Evaluating Progress in Pesticide Risk Reduction:

Summary Report of the OECD Project on Pesticide Aquatic Risk Indicators

Paris, 2002

Contents

Project Origins and Purpose 3

Development of the OECD Aquatic Risk Indicators 3

Lessons Learned at Braunschweig Workshop 8

Pilot Testing Phase 9

Summary Conclusions from Pilot Testing 16

Status and Next Steps for the OECD Aquatic Risk Indicators 17

Project Participants 18

Project Origins and Purpose

The idea for the OECD Aquatic Risk Indicator (ARI) project emerged from the first OECD Workshop on Pesticide Risk Indicators, held in Copenhagen in April 1997. At that time some member countries had committed to pesticide risk reduction, as measured by indicators mainly driven by quantities of pesticide sold. Many countries at the Workshop were interested in development of 'science-based' risk indicators linking use, fate, and hazard data to measure changes over time in aggregate risk and the impacts of risk reduction initiatives.

The Workshop defined a list of "principles" for pesticide risk indicators. Salient among them were these:

- Indicators should be scientifically robust and user friendly
- Indicators should link hazard and exposure data with use data
- Indicators should complement but not replicate or compete with the tools used for registration and risk assessment
- Separate indicators of risks to human health and to different compartments of the environment would be better than a single indicator of pesticide risks

The summary specification for pesticide risk indicators called for them to be neither too simple nor too complex, to be easy to use, to provide accurate and meaningful information, and to be sound analytical tools with sophisticated inputs but easily understood outputs.

Development of the OECD Aquatic Risk Indicators

Following the Copenhagen workshop the Risk Reduction Steering Group of the OECD Pesticide Working Group created an Expert Group (EG) in March 1998. The Expert Group was asked to develop and test indicators of aggregate aquatic risk consequent to agricultural pesticide use[1] which would meet the specification developed at Copenhagen. The EG included members from seven countries, supported by consultants from the Central Science Laboratories of the UK Ministry of Agriculture, Fisheries and Food.

The EG began by surveying the field to see what indicators of potential utility already existed. Many were found and evaluated. Most were designed to support choice of a pest management tool by individual users; others were designed for another relatively narrow purpose. None were found to be suitable for aggregate trend analysis.

In the next step the EG developed prototypes for new indicators. First, the EG identified a set of potentially relevant variables based on knowledge of the mechanisms by which pesticides move from the site of application to surface water, and of the hazard of

[1] Aquatic risk was chosen for the initial focus of the project, rather than terrestrial environmental risk or human health risk, because it would be simpler, and would be a reasonable first step toward a comprehensive suite of indicators.

pesticides to aquatic organisms. Table 1 below summarizes the range of variables considered.[2]

Table 1: Variables Used in Constructing Aquatic Risk Indicators

Category	Examples	Comments
Use data	Pesticide Crop Area treated Dose rate Frequency of treatment Method of application Observance of spray drift buffer	May be derived from surveys, statistically 'raised' to national level. May also be estimated by allocating the total quantity of a pesticide sold over all registered crop uses in proportion to the area planted to each crop.
Fate variables	Persistence Mobility Solubility	Generally these are data points also used in registration assessments
Application site variables	Adjacency to surface water Slope Presence of runoff buffer Precipitation	Can be replaced by constants at national level; potentially valuable at regional level to support comparisons between regions, but only if regional usage data is also available.
Hazard variables	Acute toxicity to algae, *Daphnia*, & fish Long-term toxicity to algae, *Daphnia*, & fish	

The next challenge was to consider key structural issues for the indicators:

- Are variables continuous or are they reduced to scores? For example, is area treated measured continuously in hectares, or is it expressed as a value of 1, 2, or 3, signifying small, medium, or large area?

- Are variables linked mechanistically, through simplified versions of the models used in registration reviews? Or are scored variables added together? Or linked logically?

The EG defined three new indicators, representing different answers to these questions about structural options. All three roughly estimate how much of an applied pesticide migrates from the site of application to surface waters, and its significance for aquatic organisms. But they do this in very different ways.

REXTOX[3] links use data, fate variables, and application site variables using simplified mathematical models of the mechanisms of pesticide fate and movement in the environment, to estimate pesticide concentrations in surface waters as a function of use. This estimate is then multiplied by the total amount used to obtain scaled[4] estimates of exposure. These exposure estimates are linked to hazard data to estimate risk, and finally the exposure

[2] The final list of variables and the linkages among them are specified in the Report of the OECD Pesticide Aquatic Risk Indicators Expert Group.

[3] The name "REXTOX" is derived from **R**atio of **EX**posure to **TOX**icity, "ADSCOR" from **AD**ditive **SCOR**ing, and "SYSCOR" from **SY**nergistic **SCOR**ing.

[4] The "unscaled" results of the indicators estimate use and risk on hectare; the "scaled" results estimate total use and risk at a national level.

and risk values are combined across all uses of all pesticides to yield the aggregate indicator.

ADSCOR uses tables to convert true values to scores for use variables including the method of application, the dose rate, the frequency of application, and observance of buffer zones. These scores are added together to obtain an unscaled exposure score for each use, multiplied by the actual area treated to obtain a scaled exposure score for each use, and summed across all uses to get an aggregate exposure score for each pesticide. This is then linked to hazard data just as in REXTOX, and exposure and risk values are combined across all uses of all pesticides to yield the aggregate indicator.

In SYSCOR, all exposure-related variables (including area treated) and all hazard variables are converted to scores. The scored variables are combined logically, using predefined tables, to yield an overall ‘penalty’ score for each use; these ‘penalties’ are then summed across all uses and all pesticides to yield the aggregate indicator. The rules for combining scores provide for synergy between factors; for example, if it’s bad for a pesticide to be persistent, and if it’s bad for it to be mobile, it’s especially bad for it to be both persistent and mobile. Where ADSCOR would add the persistence and mobility scores, SYSCOR adds an additional ‘penalty’ when both persistence and mobility are high.

Table 2: Structure of OECD Aquatic Risk Indicators

<table>
<tr><td rowspan="2">UNSCORED
Variables
Continuous</td><td>REXTOX</td><td>Π Uses true values for all variables
Π Links variables in mechanistic models
Π Calculates either scaled or unscaled</td></tr>
<tr><td rowspan="2">ADSCOR</td><td>Π Uses true values for area and toxicity
Π Uses scored values for other exposure variables</td></tr>
<tr><td rowspan="2">SCORED
Sensitive to
Breakpoints</td><td>Π Links Exposure scores by addition
Π Calculates either scaled or unscaled</td></tr>
<tr><td>SYSCOR</td><td>Π Uses true values for toxicity and scored values for all other variables
Π Always scaled
Π Links scores logically with synergistic effect</td></tr>
</table>

The consultants from the UK Central Science Laboratories then wrote software to calculate the prototype indices as defined by the Expert Group, compiled test data sets for all required variables, and tested all three indices. The scope of this initial testing was limited to arable crops and orchard crops in England and Wales, with use data based on surveys in seven years from 1977 to 1996 for arable crops and in four years from 1983 to 1996 for orchard crops. The use data showed 307 pesticides were used on these crops, and fate and hazard data were compiled for all these pesticides.

After initial testing the EG reconvened to review preliminary results, refine the models, and specify further analyses and reports. CSL implemented the specified refinements and re-tested the indicators with the same range of use data. The EG met once more to assess the results of this second round of testing and prepare its final report.[5]

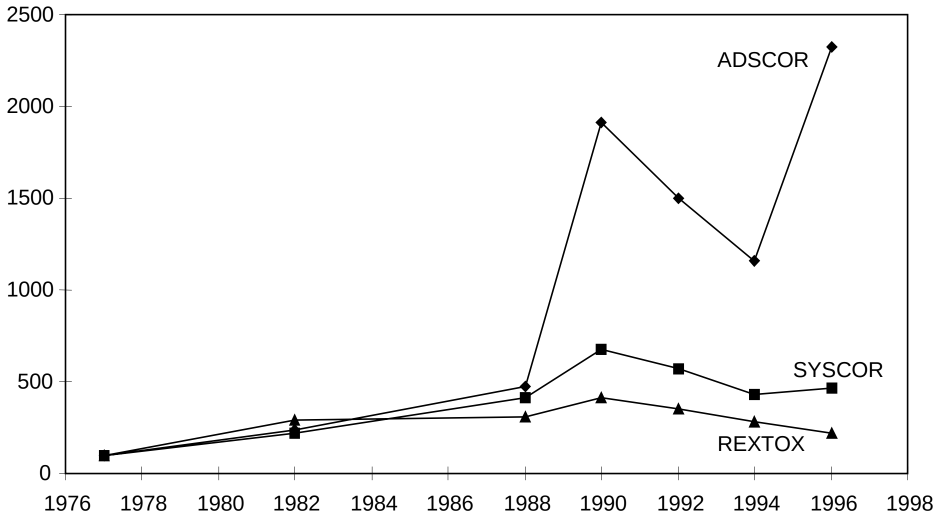
[5] The Report of the OECD Pesticide Aquatic Risk Indicators Expert Group was published in 2000.

The EG's most important finding was that different indicators can produce different risk trends, even when using the same input data. This is mainly due to differences in indicator structure, which give different variables more or less weight. Risk trends calculated by REXTOX, ADSCOR and SYSCOR diverged in several instances. The structural causes of the divergence were identified–for example, REXTOX and SYSCOR give more weight to spray drift buffers than does ADSCOR–but it is not clear which representation is more nearly correct.

Figure 1 below illustrates dramatically divergent trends for the combined acute risk to algae, *Daphnia,* and fish from arable crop use, as calculated by the three new indicators.

Figure 1: Divergence of Indicator Results

Arable - Acute Scaled (Normalised Values, All Taxa)



All three indicators show a gradual but steady rise in aggregate aquatic risk from 1977 to 1988, a clear peak in 1990, and a decline in 1992 and 1994. The risk trend for ADSCOR is quite different from the other two indicators, showing much greater fluctuation after 1988. The indicators differ in the period from 1994 to 1996, with REXTOX continuing to decline, ADSCOR rising again steeply to a new peak, and SYSCOR rising slightly.

To understand the different trends, the EG looked to see which pesticides contributed most to risk in each survey year, and then at usage data for those pesticides during the years of concern. Arable crop use of cypermethrin, the pesticide that contributed most to risk for REXTOX and ADSCOR and figured importantly in SYSCOR, rose steeply between 1988 and 1990, dropped in 1994, then rose again sharply in 1996

What explains the divergence among the three indicator trends? Why, if cypermethrin use jumped in 1996, did the REXTOX risk trend decline and the SYSCOR trend increase only slightly? The answer is that the three indicators give different weight to the use of spray drift buffers, which were required for cypermethrin used in the UK beginning in 1992. REXTOX assumes that required buffers are always observed and completely effective in preventing pesticides from leaving the application site; SYSCOR and ADSCOR are less optimistic.

In reporting these results the EG noted that aggregate indicators are by their nature and design necessarily crude tools–much less sensitive than those used in risk assessment of individual pesticides. When different models give different results, we don't know which answer is closest to indicating real risk.

The EG also found that:

- It is entirely feasible to build indicators of aggregate pesticide risks over time based on actual usage, fate, and hazard data rather than on worst-case assumptions. Compiling the data sets is the main cost, but much of the data (i.e., use data and pesticide fate constants) is potentially valuable for many other purposes. Computing the indicators is easy, fast, and cheap once the models are built.

- Indicators can be calculated at varying scales and with varying scopes in addition to national aggregates. When aggregated across pesticides, uses, or regions, they can give insights not available from case-by-case registration risk assessments. They can also be used to compare the risk potential of alternative pest management regimens, or to simulate impacts of at least some policy measures before they are implemented.

- Indicator work should begin at a low level of aggregation, calculating the risks to individual aquatic organisms in specific regions caused by individual pesticides, crop uses, or application methods. Risks calculated at a low level can later be combined to yield higher-level aggregates, but unless the work begins at a low level, the underlying causes of trends cannot be discovered.

- Specific pesticides and uses contributing disproportionately to an aggregate indicator can be identified for close attention. “Top ten” lists of the pesticides contributing the most to the index are useful for this purpose, as are graphic displays of detailed information underlying the aggregate indicator. There are many options for displaying the wealth of information contained in the ARI, and different presentation tools are needed to suit different purposes and different audiences.

- In general, existing data sets are inconsistent, with many missing values, especially for older pesticides and for long-term hazard. Bringing the necessary data to a consistent standard usable for calculating the indices is time-consuming and difficult. Estimating missing values for old pesticides lessens the accuracy and usefulness of long-term trend analysis.

- The data on intrinsic properties of pesticides–hazard, fate, etc.–can be shared among countries. Pesticide use data and application site variables are necessarily locally specific, and must be compiled anew by each participating country.

- Some features of the models–specifically the breakpoints in the scoring tables for dose rates, treated area, and other exposure variables–should be customized by each participating country to reflect their scale and practices of agriculture.

- The method of Gutsche and Rossberg[6] for converting sales data to use estimates is

[6] Gutsche, V. and Rossberg, D. (1999) A Proposal for Estimating the Quantity of Pesticide Active Ingredients Applied by Crop on National Sales Data, is contained in Annex 2 of the Report of the OECD Pesticide Aquatic Risk Indicators Expert Group ***(see Attachment 1 at the end of Annex 2)***.

laborious and has limitations, but produces usable results from relatively little input data. Estimated use data, however, gives less reliable indicator results than actual usage data. If you must estimate use data, use the estimates with care.

Lessons Learned at Braunschweig Workshop

The Expert Group presented their report to the second OECD Workshop on Pesticide Risk Indicators in Braunschweig in June 1999. Based on their discussion of the OECD aquatic indicators project and of individual countries' experiences with pesticide risk indicators, the workshop participants expanded on the Copenhagen principles with the following 'lessons learned':[7]

Indicator design and interpretation

- When designing an indicator, it is very important to be clear about the indicator's purpose. The purpose will help to determine how sophisticated a methodology is needed and how much and what types of data are needed.

- When interpreting and communicating indicator results, it is again important to be clear about the indicator's purpose. Different purposes may require different ways of presenting results.

- OECD governments need to learn how to use pesticide risk indicators, and how to interpret and communicate their results and limitations.

Missing data

- Missing data can be a significant barrier to the use of indicators. Both the Expert Group and individual countries reported problems with missing data. Activities that would lead to more complete and useful data would be very helpful to the successful use of indicators. For example, it would be useful to develop a shared database on pesticide properties, starting with existing databases (including that developed for the aquatic indicators project), although the work to create and maintain such a database would be substantial. In the meantime, indicator users should state clearly what they did to estimate, or compensate for, missing data.

Limitations and drawbacks of indicators

- Indicators can be useful as signposts and analytic tools, but should not be the sole basis for decision-making.

- Indicators are just relative measures, not exact measures of real risk, and some see this as problematic. Whether or not it is important for an indicator to correspond closely to real risk depends upon the purpose of the indicator and how it is used.

---

[7] More information is given in the Report of the 2nd OECD Workshop on Pesticide Risk Indicators: Braunschweig, Germany, 1-3 June 1999.

- Different indicators provide different results, and this can be seen as either positive or negative. One view is that this provides richer information than could be obtained from using just one indicator, and can promote improved decision making. For example, it can lower the chance of missing important changes in the underlying data (two ‘alarm systems’ are better than one). Another view is that the conflicting results of different indicators might be difficult to communicate, and might reduce confidence in the indicators’ reliability.

Validation of indicators

- It would be valuable to validate indicators, but this might not be possible in the classic sense. Monitoring should be done to check indicator results.

- Apart from being consistent with regulatory risk assessment, indicators should complement (and not contradict) other risk reduction policy tools.

Hidden benefits of indicators

- Working with indicators can increase knowledge about how pesticides are used, and can lead to better coordination among agencies within individual countries or to creation of better databases.

Pilot Testing Phase

The participants at the Braunschweig workshop agreed to continue refinement of the OECD Aquatic Risk Indicators in an expanded pilot testing phase, with these purposes:

- To give more countries and more people the opportunity to work with and test the aquatic risk indicators

- To test the responsiveness of the indicators to different sets of uses and different kinds of use data

- To compare the results of the new indicators directly to those from existing national pesticide risk indicators

Seven countries committed to pilot the OECD indicators with their national use data. The pilot team met for the first time in June 2000 for demonstrations of the risk indicator software and to discuss the scope and focus of their respective pilots. In the fall of 2000 each participating country designed their own pilot test program and compiled the necessary data.

In November 2000 Germany led a workshop for several interested countries on the technique of Gutsche and Rossberg for estimating site-specific use data from national sales data. By January 2001Volkmar Gutsche of Germany had completed extensive revisions to the original software developed for the Expert Group, making it easily portable, and distributed it with complete user documentation to all pilot participants. The pilot participants met again in April 2001 to share preliminary results, and finally in October 2001 to plan the final report of the overall aquatic risk indicator project.

Table 3a below presents the scope of the national pilot tests, showing the indicators tested, the scope of hazard data considered, and the number of specific pesticides included in the test. In the "Indicators Tested" columns "R" indicates REXTOX, "A" indicates ADSCOR, and "S" indicates SYSCOR. Parenthetical entries indicate less than full testing of that aspect of the pilot indicators.

Table 3a: Indicators Tested, Hazard Scope, and Number of Pesticides

Country	Indicators Tested				Hazard Scope: Acute/ Long Term	Number of Pesticides in Test
	R	A	S	Other		
OECD Expert Group	X	X	X		A / L	307
Denmark	X	X		Frequency of Application Load Index	A	106
France	X	X	X		A	30 to 70, depending on year
Germany	X	X		SYNOPS	A	Top 15 Herbicides & Fungicides Top 12 Insecticides
Japan		X	X		A	85
Norway	X	X	X	Norwegian Aquatic Risk Indicator (NARI)	A	56
Switzerland	X	X	(X)		A (L)	150

Table 3b summarizes additional characteristics of the national pilot tests.

Table 3b: Use and Time Scope of ARI Pilot Testing

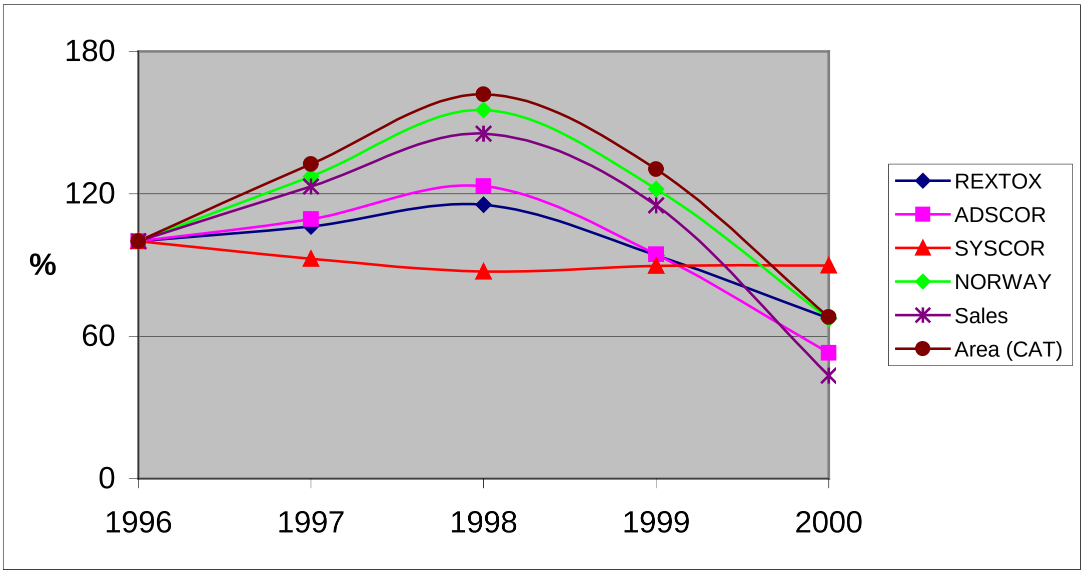
Country	Use Data	Crop Range	Time-Span	Geog. Scope (portion of ag. land)
OECD Expert Group	Actual use data; and back-estimate from sales	Arable crops	1977, 82, 88, 90, 92, 94, 96	England & Wales (large)
		Orchard crops	1983, 87, 92, 96	
Denmark	Est. from sales	Arable crops	1992-2000	National (large)
France	Survey not extrapolated	All crops	1996, 97, 98	Regional (small)
Germany	Est. from sales	Arable crops	1987, 94, 98	National (large)
Japan	Raw sales data only	Rice	1989-1998	National (large)
Norway	Est. from sales	All major crops	1996-2000	National (large)
Switzerland	Survey extrapolated	Arable crops	1997-1998	3 Lake Basins (88% these basins; 25% national area)

The final reports of each national pilot test program are compiled in the Technical Report of the OECD Pesticide Aquatic Risk Indicator Pilot Project. They contain a wealth of information of potential use to anyone considering working with indicators.

Among the salient findings in the pilot testing program are the following:

- The simpler national indicators used in Denmark and Norway and the more complex OECD indicators showed the same general trends in aggregate aquatic risk over time, because they are largely driven by the same variables—area treated, amount of pesticide, and toxicity. Figure 2 below, taken from the Norwegian report, illustrates the great similarity of the trends calculated by REXTOX, ADSCOR, and the Norwegian NARI indicator, and demonstrates as well how closely they follow changes in annual sales volume and in cumulative area treated with pesticides.

Figure 2: Normalised Results of Four Scaled Indicators, Sales Volume, and Cumulative Area Treated (CAT)



Danish testing also showed similar trends from multiple indicators, as summarized in Table 4 on the next page. Results are shown for the Danish Frequency of Application indicator, the Danish Load Index for fish, *Daphnia*, and algae, and for each of the three OECD indicators, again separately for fish, *Daphnia*, and algae. Each indicator was calculated separately for each year, for fungicides, growth regulators, herbicides, and insectides, and for each of the three indicator species.

The direction of changes in the calculated value of each indicator from one year to the next is shown in the table by a '+' for an increasing trend and a '—' for a decreasing trend. Reading across a line in the table, consistency in the direction of change across the columns—i.e., across the ten indicators tested—indicates correspondence in the calculated trend from one year to the next by different indicators. Reading down the columns, a similar sequence of signs from one column to another indicates a parallel shape to long-term trends across different indicators.

Table 4: Indicator agreement on trends

This Danish analysis includes every type of pesticide
and shows increasing (+) and decreasing (-) values of each indicator from one year to the next

Years		Frequency of Application (FA)	Load Index (LI)			REXTOX			ADSCOR		
			Fish	Daphnia	Algae	Fish	Daphnia	Algae	Fish	Daphnia	Algae
Fungicides											
1992	1993	-	-	+	-	-	+	-	-	-	-
1993	1994	-	-	+	-	-	+	-	+	+	-
1994	1995	+	-	-	+	-	-	+	-	-	-
1995	1996	-	-	-	-	-	-	-	-	-	+
1996	1997	+	+	+	+	-	-	-	+	+	-
1997	1998	+	+	+	-	+	+	+	+	+	+
1998	1999	+	-	+	+	-	+	-	-	+	+
1999	2000	-	-	-	-	-	-	-	-	-	-
Growth Regulators											
1992	1993	-	+	+	-	+	+	-	-	-	-
1993	1994	-	-	-	-	-	-	-	-	-	-
1994	1995	+	+	+	+	+	+	+	+	+	+
1995	1996	-	-	-	-	-	-	-	-	-	-
1996	1997	+	+	+	+	+	+	+	+	+	+
1997	1998	+	+	+	+	+	+	+	+	+	+
1998	1999	+	+	+	-	-	+	-	-	+	-
1999	2000	-	-	-	-	-	-	-	-	-	-
Herbicides											
1992	1993	-	+	+	+	-	-	-	+	-	-
1993	1994	-	+	+	+	+	+	+	+	-	+
1994	1995	+	+	+	+	+	+	+	+	+	-
1995	1996	-	-	+	+	-	+	+	-	+	+
1996	1997	+	+	+	+	+	+	+	+	+	+
1997	1998	+	-	+	-	-	+	-	-	+	-
1998	1999	+	-	-	-	-	-	-	-	-	-
1999	2000	-	+	+	-	-	+	-	+	+	-
Insecticides											
1992	1993	-	+	+	-	+	-	-	-	-	-
1993	1994	-	-	-	-	-	-	-	-	+	+
1994	1995	+	+	+	+	+	+	+	+	+	+
1995	1996	-	-	-	-	-	-	-	-	-	-
1996	1997	+	+	+	+	+	+	+	+	+	+
1997	1998	+	-	-	-	-	-	-	-	-	-
1998	1999	+	+	+	+	-	-	-	-	+	+
1999	2000	-	-	-	-	-	-	-	-	-	-

- One of the characteristics that distinguishes the OECD indicators from the Danish and Norwegian national indicators is their much greater reliance on specific use data. But in nearly all the pilot testing, actual use data has not been available, and use values for the indicators have been estimated. Simplifying assumptions required by the estimating techniques used have resulted in much smoother distribution of pesticides, dose rates, and treated areas than would be expected in actual use data, with the consequence that the potential for the OECD indicators to provide insights into risks resulting from specific uses has been masked.

- Relatively simple aggregate indicators like the Danish Frequency of Application index and Load Index or the Norwegian NARI may prove adequate for policy use, especially if actual use data is not available. But greater complexity is needed if an indicator is used for more narrowly focused scientific assessments of, for example, the risk impact of alternative regulatory strategies or pest management practices. This suggests that the paths of future development for policy and scientific indicators may diverge, and reinforces the lesson learned at Braunschweig, that the design and complexity of an indicator should follow from a clear understanding of its intended audience and purpose

- The SYSCOR indicator did not follow the typical pattern. Although the software could be executed without difficulty, the results suggest some errors in the internal penalty tables. The three figures below, taken from the Japanese report, illustrate an experience shared by several pilot participants. Figure 3a shows the risk trend for *Daphnia* in the *unscaled* ADSCOR indicator, not incorporating the area treated. Figure 3b shows the trend in the *scaled* ADSCOR indicator, incorporating area treated. Figure 3c shows the trend in the SYSCOR indicator, which is always scaled. The similarity of the unscaled ADSCOR and the (scaled) SYSCOR trends shows that the tested version of SYSCOR does not respond plausibly to changes in treated area. Revisions to the SYSCOR software are being developed by the French team, and will be incorporated into the OECD software suite when they become available.

Figure 3a: Unscaled ADSCOR Risk Trend for Daphnia

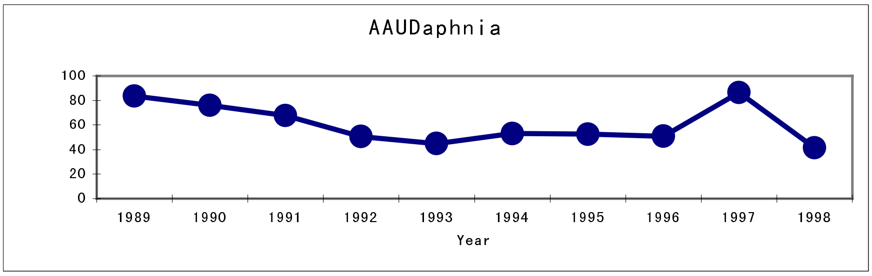


Figure 3b: Scaled ADSCOR Risk Trend for Daphnia

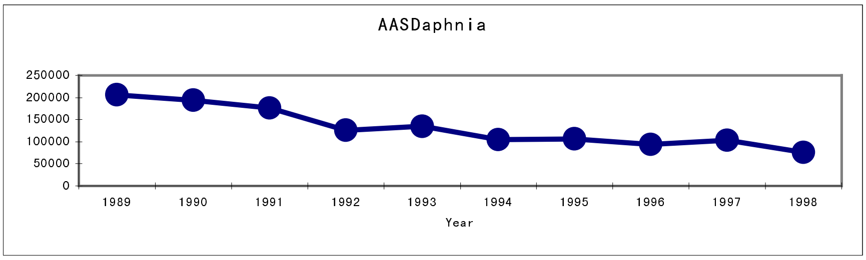
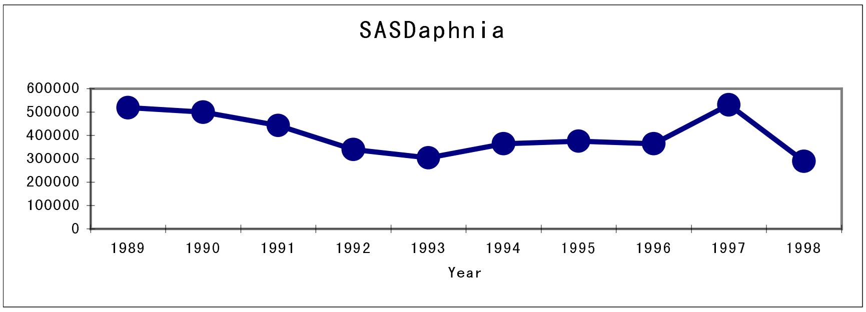
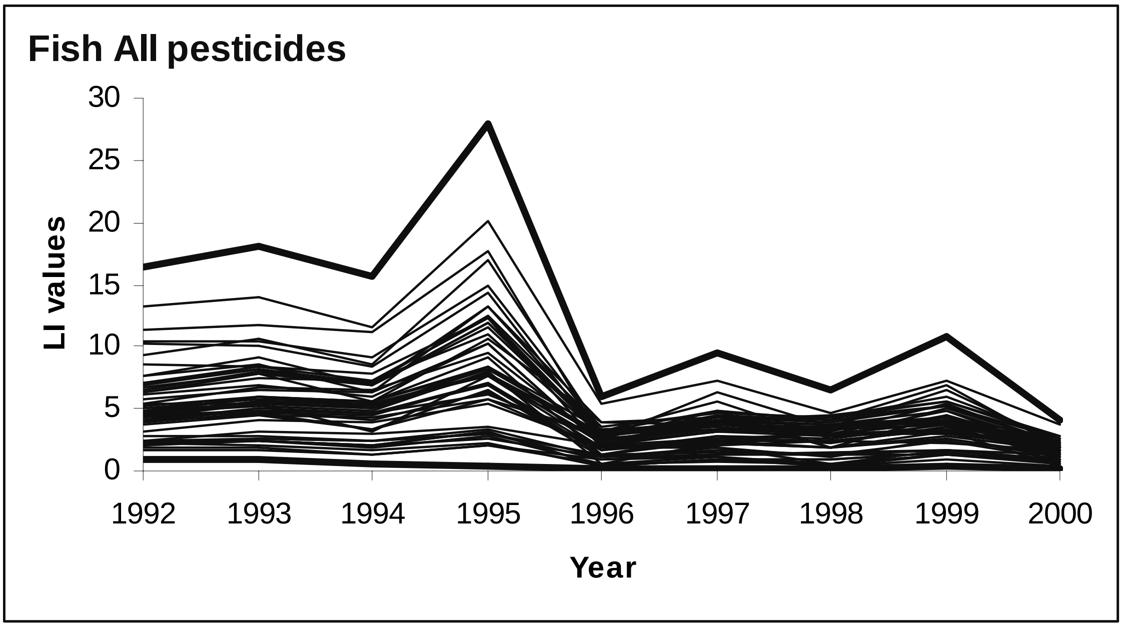


Figure 3c: (Scaled) SYSCOR Risk Trend for Daphnia



- When more than one value for a variable needed for the indicators is available, it is not obvious which should be selected for use. Testing has shown that while the choice of data can significantly affect the calculated value of an indicator, so long as you use the ***same*** rules for selection of data throughout an analysis, the choice of "worst case", "best case", or "most likely" data points doesn't seem to affect the shape of calculated trends. Figure 4 below, from the Danish report, demonstrates this clearly. Each of the 50 fine lines in the figure depicts the trend for risk to fish from all pesticides from 1992 to 2000, based on a different selection of toxicity data. The bold lines show the extremes of the distribution.

Figure 4: Similarity of Trends Calculated Using Differing Toxicity Data
Danish Load Index Indicator



Some pilot participants suggested that the best choice of data for use in indicators would be the preferred values selected for each pesticide in the registration process. Because not all countries select and record preferred values, this isn't always possible. On the other hand the pilot results suggest that the choice of data doesn't much affect calculated trends so long as consistent rules for selecting data are applied throughout an analysis.

- Sometimes the problem is still too little data, not too much. Although the set of data required to calculate acute risk indices is relatively complete, there remain significant gaps, especially for chronic toxicity data. There may still be a need for joint efforts to fill gaps in the supporting database, and until these critical gaps are filled there is little merit in calculating the long-term indicators.

  Table 5 illustrates the relative completeness of the acute database and the prevalence of gaps in the long-term database, as calculated in the Danish pilot testing program.

Table 5: Distribution of Data Gaps by Variable

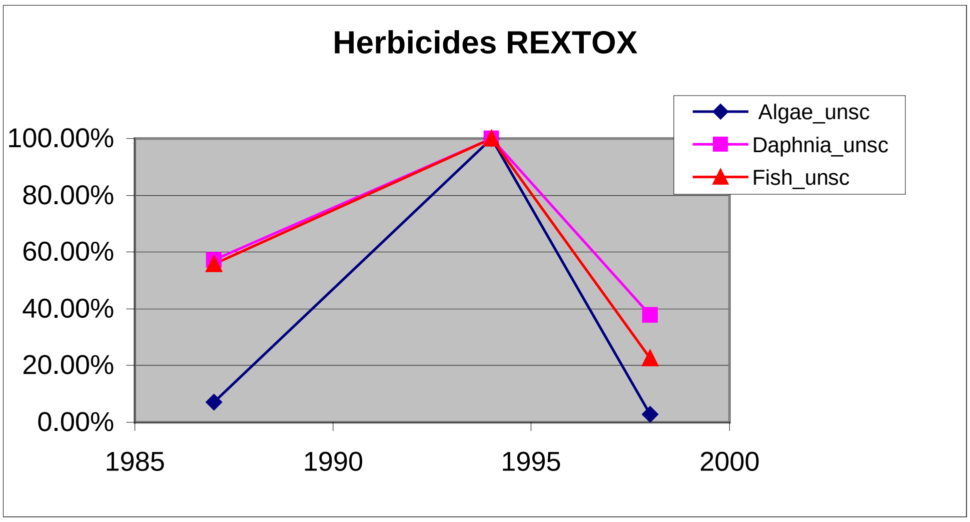
Class	Variable	Percent of Pesticides with Data Gap
Fate	DT_{50}	6%
	K_d	11%
Acute Toxicity	Fish	8%
	Daphnia	12%
	Algae	14%
Long-Term Toxicity	Fish	78%
	Daphnia	65%
	Algae	76

- The more years for which data are available the stronger and more confident the trends shown by the indicators. With the indicators and data in the pilot, it required at least five years' data to establish and confirm a robust trend at high levels of aggregation. The number of years required varies by indicator; in the Danish pilot the simplest indicator required the most years to establish a significant trend.

- Even though trends may be robust in terms of their direction, great care is needed in interpreting the numerical values of these calculated indices. Their absolute values are essentially meaningless, driven by the choice of data used, and reflecting unitless sums of numbers derived from the scores assigned to each use and exposure factor. Even the REXTOX indicator value, derived from true values of variables and scientifically based mechanistic models, is difficult to interpret at high levels of aggregation.

- If fewer than all pesticides are included in the analysis, care is needed in choosing the subset of pesticides to include. German pilot testing included the top 15 herbicides (by overall sales volume) in 1987, 1994, and 1998. When the REXTOX indicator was calculated for these herbicides, the results looked like Figure 5 below.

Figure 5: Aquatic Risk from Top 15 Herbicides 1987-1998



What is the explanation for the sharp peak in 1994? Analysis of the three subsets of herbicides used showed that two pesticides, phenmedipham and bifenox, both highly toxic to aquatic organisms and especially to algae, were included among the top 15 only in 1994. Had a larger set of herbicides been selected such that these two were included in all three periods, the relatively modest changes in their usage from year to year would not have triggered such dramatic changes in the indicator. This kind of artifact can also be minimized by defining the subset of pesticides as those accounting together for a given (and large) percentage of total use.

- In the scoring indicators the choice of breakpoints is critical. Breakpoints should be reset by each country using the scored indicators—at the national level for the area treated and the actual dose rate and, if regional analyses are performed (which in turn depends on the availability of regional use data), at the regional level as well. Dose rate break points may also need to be reset for different types of pesticides. For best results in the scored indicators, the breakpoints for a scored variable with, for example, three possible values would divide the total of all true values into three roughly equal subgroups.

Summary Conclusions from Pilot Testing

Simple indicators requiring relatively little data may be entirely adequate for policy purposes. More complex indicators may be required to serve other purposes, such as comparison of alternative pest management regimens.

The OECD indicators are likely to give results different from the simple national indicators if actual use data is available and used. Estimating pesticide use from sales data is laborious and time-consuming, and the assumptions required for simplicity reduce the impact of use data on the results of the indicator. Other methods of estimating pesticide use, such as those used in the Swiss pilot to scale up from sample data, can affect the indicators even more, to the point of effectively taking use variables out of the calculation altogether. The additional complexity and cost of the OECD indicators is probably not justifiable in the absence of real data on pesticide use.

While the indicators work, in the sense they produce results, more work is needed to explore better ways to present those results, so that they will be clearly and accurately understood. The development and refinement of presentation tools can be a shared effort among all those working with the OECD indicators.

It remains difficult to interpret the results of the indicators. Much work remains to be done to verify the plausibility of the indicators and to increase confidence in indicator results. For the most part this work must be done nationally, with specific attention to the availability of data for local variables, proper placement of breakpoints in the scored indicators, and the policy concerns of each country.

Status and Next Steps for the OECD Aquatic Risk Indicators

The original charge for the OECD Aquatic Risk Indicator project has been satisfied. A suite of aquatic risk indicators robustly rooted in the relevant science has been developed, and has proven easy to use in the several countries participating in pilot testing. These indicators link hazard and exposure data, as other simpler indicators do, and also link specific use data into the exposure calculations. Apart from the use data, the new indicators consider a range of data drawn from the set required to support registration, and link these data through simplified versions of models used in registration assessment, thereby complementing the tools used for registration assessment.

The new indicators fall short of achieving the full specification for pesticide risk indicators developed at Copenhagen. As cited at the beginning of this paper, this called for indicators to be neither too simple nor too complex, to be easy to use, to provide accurate and meaningful information, and to be sound analytical tools with sophisticated inputs but easily understood ouputs.

Many questions remain, not just about the indicators developed in the ARI project but about indicators in general:

- The ARI project has not resolved the debate over the appropriate degree of complexity in an indicator, although agreement has been reached that indicators for different purposes are likely to require different degrees of complexity.

- The OECD Indicators have proven easy to use in pilot testing, but gathering all the necessary data has proven difficult and expensive.

- It is very difficult to assess the accuracy of indicator results, because since aquatic risk cannot be directly measured there is no standard of comparison. Most testing has relied instead on plausibility testing, seeking plausible explanations for surprising results, testing the responsiveness of the indicators to variations in inputs, and comparing results to the expectations of experts. In general, the REXTOX indicator has produced the most plausible results, followed by ADSCOR. The SYSCOR indicator requires further refinement to produce accurate results.

- Debate over what the indicators mean, and their soundness as analytical tools continues. The Expert Group and the Pilot Participants all agree that the results of indicators should be taken as suggestions for further investigation, rather than as final answers, and that unless they were independently confirmed, the results of the indicators themselves would never provide an adequate basis for action. But the indicators have, on the other hand, identified the pesticides and use patterns contributing the most to aquatic risk, and thereby have provided a valuable analytical tool.

This project has accomplished a great deal, demonstrating that sound, effective, and timely technical work can be done through a multilateral team, that concurrent testing can be coordinated by the participants themselves, and that the tools and knowledge needed to develop, test, refine, and use complex indicators are available in many countries. The basic technical work needed to support simple indicators of aggregate aquatic risk from pesticides is complete. Software to calculate the indicators has been developed and tested, and is easy to install and use (a user guide is available).

With regard to future work of the OECD Working Group on Pesticides on the development of indicators of terrestrial environmental risk from pesticides, much of the data concerning pesticide use and properties compiled for the aquatic indicators as well as lessons learned would be directly relevant.

Project Participants

The Expert Group:

Herbert KOEPP, Chairman, German Federal Biological Research Centre for Agriculture and Forestry (BBA)

John CARLEY, United States Environmental Protection Agency

David Raymond FORNEY, Dupont Agricultural Products Company

Volkmar GUTSCHE, German Federal Biological Research Centre for Agriculture and Forestry (BBA)

Valerie HODGE, Canadian Pest Management Regulatory Agency

Yasuo ISHII, Japanese National Institute of Agro-Environmental Sciences

Jean-Michel JOUANY, French Faculté Medecine et Pharmacie

Enrico KIEFER, Novartis Crop Protection Ltd

Dik. VAN DE MEENT, Dutch National Institute for Public Health and the Environment (RIVM)

Mario NICHELATTI, French Ministère de l'aménagement du territoire et de l'environnement

Takehiko YOKOYAMA, Japanese Ministry of Agriculture, Forestry and Fisheries

Consultants to the Expert Group from the MAFF Central Science Laboratory, United Kingdom

Andy HART, Graham SMITH, Miles THOMAS, David WILKINSON

The Pilot Project:

Marianne BALMER, Swiss Eidg. Forschungsanstalt fur Obst,-Wein- und Gartenbau

John CARLEY, United States Environmental Protection Agency

Silvia FREY, Wasser & Landwirtschaft, Swiss EAWAG

Lene GRAVESEN, Danish Environmental Protection Agency

Volkmar GUTSCHE, German Federal Biological Research Centre for Agriculture and Forestry (BBA)

Flemming MOEHLENBERG, DHI Water and Environment, Denmark

Thomas MOUSSEAU, French Ministère de l'aménagement du territoire et de l'environnement

Elisabeth POITRINEAU, French Ministère de l'aménagement du territoire et de l'environnement

Katsuya SATO, Japanese Agricultural Chemicals Inspection Station

Peter SORENSEN, Danish National Environment Research Institute

Erlend SPIKKERUD, Norwegian Agricultural Inspection Service, Pesticides Section