and labelling directive also lists the chemicals that have a harmonised and legally binding classification in accordance with the directive (in Annex I).

A new system for the classification and labelling has been developed within the United Nations: the Globally Harmonised System of Classification and Labelling of Chemicals (GHS). This system is meant to replace the system(s) currently in force and it is to be implemented during the next few years. In this process, the European Commission has made a proposal for the new regulation regarding classification and labelling in accordance with GHS (Commission of the European Communities, 2007).

### 2.3.2 New and Existing Chemicals

To deal with the fact that such a large number of chemicals did not have any, or not enough, data on their toxicological properties, it was decided in 1979 that all new chemicals should be tested and notified before being allowed on the European market. To distinguish new chemicals from those already existing, an inventory of all chemicals that were available on the European Market between 1 January 1971 and 18 September 1981 was performed (Directive 79/831/EEC). These chemicals have since then been called “existing” chemicals and the list of existing chemicals, the European INventory of Existing Commercial chemical Substances (EINECS), contains 100,204 substances (see http://ecb.jrc.it/esis/). All substances not included in EINECS were then regulated as “new” chemicals, for which pre-market notification was required. The new substances were registered in the European LIst of

1 EINECS does not include synthetic polymers, impurities as such, intentional mixtures, medical products, cosmetic products and pesticide products, food, feedstuffs, alloys or most naturally occurring raw materials.
Notified Chemical Substances (ELINCS) and today there are about 4,381 substances registered in ELINCS\(^2\) (see http://ecb.jrc.it/esis/).

In order to notify a new chemical substance and to be allowed to put it on the market, the producer had to fulfil some data requirements. The amount of data required depended on the volume produced per year. Approximately 68.6\% of the new substances were notified in the category for chemicals produced or imported in volumes greater than 1 tonne per year and producer (see http://ecb.jrc.it/new-chemicals). For chemicals in this category, a “base set” of data had to be submitted, including data on physico-chemical properties, toxicological studies, ecotoxicological studies and the possibility of rendering the substance harmless\(^3\). For substances produced in higher volumes the submission of additional test data was required. The seventh amendment to Directive 67/548/EEC (Council Directive 92/32/EC) introduced principles of risk assessment for new chemicals.

Among the total number of chemicals on the market the vast majority falls into the category of existing chemicals for which no general test requirements have been in force. In order to deal with the fact that these chemical substances were not yet properly tested or assessed two new regulations were implemented: Council Regulation (EEC) 793/93 on the Control and Evaluation of the Risks of Existing Substances, and Commission Regulation (EC) No 1488/94 laying down the principles for the assessment of risks to man and the environment of existing substances. These two regulations state how existing chemicals should be prioritised for risk assessment and how the risk assessment should be conducted. The principles of risk assessment are further explained in the Technical Guidance Document (TGD) (European Commission, 2003).

According to the Council Regulation (EEC) 793/93 basic information on all existing substances produced in more than 10 tonnes per year\(^4\) should be reported to the Commission. For those produced in 10 - 1,000 tonnes per year and manufacturer, the so called lower production volume chemicals, a limited set of information was required (not including any physico-chemical or toxicological data), and for substances produced in more than 1,000 tonnes per year and manufacturer, the so called high production volume chemicals, available data on physico-chemical properties, pathways and environmental fate, acute and subacute toxicity, ecotoxicity, carcinogenicity, mutagenicity and toxicity for reproduction had to be submitted. However, if these data were not already available, no new tests on animals had to be performed.

The purpose of gathering this information was to perform risk assessments of each of the existing chemicals. Based on this information the Commission made priority lists specifying which of the over 100 000 chemicals to start with. According to Council Regulation 793/93 special attention should be given to substances that might have chronic effects. The producers

---

\(^2\) Types of substances that are excluded from ELINCS are pharmaceuticals, cosmetics, foodstuffs, substances used in scientific research, radioactive materials, wastes and pesticides, as these substances are regulated by other legislation.

\(^3\) Toxicological studies: acute toxicity, subacute toxicity (repeated dose), skin and eye irritation, skin sensitization, mutagenicity, screening for reproductive toxicity and assessment of toxicokinetic behaviour. Ecotoxicological studies: growth inhibitor on algae, bacterial inhibition, degradation and adsorption/desorption screening test.

or importers of substances on the priority lists had to submit all available data, and did they not already have all data that was requested in the first data collection step they had to perform the tests needed to generate this data. The next step was the evaluation of risk posed by the priority substances, which was performed by the member states of the European Community. When a substance was considered to present a risk, the evaluating member country should propose appropriate measures to reduce and control the risk. The priority setting among the existing substances has resulted in four priority lists that contain in total 141 substances, for which complete risk assessments was intended. For 77 of these substances final risk assessment reports have been produced so far (see ecb.jrc.it). As REACH has entered into force Council Regulation 793/93 is repealed with effect from 1 June 2008.

In the systems for new and existing chemicals the main parameter for priority setting was the length of time a substance had been on the market, resulting in a division of chemicals in the categories “existing” and “new” chemicals. This criterion has prioritised approximately 4,500 new substances for toxicological testing and left a majority of the 100,000 existing substances without such requirements. The second most important priority setting factor has been the production volume, determining the information requirements, reflected for instance by the 68.6 % of the new chemicals produced in volumes higher than 1 tonne per year prioritised for a larger test battery than those produced in smaller amounts.

2.3.3 REACH – the new chemicals legislation

On 1 June 2007 a new legislation for industrial chemicals, regulating the Registration, Evaluation and Authorisation of Chemicals (REACH), entered into force (Regulation (EC) No 1907/2006). With this regulation there is now less focus on how long a substance has been on the market, as the previous regulatory division between new and existing industrial chemicals no longer exists. However, so called phase-in substances (e.g. existing chemicals in EINECS) and non-phase-in substances have different deadlines for registration in REACH and there is still some transitional provisions for phase-in substances and the substances previously called new (i.e. notified). Anyhow, the production volume is still the main means for prioritisation of testing and risk assessment of chemicals, determining what substances are covered by the legislation and what data to submit for the registration of the chemicals included in REACH. For an estimated 30,000 substances, produced (or imported) in more than 1 tonne per year and producer (or importer) (Commission of the European Communities, 2001), registration will be required. The principle of this system is the same as in the old legislation for new substances; the higher the production volume, the more information about the substance is required.

The former legislation prioritised approximately 4,500 substances for testing. The scope of REACH is approximately 30,000 chemicals. This volume based inclusion criterion represents a first significant priority setting step, leading to the exclusion of about 40,000 substances from registration. Toxicity testing and chemical safety reports are however only requested for the about 10,000 substances produced in more than 10 tonnes per year (Commission of the European Communities, 2001), thus REACH requirements will leave about 60,000 industrial chemicals insufficiently tested and risk assessed.
REACH will be implemented gradually up until 1 June 2018. The REACH regulation does not concern classification and labelling and the before mentioned classification and labelling system (67/548/EEC) is still in force.

REACH applies tiered approaches to testing as there are three different possible implications, based on the results of the tests that have to be conducted according to the requirements based on production volume. Based on these test results in lower tiers (a) the available test data can be considered to give enough information regarding a certain type of hazardous property or effect, hence (some of) the subsequent planned tests could be regarded unnecessary to perform, (b) the planned testing could be unaffected, or (c) more tests could be requested, outside the original test strategy, due to results indicating other effects that needs to be investigated. In other words, some tests can be waived if certain other tests have shown results complying with specifications in the regulations, e.g. if a substance is found to be a strong acid (with a pH < 2) the in vivo test (i.e. tests on living organisms) for skin sensitisation should not be conducted (Regulation (EC) No 1907/2006, Annex VIII). On the other hand, additional tests can also be requested on the basis of results from the obligatory tests, e.g. if the result of a subacute or subchronic study shows certain effects then a two-generation study of reproduction toxicity shall be performed on one species. Based on the result from this study, further studies can in turn be deemed needed (Regulation (EC) No 1907/2006, Annex IX).

Another example of a tiered approach or priority setting criterion in REACH is that special attention should be paid to substances meeting the criteria for being persistent, bioaccumulating and toxic (PBT), or very persistent and very bioaccumulating (vPvB) (Regulation (EC) No 1907/2006, Annex XIII); If, for a particular substance, a need for more information has been identified, the registrant has to submit a so called testing proposal to the new European Chemicals Agency (ECHA). In their examination of testing proposals, PBT and vPvB substances are prioritised. These substances are also prioritised when ECHA shall recommend which substances that need further evaluation or authorisation. This prioritisation is based on the assumption that substances with the ability to persist in the environment or in living organisms are presenting a risk due to the prolonged exposure and an accumulation in the environment that might be hard to reverse (European Commission, 2003, part II).

Other tools for priority setting that are included in REACH are mathematical models describing correlations between a chemical substance’s molecular structure and its toxicological properties and reactivity, so called (Quantitative) Structure-Activity Relationships ((Q)SAR). There are several (Q)SAR models that are being used as priority setting tools within the EU and the US (Cronin et al 2003a, 2003b). According to TGD, there are four main uses of these models: (a) to generate necessary data that are missing, (b) to evaluate existing experimental data from tests not conducted according to guidelines or with incomplete information on the tests, (c) to give supporting information for the determining of whether more tests should be conducted, and if so which test strategy to use, and (d) to predict what effects, where data are missing, that might be of concern (European Commission, 2003, part III). The TGD gives guidance on how to use (Q)SARs in risk assessments and recommends some (Q)SAR models for certain environmental endpoints. There are furthermore several (Q)SAR models for the prediction of human health endpoints, e.g.
carcinogenicity, mutagenicity, acute toxicity, reproductive toxicity, eye and skin irritation and skin sensitisation (Cronin et al 2003b). However, in the recent edition of TGD, there are no recommendations for (Q)SARs regarding human health endpoints (European Commission 2003, part III).

Historically, there has been some reluctance in using (Q)SARs for risk assessment in Europe due to lack of validated methods, but REACH opens up for an increase in the regulatory use of (Q)SAR. One of the major factors determining the regulatory acceptance of (Q)SAR data concerns the validation of the models. An important step towards a harmonized process for model validation is the work performed by the member countries of OECD, making a first attempt to determine commonly agreed upon principles for validation of such models (Netzeva et al, 2005).

2.4 Animal Welfare

There is ongoing extensive work about how to improve animal welfare with regards to the testing of chemicals. One often mentioned way of meeting these goals are the three Rs: Refine, Reduce and Replace, coined by Russell and Burch (1959). Council Directive 86/609/EEC regarding the protection of animals used for experimental and other scientific purposes was adopted with the purpose to reduce the number of animals used in experiments and to ensure that the pain and distress to these animals are kept to the minimum. The reduction of the use of experimental animals is also mentioned as an important aspect in REACH. The strong emphasis on the reduction of animal experiments in REACH is demonstrated by the fact that if the required information for a substance is not already available, then all available data from in vitro and in vivo studies, historical human data, (Q)SARs and data on structurally related substances shall be assessed first. If these data are not sufficient, depending on the annual production volume per producer, either new data has to be generated through new tests (if produced in 1 - 100 tonnes per year and producer) or a testing strategy has to be submitted (if produced in more than 100 tonnes per year and producer).

Another important judicial change, influencing the work with finding alternative methods, is the seventh amendment to the directive for cosmetic products (Directive 2003/15/EC), entering into force in 2009. According to this directive, no tests on animals will be allowed of the final product or any of its ingredients. The methods developed to replace in vivo models are mainly in vitro methods. As before mentioned, there is also an intense work with the development of (Q)SAR models, and to develop generally accepted methods and principles for the validation of these methods.

---

5 With the exception of test for repeated-dose toxicity, reproductive toxicity and toxicokinetics that will be be allowed at least until 2013, due to lack of validated alternatives.
3 Study of priority setting strategies

3.1 Aim and methods of the thesis

As outlined above, several priority setting strategies have been applied in previous and existing legislations and one of the most influential is production volume; the higher the production volume, the more regulatory requirements. However, even though production volume will give us some information relevant for assessing the probability of exposures, it will say little about the toxicological properties of the chemicals, and hence about the risks that exposures may pose (Hansson & Rudén, 2006). The aim of this thesis is to contribute to finding new ways to evaluate priority setting methods in chemicals control, and the overall objective is to generate knowledge and methods that can help to improve the efficiency of test strategies for industrial chemicals.

The work presented below draws upon criteria and information generated within (1) the system for classification and labelling of dangerous chemicals in the European Union, (2) the European legislation for chemicals notified as new substances and (3) the REACH legislation. The data obtained from these sources for a reasonably large number of chemicals have been evaluated statistically with the purpose to identify potential correlations between the outcomes of tests and different risk management consequences. In particular data on the classification and labelling of individual compounds, bioconcentration factors, and test data availability as indicated by notification category (i.e. production volume range) were used.

Summaries of the two papers included in this thesis are presented in the following section.

3.2 Preview of the papers

3.2.1 Paper I

Materials and methods

The testing procedures and the decisions on toxicity classifications and warning labelling taken for new industrial chemicals introduced on the European market were explored, using a database containing information on all substances notified as new industrial chemicals according to the European legislation between 1994 and 2004, made available to us by the European Chemicals Bureau. For this study we have utilized data for 1 409 substances produced (or imported) in production volumes greater than 1000 kg per year. The database contains the following information: CAS-number, year of notification and its classification and labelling as proposed by the Competent Authority to which the notification was submitted.

The risk phrases reveal which type of test was the basis of the classification. Hence, if a substance is classified as harmful and labelled with a particular risk phrase indicating that it is harmful if swallowed, then it can be concluded that a positive result has been obtained in a test of oral acute toxicity. If it is assigned another particular risk phrase indicating it being harmful after prolonged exposure, then it can be concluded that it has yielded a positive outcome in a repeated dose test. If it has both these risk phrases, then positive outcomes have been obtained in both types of tests. This crucial connection between R-phrases and test
outcomes (that was also exploited in Rudén and Hansson, 2003) is an essential part of the method used here.

The reason for the study of chemicals notified with a production volume exceeding 1000 kg per year and manufacturer is that several types of tests, including tests for acute toxicity, subacute (28-day) toxicity, irritation and sensitisation, had to be performed for chemicals in this category. The notification rules gave little possibilities to avoid any of the required tests and hence it can, with reasonable confidence be assumed that a substance in this group that has not received a risk phrase indicating acute toxicity still were tested for acute toxicity, but with a negative outcome.

The prevalence of a certain risk phrase among these substances was calculated and used to estimate the probability of meeting the criterion for this risk phrase. This information was combined with information on the test requirements and the cost of the testing required to arrive at that particular risk labelling. Based on this information an efficiency ratio was developed for the purpose of this study. The efficiency ratio of a test or combination of tests is thus the ratio between (i) the probability of meeting a criterion for a certain classification when testing a substance and (ii) the cost of performing that particular test. This ratio should be as high as possible, so that as many toxic substances can be identified as possible in relation to the cost of testing.

For the purpose of this study, it is the proportions between the prices for different tests, rather than their absolute values, that are important and a monetary unit of €1,000,000 was used in the calculation of efficiency ratios, i.e. how many chemical substances eligible for a certain risk phrase would be expected to be found when carrying out tests worth €1,000,000.

As estimates of the costs of different types of tests we used the costs of individual tests provided in a report for the European Commission (RPA, 2003). These calculations were made for three types of test systems, given that we would only have access to two tests ($T_1$ and $T_2$): (a) the single test where only one of the tests $T_1$ and $T_2$ is performed, (b) the full combination of the two tests when both $T_1$ and $T_2$ are performed on all tested substances and (c) a combination with a negative coupling where $T_1$ followed by $T_2$ is performed on those substances that yield a negative test outcome in $T_1$ (or the other way around). We focused on cases when outcomes from two tests are used as alternative criteria for the same risk management decision, namely the classification of a substance as dangerous to health.

**Results**

Under the assumption that the frequency of classifications in the studied set of chemical substances is representative for untested substances we should expect to discover 112 substances that should be classified and labelled at least as harmful due to acute oral toxicity if we would spend €1,000,000 on testing (a random selection of) untested substances. The corresponding efficiency ratios for subacute toxicity, irritation and sensitisation are listed in Table 1.
Table 1. Prevalence of classifications and the efficiency ratios for single tests, for substances notified in category VIIA (> 1000 kg/year and producer) between 1994 and 2004. The total number of substances in this dataset is 1409.

<table>
<thead>
<tr>
<th>Test</th>
<th>Prevalence of corresponding classification</th>
<th>Efficiency ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute toxicity</td>
<td>20%</td>
<td>112</td>
</tr>
<tr>
<td>Subacute toxicity</td>
<td>9.3%</td>
<td>2</td>
</tr>
<tr>
<td>Irritation</td>
<td>29%</td>
<td>156</td>
</tr>
<tr>
<td>Sensitisation</td>
<td>28%</td>
<td>72</td>
</tr>
</tbody>
</table>

If the testing of untested substances aims at discovering as many substances as possible that should be classified as harmful, or in other ways dangerous to health, then the tests for acute toxicity, irritation and sensitisation are drastically more efficient for this purpose than the test for subacute toxicity.

The efficiency ratios for full combinations of tests and for negatively coupled tests were also calculated. From this calculation it could be concluded that in the negatively coupled test systems, the order between the two tests did not have any major impact on the efficiency ratio.

Testing for subacute toxicity is much more expensive than the other tests for health classification under consideration here and the prevalence of substances classified as subacutely toxic are much lower than the other tested health effects, hence those test systems that contain this test have much lower efficiency ratios than those that do not.

Another thing of interest is to see to what extent acute toxicity “covers” the other types of toxicity, i.e. the percentage of substances with other classifications that are also classified as acutely toxic. Approximately 60.8 % of the subacutely toxic substances are also acutely toxic. Less than a third of the irritating substances exhibit acute oral toxicity, and the same applies to sensitising substances (Table 2). However, even though 39.2 % of the subacutely toxic substances are not classified as acutely toxic by oral intake, as many as 83.8 % of the subacute substances (109 out of 130) have a classification for irritation, sensitization and/or acute toxicity and as mentioned above and only 1.5 % out of all substances were classified as subacutely toxic without having a classification for any of the other three effects (Table 3). In a situation where data on acute toxicity is already available, the results shown in Table 2 and 3 can be used as an argument for giving higher priority to additional tests for irritation and sensitisation than to additional tests for subacute toxicity.

Table 2. The proportion of different toxicants in notification category VIIA (>1000 kg) that are also being classified as acutely toxic (oral)(R22, R25 and R28). This table shows that 60.8 % of the subacutely toxic substances are also acutely toxic, etc.

<table>
<thead>
<tr>
<th>Classification</th>
<th>Prevalence of labelling for acute oral toxicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subacutely toxic</td>
<td>60.8%</td>
</tr>
<tr>
<td>Irritating</td>
<td>27.7%</td>
</tr>
<tr>
<td>Sensitising</td>
<td>30.1%</td>
</tr>
</tbody>
</table>
Table 3. Prevalence of classified substances in notification category VIIA (>1000 kg), that are tested both for (i) subacute toxicity (R48) and (ii) acute oral toxicity (R22, R25 and R28), irritation (R36, R38 and R41) and sensitisation (R43). The total number of substances in this dataset is 1409.

<table>
<thead>
<tr>
<th>SENSITISING OR IRRITATING OR ACUTE TOXIC (ORAL)</th>
<th>NOT (SENSITISING OR IRRITATING OR ACUTELY TOXIC (ORAL))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subacute toxic</td>
<td>109 (7.7%)</td>
</tr>
<tr>
<td>Non-subacute toxic</td>
<td>656 (46.6%)</td>
</tr>
<tr>
<td></td>
<td>21 (1.5%)</td>
</tr>
<tr>
<td></td>
<td>623 (44.2%)</td>
</tr>
</tbody>
</table>

Discussion and conclusion

The efficiency ratios, as estimated on the basis of toxicity data for a large number of industrial chemicals, reveal significant differences among standardized tests and the corresponding criteria for classification and labelling for acute toxicity, irritation, sensitization and subacute toxicity in their capability of identifying chemicals eligible for classification and warning labelling.

The most striking result of our investigation is the low cost efficiency of the 28-day study compared to the short-term toxicity tests in giving rise to risk management measures in the form of warning labelling according to the classification and labelling directive. The tests needed for labelling a substance as being dangerous after a prolonged exposure, i.e. repeated dose studies (in this case a 28 day-study), lead to far less substances actually being assigned a risk phrase than after testing for acute toxicity, irritation, or sensitisation, given that the same amount of resources is spent on testing. Given that a positive acute toxicity test can give the same danger symbols, the relatively high prevalence of substances being classifiable as acutely toxic and the scarcity of resources, it seems to be more cost efficient to perform acute toxicity tests on a larger number of substances rather than to perform additional subacute toxicity studies on the substances already tested for acute toxicity.

It should be noted that the criteria for how to classify substances on the basis of test results play a significant role in this analysis since they dictate both what kind of tests that are required and how the results of these tests should be interpreted. The efficiency of the different tests in terms of resulting in a classification is thus highly dependent also on how these criteria are defined.

Data generated from regulatory testing are however not just used for the purpose of hazard identification as performed within the classification and labelling system. The data generated in these different tests are also used in more comprehensive risk assessment of chemicals and an additional aspect is therefore of what relevance the different tests are for the risk assessment process. Currently more complex tests like the 28-day study are likely to be of more use in these types of evaluations than in the classification and labelling system. Moreover, less complex tests usually have less relevance for human health risk assessment which may result in over-regulation or under-regulation, as well as unwanted proportions of either false positive results or false negative results.
3.2.2 Paper II

The aim of this study is to discuss potential consequences of having a tiered test system that uses characteristics describing a substance’s potential to be bioaccumulating as a tool for prioritisation. Thus, the correlation between a chemical’s bioaccumulating potential as estimated by its bioconcentration factor (BCF) and its toxicity as described by its classification according to the classification and labelling directive (Directive 67/548) was investigated. The study focused on chemicals classified for (a) acute toxicity, or (b) carcinogenicity, mutagenicity and/or reproductive toxicity. A group of unclassified chemicals was also included for comparison.

Materials and method

For the purpose of this study, a database that includes a total of 437 chemicals was compiled. The database includes information on 257 chemicals classified for cancer, mutagenicity and/or reproductive toxicity, 90 chemicals classified for acute toxicity and 90 chemicals identified as unclassified.

To identify carcinogenic, mutagenic and reproductive toxicants, the list of chemicals with harmonized classification in Annex I to Directive 67/548 on the classification and labelling of dangerous substances was used. From this list all chemicals classified for carcinogenicity (C), mutagenicity (M) or reproductive toxicity (R) with CAS-numbers were identified. Those 257 substances that also had BCF data in the CAS Registry database were included in our database.

This database was thereafter supplemented with chemicals classified as acutely toxic. Out of all chemicals classified for acute toxicity in Annex I, the 90 substances having BCF data in the CAS Registry database (out of the 100 most acutely toxic chemicals not classified for carcinogenicity, mutagenicity or reproduction toxicity) were added to our database.

Finally, a random sample of 90 substances being unclassified (according to the Chemical Substances database administered by Prevent; http://kemi.prevent.se) were added to the database, together with their BCF values (in the CAS Registry Database).

We used the information in this database to investigate whether the BCF values correlated with hazard classification and to make a preliminary assessment of the consequences of the use of the bioaccumulating potential as a criterion for priority setting in REACH. In particular; will chemicals with any of these particular hazard classifications be over- or underrepresented among the bioaccumulating chemicals?

Results

Of all chemicals in our database, 12 % had potential for bioaccumulation according to the REACH criterion, i.e. BCF > 2000. Among the compounds only classified for mutagenicity, 7 % had a BCF > 2000, and among substances only classified as toxic to reproduction or only
as carcinogenic, 15% and 12% respectively had a BCF > 2000.⁶ Among the acutely toxic and the unclassified chemicals, 12% and 14% respectively had a BCF > 2000.

Several comparisons were made, between the number of substances that had a BCF ≤ 2000 and BCF > 2000 in the different groups of chemicals. None of the comparisons showed any significant difference between the different groups ($\chi^2$-test, $p > 0.05$).

**Conclusion**

When comparing the number of bioaccumulating chemicals classified as (a) acutely toxic, (b) carcinogenic, mutagenic, and/or reproductive toxicants, or (c) lacking a toxicity classification no correlation between the chemicals’ BCF and their classifications was found.

Based on these results it can be hypothesized that prioritizing chemicals with bioaccumulating properties is neutral with respect to these types of toxicity. Or in other words, if these data are representative, the use of these properties for prioritization of long-term testing is not likely to have a bonus effect of also prioritizing chemicals that have higher than average probability of being acutely toxic, carcinogenic, mutagenic and/or reproductive toxicants. Neither is it likely to have the disadvantageous effect of prioritizing chemicals that have lower than average probability of having these toxicological properties.

This result thus supports the use of bioaccumulation data at first tier for selecting chemicals that should be submitted to further testing. Prioritizing chemicals with bioaccumulating properties for long-term testing has the potential to be part of a scientifically sound test strategy for determining the toxicity of these substances, and hence to make a reasonably robust PBT assessment.

**3.3 Concluding remarks and future directions**

The results in this thesis are highly dependent on the criteria for different types of classifications, as the criteria dictate both what kind of tests that are required and how the results of these tests should be interpreted. The risk phrases and safety phrases are specific for most endpoints used in the classification and labelling criteria, but the criteria also suggest which endpoints that are considered as indicating equally great risks by assigning the same classification to the positive outcome of these tests. This fact is used in the first paper, when comparing two different tests, possibly giving the same type of classification, but to very different costs and with different probabilities of resulting in a classification.

The evaluation of the 28-day test for subacute toxicity in paper I is not an evaluation of the inherent, scientific qualities of that test. The evaluation is limited to its regulatory efficiency when used for hazard identification in the current classification and labelling system. The finding that the 28-day test has a low efficiency ratio compared to the acute toxicity can thus not be seen as an indicator of the overall significance of this test in contributing information to the more comprehensive risk assessment process.

---

⁶ Only, as in mutagenic and not carcinogenic and not toxic to reproduction. Other classifications could be possible, but those are not taken into account for this group in this study. A similar explanation applies to statements about chemicals classified as only carcinogenic or only toxic to reproduction.
A problem with using the classification and labelling of chemicals in paper I and II is that we do not have access to the toxicological data or, as in the case of Paper II, not even information about whether the substances are tested at all or not. In paper I we assign the probability of a substance being classified with the prevalence of the same classification. This is possible as we have good reasons to believe that all substances included in the study have actually been tested for the same endpoints, due to the legal requirements for this specific dataset. With the existing chemicals, that assumption can usually not be made.

Finally, in order to increase the efficiency of chemicals control it is crucial to continue to improve procedures and principles of toxicity testing, and in this endeavour priority setting principles play a significant role. This thesis represents an attempt to evaluate different priority setting strategies empirically. The main obstacle in performing such studies is the lack of sufficiently large, and publicly available, sets of systematically obtained data for a reasonably large number of chemicals. With the implementation of REACH there is good hope that such data sets within the next decade will be generated for a limited proportion of the industrial chemicals. However, these data are likely to be covered by rules of confidentiality and will probably, and unfortunately, not be available for scientific purposes. If, on the other hand, the data generated within REACH were allowed to be subject to the scrutiny of scientists, these data could significantly contribute to the possibility to perform empirical evaluations of the regulatory outcomes of testing, and ultimately to improve priority setting principles for the substances that remain to be tested, thereby increasing the efficiency of the legislation towards the aim to protect human health and the environment.
References


