Validation of Assays for the EDSP

Briefing for EDMVAC

November 30, 2005

(Nov 23 Draft)

Purpose of today's briefing

- Present and discuss EPA's approach to validation of the assays for the EDSP
- Obtain the EDMVAC's comments and perspectives on EPA's validation approach

The EDSP

- A two-tiered screening program
  - Tier I—to identify the potential of chemicals to interact with the estrogen, androgen and thyroid systems.
  - Tier II—to identify and characterize the adverse effects resulting from that interaction and the exposures required to produce them
- Tier I should be composed of both mechanistic in vitro assays and in vivo apical assays. (EDSTAC)

Statutory Requirements for Validation

- EPA must use valid screens and tests in the EDSP (Section 408 (p) of the FFDCA)
- Federal agencies must ensure that new and revised test methods are valid prior to their use (ICCVAM Authorization Act 2000)

What is Validation?

Validation is an assessment of the reliability and relevance of a test method for a particular purpose.

- Relevance
  - –The extent to which a test method will correctly predict or measure the biological effect of interest.
- Reliability
  - –The extent to which a test can be performed reproducibly within and among laboratories over time.

History of Alternative Test Method Validation

- Formalization of process for validation of alternative methods stemmed from efforts to obtain acceptance of non-animal alternatives
- OECD Solna Workshop 1996
- Consensus on validation criteria reached
- Solna principles nearly identical to ICCVAM validation criteria
- Agreement among participants that the criteria should apply to validation of all toxicological methods

Validation Criteria

1. Scientific and regulatory rationale
2. Relationship of endpoints to biological effect or toxicity
3. Formal detailed protocol
4. Assessment of variability
5. Assessment of performance with reference chemicals

Validation Criteria

6. Comparison of the performance of the replacement test to the original test
7. Description of limitations
8. Data quality/use of GLPs
9. Availability of data and independent scientific peer review

OECD Guidance Document 34

- OECD recognized the need to update the Solna principles “ to provide practical guidance” on the validation of test methods. Result is GD 34.
- Stockholm conference in 2002
- Berlin Workshop to deal with definition and role of prediction model and data interpretation procedure in test development and validation

Additional Criteria for Alternative Methods

Post-Solna practitioners of alternative method validation agreed on the following additional criteria for validation:

- An alternative test method consists of two parts: the test system and a prediction model.
- A prediction model is an algorithm for converting in vitro data into a prediction of in vivo toxicity.
- Validation is a test or measure of the performance of the prediction model.
- The prediction model needs to be developed prior to validation to allow a prospective evaluation of the prediction model.
- The set of test chemicals used in validation should be different from the set used for model development.

Validation of Ecotoxicity Test Methods

- New test method reproducibility measured across labs with limited number of chemicals (ring test).
- Relevance is assumed because an environmentally relevant species is selected for testing.
- No prediction model; direct observation of toxicity of interest (e.g., critical life processes).
- Standardized method is based on protocol assessment in ring test rather than prevalidaton.

Flexibility in applying criteria

- The amount and kind of information needed and the criteria applied to a new test method depends on…
  - Scientific and regulatory rationale
  - Type of test (new, replacement)
  - Use of the method (screening, definitive, adjunct)
  - Domain of applicability
  - Relationship to the species of concern
  - Mechanistic basis of the test
  - History of the use of the method within the scientific and regulatory community

OECD GD 34

Validation Realities for the EDSP

- Tier I is for screening, not for prediction of in vivo toxicity
- Battery of Tier I assays—assays compliment each other
- Assays are “new” assays, not replacements of existing screens or tests
- Limited number of reference chemicals available
- Practical limitations regarding numbers of tests that can be run during validation

Application of the Validation Criteria to the EDSP

1. Scientific and regulatory rationale
2. Relationship of endpoints to biological effect or toxicity
3. Formal detailed protocol
4. Assessment of variability
5. Assessment of performance with reference chemicals
6. Comparison of the performance of the replacement test to the original test
7. Description of limitations
8. Data quality/use of GLPs
9. Availability of data and independent scientific peer review

Demonstrating Relevance in the EDSP

- Relevance can be based on three factors:
  - Scientifically accepted theory
  - Empirical demonstration of test performance
  - Direct observation of inherently relevant endpoints
- Contribution of each factor differs according to the assay being validated.

Demonstrating Relevance in the EDSP (2)

- Relevance of EDSP assays rests mainly on theory and direct observation of relevant endpoints.
  - Theoretical basis must be clear and well accepted
  - Less need for empirical proof

Demonstrating Relevance in the EDSP (3)

- Role of empirical data
  - Assess the sensitivity of the assay
  - Assay should correctly detect benchmark chemicals
  - Negative chemicals must demonstrate that the assay can discriminate between positive and negative chemicals
- Reference chemicals
  - Chemicals with historical information and known mode of action (i.e. estrogen receptor agonists, aromatase inhibitors, androgen receptor antagonists, anti- thyroid)
  - Test more in a single lab than across labs
  - Limited number
    - Repeat the same chemicals in prevalidation and interlaboratory studies

Reliability

- Within-test, intralaboratory and interlaboratory variability will be measured for each endpoint
- Assay/endpoint is sufficiently reliable if overall variability is low enough to give a level of sensitivity or power consistent with the purpose of the assay.
- Endpoints shown to be of low reliability will be dropped from the protocol or made optional

Tailoring Validation Studies to Different Types of Assays

- *In vitro* single mode-of-action screens
  - Examples: binding assays, aromatase inhibition
  - Scientific understanding plays a big role in establishing relevance
  - Positive and negative controls used
  - 10 or more chemicals of varying potency (strong, moderate, weak positive and negative) will be tested in interlaboratory studies

Tailoring Validation Studies to Different Types of Assays (2)

- Single mode-of-action *in vivo* screening assays
  - Examples: Uterotrophic, Hershberger
  - Positive and negative controls used
  - Fewer reference chemicals used due to practical considerations (animal welfare, higher costs, etc)
    - 7-10 chemicals of varying potency
  - Key criterion is whether assay detects chemicals that interact by the defined mode of action and not others

Tailoring Validation Studies to Different Types of Assays (3)

- Multiple mode-of-action, *in vivo* screening and definitive assays
  - Examples: All Tier II tests, pubertal assays, adult mammalian screens, fish reproductive screen
  - Only negative (vehicle) controls used
  - For Tier I assays, each basic mode of action will be tested with one or more reference chemicals. One or more negative reference chemicals tested.
  - For Tier II tests, reference chemicals will be chosen to evaluate each endpoint.
  - Special problems with coding chemicals for ecotoxicity testing
  - Tier II tests will be accepted as definitive for risk assessment
  - Validation status of multi-modal assays will be reviewed after data are available on 50-100 chemicals

Peer Review Approaches Considered for Individual Assays

Approach	Pro	Con
Letter review under contract	Fewer resources Fast turnaround	Less transparency Less interaction
Public panel under contract	More transparency More interaction	More resources Slower turnaround
SAP/SAB	More transparency More interaction	More resources Slower turnaround

Peer Review

- An independent peer review panel will be convened to review groups of related assays
- Peer review will most likely be conducted under a contract mechanism
- All reports would go to the peer review panel. Raw data would be available upon request.
- The Tier I battery will be proposed by EPA and reviewed by the SAP/SAB.

Questions asked of EDMVS and EDMVS answers

Issues /Answers

EPA Question:

- Should the primary demonstration of relevance, i.e., the multi-chemical study, be performed during prevalidation rather than during the validation phase?

EDMVS Answer:

- EDMVS endorsed this concept but noted that since some new chemicals should be used in the interlaboratory validation studies, data on relevance would be acquired at this phase too.

Issues /Answers

EPA Question:

- Is three a reasonable minimum number of laboratories to use during validation?

EDMVS Answer:

- In general 3 labs is a reasonable minimum, but the answer to this question is really assay specific. Some may require more and it may be reasonable to have fewer than three in other cases.
- A power analysis should be conducted at the end of prevalidation to assist in determining the optimum number of laboratories.

Issues /Answers

EPA Question:

- How critical is it to include more than one laboratory at the prevalidation stage?

EDMVS Answer:

- In general, it is important to get a sense of the transferability of protocols and variability between laboratories before beginning validation. One needs this information to determine how many laboratories are necessary in validation (see power analysis comment).
- “Validation should be thought of as a confirmation of what one has learned in prevalidation.”

Issues /Answers
asked of EDMVS

EPA Question:

- Is it reasonable for the validation of the Tier 1 screening battery to be a paper exercise in which the performance of assays on a core group of chemicals is compared?

EDMVS Answer:

- This is a reasonable expectation and should be the Agency's goal; however, EPA should be prepared to conduct additional studies in which chemicals are run through the complete Tier 1 battery if validation data on individual assays data do not support a clear determination on the composition of the battery.

Issues for the EDMVAC

Please comment on the following issues

Topic 1: Determination of Relevance

- Relevance can be based on three factors:
  - Scientifically accepted theory
  - Empirical evidence
  - Direct observation of relevant endpoints
- Less empirical evidence is necessary when relevance is grounded in one of the other factors.

Topic 2: Reference Chemicals

- Use only chemicals with high quality data whose mode of action is understood.
  - Others believe chemicals should be representative of domain regardless of the availability of high quality data or expectation of results
- Challenge the assay with carefully selected benchmark chemicals
- Number of chemicals will vary with assay
  - 10-25 for *in vitro* screens
  - 5-10 for in vivo screens
  - 1-3 for in vivo Tier II
  - 10-25% of chemicals will be negatives

Topic 3: Assay Performance

- Determine ability of labs to measure endpoints (e.g., organ or tissue weights, AGD).
- Calculate variability.
- Variability satisfactory if it allows adequate sensitivity or power for the assay to fulfill its intended use.
- Tier I assays judged by performance against benchmark reference chemicals.
- EPA does not plan to compare the predictivity of screens with higher tier tests as part of validation.
- Not enough data for meaningful statistical descriptions of certain performance measures.

Topic 4: Statement Regarding Use of Tier 1

- Tier 1 is used for screening, not for prediction of in vivo toxicity

The Future

- Mid-course review
  - EPA plans to review the results of the first 50-100 chemicals to assess the performance of the battery as a whole.
  - While not a part of validation, the review will provide useful feedback on individual assay performance and may lead to a re-evaluation the need for, modification of, or replacement of an assay within the battery.
- Replacement assays are being developed
  - May need full validation, or
  - May be substituted if they meet the performance criteria for the assays they are replacing