Current OECD Initiatives and Progress in Endocrine Disrupters Testing and Assessment

Herman B.W.M. Koëter
Organization for Economic Cooperation and Development

Thank you very much for your nice introduction. It is my pleasure to be here again, it is the 5th symposium. I have had a pleasure to be at most of them and I am always very impressed with the large interest of the scientific community in Japan.

I am also told that we do not have that much time, we are running late anyway, so I will try to push my presentation into 15 or 20 minutes from now.

I would like to start to say that internationally the work on endocrine disrupters has been and is very much coordinated. It is not an activity that is taking place in one country or a few. It is widely international.

As you see on this slide here, there is the Inter-Organization Programme for the Sound Management of Chemicals where a number of international organizations, as you can see, here six UN organizations on OECD work together on a number of issues related to chemicals. They decided together that the work on endocrine disruptors should be shared by OECD and WHO to take the lead.

Where do we end up with this activity? On the left hand side, you see generally what the OECD works on, and on the right hand side, you see what IPCS works on and Dr. Meredith from WHO will later explain that part of the work. You also see that harmonization of terminology is on both sides. That is a joint project where we work together on the harmonization of the generic terms used in assessment in general but in endocrine disrupters in particular. We also work on technical terms.

The OECD works the development of new testing matters and the harmonization of the information exchange on testing and assessment and we try to share assessment work. All that work is very pragmatic and practical tools. Up at IPCS, WHO, they identify research gaps, also work on terminology and there is the maintenance of an inventory of research activities that Dr. Meredith will talk about later.

Going back to the OECD, in order to organize this work on endocrine disrupters, we established in 1998 the EDTA. That stands for the Task Force on Endocrine Disrupters Testing and Assessment. You see those two words, Testing and Assessment, are both keys in the work of that group.

The work is directed by our Member countries in OECD and this time I did not show the slides of who are members are. For those of you who are not well aware of OECD, we have 30 Member countries in North America, Europe and Southeast Asia that together determined our work.

The work in OECD is also based on consensus, which can be extremely difficult when you work with scientists to continue the discussions to have full consensus on all the issues. It is, at the same time, our strength because once that consensus has been reached, there usually is a very strong incentive to work according to that incentive.

The task, as you can see here, of that endocrine disrupters work is to manage all the work of the special activity to reach international consensus both on the development of test methods and on the testing and assessment approaches. I will start with the testing and assessment approaches because I think that is probably even more important than the technical progress in certain numbers of tests.

The first time EDTA met, they decided that in working towards development of new tests one needs to know where these tests could be used for, for what particular work and for what particular stage of assessment. This is a bit the chicken and the egg problem. Where do you start? Do you first start to develop tests and then use these tests in the assessment of chemicals? Or would you first determine what kind of information you would like to have in your assessments and based on that, develop your tests?

It is not very clearly either one or the other, so these two aspects go hand in hand. What the EDTA considered is, let us first develop a tool box where all the tests considered to be the tools that you would use, would fit in. You could find them back easily once you needed them. At the time, an initial framework was agreed which started with an initial assessment but then followed up with screening, and followed up with testing.

That testing period idea worked for a while until people started to say, "Yes, but now we have developed this particular screening method and according to the scheme it comes before testing and sometimes we would like to see a doctor testing because the screen could be providing information on metabolism or kinetics or otherwise provide more information." They did not want to see that flaw.

EDTA 4, a few years later, started to discuss that conceptual framework again and they then said, "Well, maybe we will always start with existing information collection of course." And we will use some simple tests and measurements like physical/chemical parameters and other things to sort chemicals, and to come up with those that probably are not such of a problem like polymers or almost that may be a problem.

Then, we have short and long term tests. We do not say screening and testing anymore. Short and long term does not necessarily mean first short and then long term.

But there was no agreement on that because experts started to say, "Yes, but what is a short term?" Is that one week test or four weeks test? Is that still short term?" It is hard to find a color for that.

This June and you cannot read that slide you are not supposed to. I only showed that slide to show you the picture and not the text because it comes later.

In June, EDTA again came together, this time in Tokyo. This time EDTA came together with not only experts from other countries, but also there were NGOs like environmental NGOs and welfare NGOs. Industry was there, there were a lot of people.

They agreed in the end to a conceptual framework that comprises five levels. You can see five bars here. Each of those bars is no longer identified by the kind of test that you would conduct but by the type of information that it would reveal. It also means, you see, there is no connection between the bars. You can just use any bar as you like, you can exit any bar, enter any bar wherever that is possible.

Very briefly, what do these bars now mean? At level 1, we could start it with level 5, but it does not really matter, but it is, and it remains quite obvious that at some point one has to sort out chemicals. There still are a massive amount of chemicals around and the numbers are very, very much, but one would generally agree that some 30,000 chemicals need to be assessed in one way or another. You have to start with sorting and prioritization. That is what at first leveled us.

On the right hand side, you see the sort of components that could be used for that. This is also a work in progress. It does not mean that what you see on the right hand side is going to be a definitive list, absolutely not. More will be added and some will probably be taken away if better methods arise.

Level 2 is a level that provides mechanistic data based on *in vitro* assays. That is what it does. On the right hand side again, you see a number of possibilities and every time when you see tests or assays or screens or whatever on the right hand side it means that there is international agreement that these screens or tests may play an important role. It does not mean that we will all have them already available as finalized and validated screens. It is only in the identification that we should work towards the development of international guidelines for these particular assays.

Level 3 starts with *in vivo* information, but the *in vivo* information is limited to single endocrine mechanisms. On the right hand side, you see for the first time now two boxes. The middle box relates to human health effects, usually work in mammalians, and the right hand side is environmental testing.

Here you see the well known uterotrophic and Hershberger assay that provides information on the single endocrine mechanism and I think it is important because often I hear the criticism that uterotrophic

assay is not a good assay because it does not really tell you anything and you do not know anything about it and so many endocrine disrupters mechanism. That is true, it only provides information on one particular element.

On the right hand side, you see as an example the fish vitellogenin assay where you focus on only one particular thing.

Level 4 is still providing information from life organisms, but it provides information from multiple endocrine mechanisms. As an example on the right hand side, you see a fish screening test that would include not only vitellogenin but also gross morphology with GSI: gonadosomatic index, gonad histology and probably other things that together give a much better picture.

The frog metamorphosis assay is there. As some would say that probably is not true because it is a single endpoint. It is not really a test for the amphibian assessment, it is only a test as a surrogate for mammalian thyroid problems. Others feel that the frog assay still provides more than only that surrogate for the human thyroid system.

In the end, level 5. I say in the end of this because it is the 5th and last light slide on that series, not because this is where one has to end up. This provides *in vivo* information but then on a combination of endocrine disrupters information and other mechanisms for risk assessment. Here we see for instance 1 and the 2-generation tests in rodents, but you see on the right hand side a partial or full life cycle assessment in fish, birds, amphibians, invertebrates and all those other species that we probably have to work on.

With the agreement on that, there were also a couple of notes that were seen as quite essential. I will run through them very briefly.

Note 1: emphasizes that one can enter the scheme at any place, at any level and you can also exit at any level. It all depends on the nature and the existing information needs that you have. So, tailor made approach, case by case approach is key.

Note 2: except for level 5, ecotoxicity, where we have that combined endocrine disrupters and other risk assessment elements. That should also, for the environment, focus on population damage, it can no longer only provide information on one particular species or individual.

Note 3: there is also an important understanding that when you use a multimodal model that covers several endpoints of a single assay, the single endpoint assays are no longer needed in that particular case. They are covered by that multi-model.

Note 4: the assessments of chemicals should be based on a case by case basis, but I said that already.

Note 5: the framework should not be considered as all inclusive, I said that as well. It is a work in progress and new tests will be added. Maybe tomorrow we have to decide if certain tests do not work, they will be deleted. It is just providing the tools that Member countries need in order to do the assessment.

Further than to the work of OECD, once we have that in place we still have to do our testing and assessments and we are active very much in the vision of existing guidelines, Test Guidelines and the development of new Test Guidelines and validation work as appropriate.

I have a long list here of assays that we are currently working on in OECD, the enhancement of Test Guideline 407 which is a 4 week general assay for systemic affects in rodents. Currently validation work is on the way. It is actually experimental work finished, we are in a stage of analyzing the results. Uterotrophic assay is the same story. The experimental work is done, the final analysis is under way and the report is almost finished.

Hershberger is sort of mid way where 1 Phase has been finalized and reported and the 2nd Phase has started in Europe now recently. And North America and Japan have already finished this summer. We worked on fish screening tests where we now finally have an agreement on a protocol and I can tell you that was not easy. To reach agreement we needed two expert meetings, one in London and one in Tokyo.

We needed a validation management group meeting in order to come up with directions on what should be included in that fish screening assay. The fish full life cycle test has a ground plan. We have not started to work yet, we wanted to make sure that the screening test for fish was first on track.

Avian reproduction tests, this is an issue has been extensive discussions on one or 2-generational species selection. More noticeably, workers on the way in the US are almost finished. That will most likely help very much to proceed with this work and to make known finally major progress in that area.

For both amphibian and invertebrate screening and testing expert groups are currently being established. I have slides on details of this, but time does not allow to go into that, so I will keep that probably for discussion later.

Validation work that we conduct in OECD is one of those that is not only labor intensive, it is also not an activity that belongs to part of our core work. This is not the work that we would like to do basically. We would like to develop methods, we would like to develop testing assessment strategies, we do not really like to do the validation work. But sometimes it is needed that OECD does that, certainly in those cases where enormous amounts of money are involved, where the countries that donate that money would really like that to be dealt with at the international level rather than one country or a group of countries taking the lead in that.

To insure that everything goes according to principles that have been set, you see here the Solna principles of 1996. That often refers to where principles for good validation and regulatory acceptance have been discussed and agreed. But also, flexibility and transparency are very important and taken into account, and recently there has been a meeting in Stockholm that refers to that. Validation Management Groups are established. Groups of experts in all kinds of aspects of validation that manage and co-ordinate that work.

These Validation Management Groups, they report to the EDTA and EDTA in its turn reports to the Working Group of National Coordinators of the Test Guidelines Programme, which is the highest level body when it comes to the work on testing and assessment of chemicals. Further down the road, there are more closely related bodies.

We currently have three Validation Management Groups, one for Mammalian Effects established in 1999 that oversees the work of uterotrophic and Hershberger 407 and later 416 and others. Ecotoxicity tests, the VMG-Eco that was established a year later.

We did that intentionally to sort of stagger that a little bit in order to avoid too much work to come all together in a short period of time. The VMG-Eco is focused mostly now on the fish testing and on the amphibian and invertebrates pretty soon.

There is going to be a 3rd VMG specifically for *in vitro* testing. There was an agreement reached by EDTA in June this year and that *in vitro,* it is probably not the right word, probably the better word would be non-animal testing because it would include things like QSARs and *in silico* methods in addition to *in vitro* method.

In all the validation, work we use a series of reference chemicals and I will not go into any detail at this point, but you can imagine that having reference chemicals where every laboratory that conducts a test uses that chemical from the same supplier, from the same source and batch, is crucial. We use that both in the mammalian and the environmental studies all those same chemicals. Here you see an example, a list of chemicals that have been identified, estrogen agonist and antagonist and negative chemicals and also androgen and androgen antagonist that we are using.

In addition to that we have a thyroid agonist and a thyroid toxicant, aromatase inhibitor and 5α-reductase inhibitor that we use at various stages of the work.

Looking at the time again I will move on to assessments. Because that is, in many ways, not more interesting topic because I see development of methods is scientifically very challenging and certainly

when we have heard the previous speaker talking about all those exchanges that may happen in us humans and at particularly low levels and the endpoints that we need to develop to cover all that. We are far from having that fully covered yet. The development of methodology is extremely important.

At the same time, one also wants to start doing assessments and we all know that certain well known chemicals, the TCDDs, the dioxins, furans, and many others that have already been regulated in Member countries, but are also others there that other countries would like to start to do that work.

It can only be done by sharing the work. I mean so much is true that there is no country today that is capable of supporting or carrying the financial burden of going for endocrine disrupters assessment alone. The amount of money involved is tremendous, that is hundreds of millions of dollars involved in that.

There is first of all the vast number of chemicals, there is the time pressure. We do not want to wait until 2020 and say to our grandchildren that we are still not there. The cost of testing as I said is extremely high, the co-ordination of the assessment may save a lot of time. But, on the other hand, there is also a difference in philosophy and priorities in Member countries.

That is not something that we even want to harmonize. It is impossible, countries have different cultures, countries have different social and economical climates that make it very specific the assessment and certainly the risk management of chemicals in the end. We would like to share the work, we do not like to harmonize to the extent that everybody does exactly the same thing.

Currently, there are several testing programs on the way where we can learn from other countries and I just have a couple of examples here. The list is certainly not all inclusive because the list is very long and you yourself just finished a Japanese meeting of a day and a half talking about your own research, and there is so much more going on in the world.

But the things that are more internationalized, you see that here the testing initiatives in Japan and the United States and you will hear much more of that later June this conference again. Also an existing chemicals program like the SIDS in the OECD where we run through existing chemicals with sort of minimal data sets. That is what the word SIDS stands for Screening Information Data Set. Part of that is a reprotoxic screening assay and it would be very helpful to have that information once we start assessing chemicals because for certain chemicals we do have that information and it may help in prioritizing.

The same holds true for the high production volume, HPV, voluntary initiative in the USA where industry on a voluntary basis provides data of all those high production volume chemicals which is in a relatively short period of time.

There is an initiative in Europe to revise EC directors for new and existing chemicals that also has lists of chemicals, has priorities and children's health programs. Not only in the United States but also in the WHO and elsewhere, help in prioritizing and providing information that is very helpful.

What can be shared today? First of all, we can share the development of methods for screening and testing. We do that to an extent, we work together in OECD, but we only work together now on those assays where we have reached international consensus of their usefulness.

There are countries that develop assays because they unilaterally think it is useful for their country. Sharing that information is extremely helpful for other countries, for scientists to learn about those.

Also, the grouping of chemicals of interest. Certain countries have a list of chemicals of concern, others call it the list of chemicals of interest. Other countries have two, three or even more lists of chemicals depending on what ministry you come from. Sharing all these lists will end up in a number of chemicals that are on all the lists and by that very nature may get a high priority.

The co-ordination of testing. As Japan wants to conduct a number of 2-generation studies as it has done recently and it will do in the near future. It is good for other countries to know with what chemicals these tests are being conducted so to avoid duplication. Share those testing results and share assessment reports and the discretion of those assessment reports. Then, we come into the more tricky ground because if

countries set up there assessment reports and come to certain conclusions as respect to their country, we do not want automatically other countries to adopt those same conclusions.

We would like to share that with other countries and to do more than that, we would like to talk about it. We would like countries to provide an opportunity to explain why they did a certain assessment and why they had the outcome that they chose. It will learn and teach. Other countries will learn from it and it may gradually leads to a more common understanding of which chemicals are the bad chemicals and should be regulated.

But rather than to again enforce it, talk about it and see if there is certain common sense or not. By that we will achieve that mutual understanding of national assessments, by exchanging we learn by doing and it is all on a voluntary basis.

It is also important that we all understand that risk assessment and risk management is not only driven by hazard assessment outcomes and by science. It is important, science rules and it is one of the slow ones in OECD, but at the same time as I said earlier there are economical factors and social and cultural factors that makes assessment different from country to country.

To give an example just outside endocrine disrupters to be on safe grounds. If we talk about metals assessment, and we do that nowadays quite extensively, it is very obvious that if your country has a very strong mining activity that is economically one of the pillars of your economy. You probably have a very different view point on how to deal with metals and metal substances, your country is only importing them. And the same holds true for many of those chemicals. Next to the science that rules there is also the policy that rules.

I will show you now at the end of my presentation a few websites because that assessment sharing is not only talking anymore, in June there was an agreement reached on how to approach that. We have these pages now working, active, they have search possibilities there. You can search for testing method development plans, for testing reports, for activities in other Member countries and a list of chemicals of concern from other Member countries.

This is just an example. If you look by chemicals what you may see, lots of what you see in here are mock-ups, are not true reports because this was basically done for a meeting that was held earlier. The same for, if you look at country, the work done by certain countries.

Important though, what I almost forgot to mention, but I want to say is the ICAPO that you see here. ICAPO stands for the International Council on Animal Welfare Policies in OECD. That is a group of 23 animal welfare organizations that now work as NGO in OECD to make sure that the work that we do, at least they try to influence that work to the extent that they could accept that also from an animal welfare viewpoint.

The web-sites of interest are seen here. The chemicals testing, you see that on top, that is a public site. To the bottom, the worksharing pages that I showed a second ago are unfortunately not yet public sites, and we do that intentionally. We want to make sure countries are eager to put there documents up even if they feel that this may trigger discretion. We do not want them to be public documents for now so you need a username and a password.

If you would like to look at those pages, then you have to see your National Co-ordinator. Who is your National Co-ordinator? You have to go that top page where there is general information and it also has information for each country about who the National Co-ordinator in OECD is.

I will end here. I would like to thank you for your attention. I hope there will be time later this afternoon or otherwise later in this week to have some more discussions on where we are specifically in testing and to answer any questions that you may have. Thank you very much.