EVALUATING EXPERIMENTS FOR CODE AND CROSS-SECTION VALIDATION FOR CRITICALITY SAFETY

Virginia A. F. Dean, Ph.D., P.E.
Consultant
HC 1 Box 407, Elgin, Arizona 85611-9710
VFDean@aol.com

ABSTRACT

The International Criticality Safety Benchmark Evaluation Project (ICSBEP) began a decade ago for the purposes of preserving and enhancing the value of critical-experiment data for the nuclear criticality-safety analyst. Experiment descriptions and benchmark models are being compiled in a handbook [1] for use by safety analysts, code developers, and cross-section evaluators to validate neutron-transport codes and cross sections that calculate the multiplication factor k_{eff}.

An ICSBEP working-group member presents her views about particulars of the project, noting a few continuing problems and making suggestions for contributors and users.

1. INTRODUCTION

1.1 BACKGROUND

Since the beginning of the nuclear era, large amounts of time, money and effort have been spent performing experiments with fissile materials to determine combinations with other materials that produce steady-state heat generation from fission, i.e. critical systems. The International Criticality Safety Benchmark Evaluation Project (ICSBEP) began a decade ago to preserve and organize the valuable experimental data that defined these critical systems. The usefulness of such data for nuclear criticality safety is as tests of the computer codes and nuclear cross sections used to calculate the safety of fissile-material storage and handling. Therefore, two primary tasks of the project are to derive benchmark models of the critical experiments for creating input files for the neutron-transport codes, and to estimate how close to $k_{eff} = 1$ the codes should reasonably be expected to calculate, based on the uncertainty of the available experiment data. The experiment descriptions, models and evaluated uncertainties are now being compiled in the ICSBEP handbook. [1]

When the ICSBEP project began, many experiments, especially those with neutrons of high energies (fission spectrum) or thermal energies (usually moderated by water), had been

published in various formats and were well known. However many other experiments had not been published. Some of these other experiments had been conducted for private purposes; perhaps publishing results of the experiments had not seemed important at the time. Other experiments were kept secret because of the cold war. Others may not have been published because calculations by commonly used computer-code packages (Monte Carlo or S_n codes employing a particular set of nuclear cross sections) produced results that did not agree with the experiments. The sources of errors causing the discrepancies – whether with the experiments or with the codes or cross sections – were not known.

Over the years, the need to validate codes and cross-section sets for designing fissile-material storage and reprocessing facilities increased, as did international cooperation in nuclear research for peaceful purposes, and public awareness of and demand for nuclear safety. The latter also motivated additional regulations that raised costs of new experiments in the United States. All of these made the timing right for the ICSBEP.

1.2 ICSBEP EVALUATION OF A CRITICAL EXPERIMENT

The description of an experiment and its benchmark model, as documented in the ICSBEP handbook, is called an *evaluation*. The author of the evaluation (the evaluator) has several tasks. He or she should collect all relevant available data about the particular experiment or set of experiments, organize it, and clearly describe it. Another task is to evaluate the data, which includes considering all collected data (some of which may be inconsistent), distilling the benchmark model from it and determining the effect on the multiplication factor, k_{eff}, of model simplifications and data uncertainties. These estimates of the effects of uncertainties and simplifications tells users how close to $k_{eff} = 1$ their code package should calculate the model in order to be in agreement with the critical experiment. Another task of the evaluator is to describe completely and clearly the benchmark model of the experiment (geometry and materials). Users of the handbook can then use the model description to create input files for their codes that calculate k_{eff}. The evaluator also creates sample inputs of the benchmark model for one or more of the widely used criticality-safety code/cross-section packages and reports the sample calculated results.

To insure that the benchmark model rightly represents the experiment, the evaluation is subject to a 3-step review. First, a member of the evaluator's organization reviews the document. During this "internal review", the reviewer compares information in the evaluation with data in the references and checks that the experimental data has been correctly copied and interpreted. He or she also checks the correctness of values derived from experimental data. Second, someone outside of the evaluator's organization performs an "independent review" of the completed evaluation to see that the description of the experiment, the data evaluation, and the model are clear, correct, and complete. The independent reviewer also checks derived values. Third, a few weeks before a meeting of the ICSBEP working group, members review the evaluation ("working-group review"), without detailed checking of numerical values, to confirm that the evaluation is complete and conclusions are reasonable.

As member of the ICSBEP working group, I have been both reviewer and evaluator. The following discussion represents my viewpoint of some of the problems we have encountered in

implementing the above process and attempting to achieve its goals – my lessons learned, from the benefit of experience and hindsight. Some of these problems continue to cause difficulties, though our methods of handling them are improving.

The sections of an ICSBEP evaluation of a critical experiment are summarized below:

Section 1 – Detailed description of the experiment
- 1.1 Overview (background, brief description of configurations and experimentally varied parameters, number of acceptable configurations)
- 1.2 Configurations and experimental method (detailed descriptions of geometry, explanation of procedure to determine critical condition)
- 1.3 Materials of the configurations (densities, chemical analyses)
- 1.4 Supplemental measurements (additional measurements of the experiment that are not relevant to the benchmark model)

Section 2 – Evaluation of the data

Discussion of inconsistencies and missing data (which of inconsistent data is chosen for the model and why)

Calculated effects on k_{eff} of uncertainties in geometry or materials

Conclusion of acceptability, or not, for a benchmark model

Section 3 – Benchmark-model specifications
- 3.1 Simplifications (differences between the proposed model and experiment, with estimated effect of the simplifications on k_{eff})
- 3.2 Geometry (complete description of the dimensions of the model in words and figures)
- 3.3 Materials (atom densities of each component of each material)
- 3.4 Temperature (of experiments, of model)
- 3.5 Expected k_{eff} (may include small bias from simplifications) and its uncertainty (from uncertainties in geometry and material data, measurement of k_{eff}, and bias calculation)

Section 4 – Sample calculated results (noting results that are more than 1% high or low)

Section 5 – References (published references about the experiments)

Appendix A – Input listings for sample calculated results

Other appendices – as needed

2. EXPERIMENT DESCRIPTION

Section 1 of an evaluation should give all known data about the experiment's geometry and materials. This includes descriptions not only of the critical assemblies but also of the surroundings (the room, assembly supports, and measuring instruments). The evaluator usually chooses the level of detail of the description of the surroundings based on how well reflected the assembly core is. Also included in Section 1 are uncertainties and inconsistencies in reported measured data, methods of measurement (such as the chemical analysis method, geometry measurement procedure, determination of the critical configuration), and the identification of data sources (such as published references, quarterly reports of a laboratory, logbooks, personal communications with the experimenter). Basically, Section 1 is an organized presentation of the

raw data relevant to creating a benchmark model that come from the evaluator's research of the experiment.

2.1 EXPERIMENT VS MODEL

A problem initially encountered by some evaluators was confusion of the model with the experiment. The model, rather than the experiment, was mistakenly described in Section 1. (They wondered why they were required to describe the model again in Section 3!) As I knew from my own experience, confusion of model with reality is easily made. We are accustomed to thinking of many things in terms of models rather than in terms of concrete physical reality – so much so that the model often seems more real than physical reality! It is the actual, physical, experimental configuration that should be described in Section 1.

2.2 INCLUDE ALL DATA

It is still a frequent problem that an evaluator will give only enough data about the experiment in Section 1 to support the particular model that he or she has chosen to represent the experiments. Although this is necessary, it is not sufficient for Section 1. The goal for Section 1 is to collect *all data that might be relevant* to develop the best model of the experiments. The conscientious user of the benchmark model will want to know that the experiment was thoroughly researched and that the evaluator was aware of all the data, including contradictory or neglected data, when he made his choices for the benchmark model. Data not used in the model might be helpful to a particular user for a particular purpose – for instance someone who wants to verify that measuring instruments, when added to the model, have negligible effect. Or someone may want to check that aluminum cross sections (in aluminum structure, which is sometimes omitted from the models) do indeed produce negligible effect in the particular configuration. Data not used in the benchmark model might be useful for someone who has acquired new information about the experiments. For example, tank dimensions might be used with a newly analyzed composition of tank material, or shape of fuel-rod end plugs might be used with their rediscovered composition. Therefore, the evaluator should provide all discovered data about the experiments, whether used in the benchmark model or not.

2.3 RESEARCH PRIMARY DATA SOURCES

Another frequent problem is that not all sources of data are researched. Evaluators are encouraged to check all references and especially not to rely solely on secondary references. My experience indicates that sometimes authors of secondary references may not have had access to as much data as the evaluator has, or they may have overlooked something or made errors in interpretation.

2.4 CONTACT THE EXPERIMENTERS

Evaluators are asked to contact the experimenters directly about the experiments. They can then ask the experimenter to clarify things that are not fully described in the references and even to confirm the published data.

Once when I contacted an experimenter concerning an evaluation that I was beginning to write, he immediately informed me that some dimensions in the published reference had later been corrected. Another time, when an evaluator was concerned about the gap between reflector and core and the experimenter was asked about it, the experimenter said that he and the other experimenters had been aware of the large effect of any gap and had taken steps to minimize it. This allowed the evaluator to estimate a smaller uncertainty for the gap width. Another time, when an evaluator asked an experimenter about a measurement method, the experimenter was able to provide more basic measurement data from which more precise benchmark data could be derived. Yet another time, an experimenter advised an evaluator to use refined data in the published reference rather than data from monthly reports, because he remembered that additional measurements were sometimes taken that corrected the data in the monthly reports. Such "private communications" from experimenters should be included in the evaluation.

2.5 STUDY LOGBOOKS AND CHEMICAL ANALYSES

Evaluators should also attempt to locate and then study logbooks and laboratory chemical analysis sheets and include the findings. Sometimes logbook or chemical-analysis data are more complete than published data, allowing more realistic modeling of the experiment (detailed model). The formally reported experiment may, in fact, have been an idealized experiment derived from several experiments. My experience has been that whenever I have contacted the experimenter or consulted a logbook, I have learned something new about the experiments, and often it was not what I had set out to learn. So researching all sources of data has the potential for contributing valuable information.

2.6 INCLUDE ALL EXPERIMENTS OF A SERIES

Another common deficiency of Section 1 is that not all reported experiments of a series are described. Instead, the evaluator will pick typical or simple experiments from the series without saying whether omitted experiments are deficient in some regard, or whether they appear to be perfectly good experiments but he does not have time to include them. Evaluators should tell the extent of their study and their judgment of the other experiments of the series (in Section 1.1 or in Section 2), to save the time and effort of future evaluators and validators, or to direct their attention to the additional, similar experiments.

If possible, it is preferred that all experiments of a series be included. If the same sources of uncertainty apply to all experiments of the series, as is often the case, then the deviation of results of calculations of the configurations by a generally reliable code package can be compared with the evaluated uncertainty as a test of the evaluated uncertainty. [2] It is recommended that the evaluator mention the similarity of individual experiments of a series. This alerts the user not to give excessive credit for agreement or disagreement of a code's calculated results with the expected k_{eff} when the experiments are used for code validation for a criticality-safety analysis, because of the possibility of systematic error in all experiments of the series.

2.7 REVIEWER SHOULD COMPARE DATA IN EVALUATION WITH DATA IN REFERENCES

An all-too-typical deficiency of Section 1 is inaccurate representation of the experiments due to an incomplete "internal review". In my experience as reviewer, it is common for evaluators to misinterpret or entirely miss valuable reported data. The internal reviewer needs to be instructed that an important part of his job is to check that the evaluation correctly represents the references. The only way the internal reviewer can thoroughly do his job is if he is supplied with all the references, and then carefully reads both references and draft evaluation and compares the references with the evaluation. It is suggested that the evaluator advise his internal reviewer to do this. The evaluator can tell his internal reviewer to simply assume that everyone sometimes overlooks data and makes typos and to please try to find the mistakes that he, the evaluator, has made.

2.8 IDENTIFY DERIVED DATA

Another problem is that sometimes derived data, rather than measured experimental data, is included in Section 1 without being identified as derived. An example of derived data is density when it is derived from measured mass and dimensions. An important task of the evaluator is to clearly differentiate between directly measured experimental data and data derived from it. It is recommended that only directly measured experimental data be included in Section 1. Then in later sections of the evaluation, quantities describing the benchmark model can be carefully re-derived from the basic experimental data. Also, users interested in creating their own models of the experiment can then clearly know which is the basic, reported, measured data.

2.9 USE TECHNICAL TERMS CORRECTLY

Reviewers and evaluators should also check for terms that are used incorrectly. For example, in the Supplemental Measurements section (Section 1.4), the evaluator (perhaps quoting the original reference) might state that "reflector savings" was measured, when actually the quantity was extrapolation length. Reflector savings and extrapolation length are often confused because both are small dimensions at the edge of the core, but they are two completely different quantities. Similarly, sometimes it is claimed that "neutron flux" was measured, when actually the measured quantity was the count rate. These misuses of terms should be corrected during the review process.

3. EVALUATING THE DATA

Section 2 is where the evaluator reconciles inconsistent data, suggests and supports assumptions to substitute for any missing data, and calculates the effects on k_{eff} of measurement error (parameter uncertainties).

3.1 EFFECTS OF PARAMETER UNCERTAINTIES ON THE BENCHMARK-MODEL k_{eff}

The purpose of calculating the effect on k_{eff} of uncertainties in measured parameters (such as material densities and compositions, sizes and positions of pieces of the assembly) is to indicate how far from the expected k_{eff} value a code/cross-section package might accurately calculate due to inexactness of values that define the experiment. If poor calculated results with the benchmark model are within the *total uncertainty in the benchmark-model* k_{eff}, then the poor results are less likely to indicate poor cross sections or poor calculating by the code, than merely incorrect input values from the (unavoidable) limitations of the model.

The *benchmark-model* k_{eff} is the k_{eff} expected from a correct calculation of the benchmark model. From another viewpoint, it is the experimental k_{eff} expected from an exact experimental mockup of the benchmark model, as it is described in Section 3 of the evaluation. An exactly critical experiment or model has k_{eff} exactly equal to 1. The k_{eff} value of a slightly subcritical or supercritical experiment or model is not so clearly definable. Also, whenever a proposed model deviates in any way from an actual critical experiment, its k_{eff} value also becomes less well defined.

All too frequently, uncertainties in measurements of dimensions, material densities, compositions, and the critical condition are not provided in the references. Moreover, it is typical that measurement uncertainties, when they are provided, are not well characterized. It is often not stated in the references whether a quoted uncertainty in a measurement is a limitation of the measurement method, a standard deviation of the results of several measurements, the range of measured values, or a rough estimate by the experimenters.

Because the uncertainties are usually not well characterized and because evaluators have different preferences for treating uncertainties, my policy as reviewer has been to accept any method that seems reasonable (after suggesting my opinion of what is best). However, the evaluator should explain his method, so that the reader may understand how the final total uncertainty in the benchmark-model k_{eff} has been determined. Then users of the evaluation can judge for themselves how well the value of the total uncertainty in k_{eff} represents the effect of what is not known about the experiments, and whether or not the value of the uncertainty is sufficiently conservative or sufficiently realistic to meet their programmatic needs.

3.2 CALCULATING THE k_{eff} UNCERTAINTIES

Evaluators are encouraged to use deterministic transport-theory codes, rather than Monte Carlo codes, to calculate effects of uncertainties (parameter variations) whenever possible. If Monte Carlo is used, relatively large variations in the parameters should be calculated, so that their effects will be evident even with Monte Carlo results. The standard deviation of the difference of any two quantities is equal to the square root of the sum of the squared standard deviations of the two quantities. This also applies to the difference between k_{eff} results of two Monte Carlo calculations. Therefore, to obtain a Δk larger than its standard deviation, the parameter variation, Δp_{calc}, must be sufficiently large. This can be achieved by calculating with a large parameter variation, or by running the two Monte Carlo calculations with a very large number of neutrons so that the statistical uncertainty of each Monte Carlo calculation is small.

This is summarized as follows:

$$\Delta p_{calc} = p_2 - p_1$$
$$\Delta k_{calc} = k(p_2) - k(p_1)$$
$$k(p_2) = k_2 \pm \sigma_2 \text{ and } k(p_1) = k_1 \pm \sigma_1, \text{ so}$$
$$\Delta k_{calc} = k_2 - k_1 \pm \sigma_\Delta = k_2 - k_1 \pm \sqrt{\sigma_1^2 + \sigma_2^2}$$

Therefore, the evaluator should choose $p_2 - p_1$ to be large enough so that $|k_2 - k_1| \geq \sqrt{\sigma_1^2 + \sigma_2^2}$. Then the uncertainty in k_{eff} due to Δp_{unc}, which is the reported error in parameter p, is estimated as

$$\Delta k_{unc} = |k_2 - k_1| \times \frac{\Delta p_{unc}}{\Delta p_{calc}}.$$

When deterministic codes are used for sensitivity studies, the evaluator should ensure that the mesh size is small enough that small mesh-size variations do not produce significant changes in results. The inputs for the base case and its variation should be as similar as possible, so that any observed difference in results will be mostly due to the parameter variation.

Because the benchmark model is commonly modified during the evaluation procedure, the question arises whether or not sensitivity studies must use the final benchmark model as the base case. The working group has agreed that it is not necessary to rerun parameter sensitivity studies with the final benchmark model, as long as the base-case model for the sensitivity studies is similar to the benchmark model. This is because the Δk_{eff} effects for small changes are expected to be similar for similar models. Whether the base-case model is similar enough to the benchmark model to be a good basis of comparison is left to the judgment of the evaluator. Evaluators are encouraged to describe any dissimilarities between the base-case model for sensitivity studies and the benchmark model, so that the user may verify that it is sufficiently similar to the benchmark model.

3.3 RANDOM VS SYSTEMATIC ERROR

The ICSBEP working group and evaluators routinely assume that uncertainties in measured parameters occur randomly, and therefore combine the various calculated uncertainties "quadratically" (i.e., as the square root of the sum of the squares of the individual Δk_{eff}'s) to determine the total uncertainty in the benchmark-model k_{eff}. This method of determining the combined uncertainty is incorrect in cases of possible systematic error in measurement. A systematic error, by definition, will be essentially the same in all components which have the error. Examples of systematic errors are uncertainties in fuel composition in fuel rods made from one well-mixed batch of fuel; in masses of fissile components weighed on one scale; in fuel-rod claddings made of the same highly uniform tubing. One error of this type will affect all similar components, and the error in each component will usually affect k_{eff} in the same sense, positive or negative. The effect on k_{eff} of the possible systematic error in all components can, therefore, be included in one Δk calculation. This is in contrast to combining quadratically the effects of the error in each component to determine the resulting uncertainty in k_{eff}.

Random uncertainty, on the other hand, comes from a large number of small uncorrelated perturbing influences. [4] Therefore, the error in each component is expected to be different from the error in a similar component. (For a truly random error and a large population of components, the error distribution over all components shows a 'normal' distribution.) An example of random uncertainty might be uncertainty in fuel-pellet diameter. Another example is uncertainty of pitch, or spacing, of fuel rods in an array due to holes in the lattice grid plates being bigger than the rods they hold. In this case, each rod's variation away from the average value is uncorrelated with the variations of other rods.

3.4 RANDOM UNCERTAINTY IN SPACING BETWEEN FISSILE UNITS

For the case of random pitch uncertainty due to grid-plate holes being wider than each fuel rod of an array of n rods, if Δk is calculated for an increase in pitch by ΔR (the difference in radii of rod and hole), the effect of each rod may be estimated as $\delta k_i = \frac{\Delta k}{n}$. Then the total effect on k_{eff} of the random uncertainty of rod positions, as limited by their holes, is $\sqrt{\sum_i^n \delta k_i^2} = \sqrt{n\left(\frac{\Delta k^2}{n^2}\right)} = \frac{\Delta k}{\sqrt{n}}$. So the effect of a random uncertainty may be estimated by dividing the calculated effect from all pieces having the same variation by the square root of the number of pieces.

3.5 SYSTEMATIC UNCERTAINTY IN SPACING BETWEEN FISSILE UNITS

The previous example of random pitch uncertainty was due to random lateral movement of fuel rods in the holes of their lattice plates. However, there is also a systematic pitch uncertainty, due to the precision with which the grid manufacturer is able to produce grid plates with the desired pitch, or due to the precision with which the experimenters were able to determine the average pitch. For example, holes in grid plates with pitch 1.32 ± 0.005 cm might actually have an average pitch of 1.324 cm. In this case, the Δk effect of a ± 0.005-cm pitch variation for the entire array should not be divided by $\sqrt{n}$ because it applies to the average pitch of the entire array. The effect of the ± 0.005-cm uncertainty may be calculated simply by increasing the pitch by 0.005 cm (or by a larger increase X and then multiplying the Δk by 0.005/X).

Typically, in ICSBEP evaluations the systematic pitch uncertainty has been ignored. However, in a recent evaluation the evaluator gives the design tolerances of grid plates and explains, "The design tolerance means the *uncertainty of the average* of the distances between the centers of neighboring lattice holes. This is to be distinguished from the standard deviation of the off-centering of the individual lattice holes with respect to the ideal lattice." [5] This off-centering of holes describes a second type of random uncertainty in pitch, besides the previously mentioned one due to holes being bigger than rods.

3.6 AN EXAMPLE OF CONSEQUENCES OF SYSTEMATIC UNCERTAINTY

A factual example of the consequences of the difference in effect of systematic versus random uncertainty was an evaluation of experiments of thin, steel-clad, highly enriched UO_2-Al fuel rods at two close pitches. These were experiments performed specifically for the ICSBEP at Kurchatov Institute in Moscow in late summer of 1997. [6] The calculated total effect of parameter uncertainties, assumed to occur randomly and therefore combined quadratically, was ~0.5% of k_{eff}, which is a typical value for ICSBEP evaluations. Calculated k_{eff} results were ~1.00, as expected, except for configurations at the larger pitch with calculated k_{eff} values of ~1.015! Luckily, components of the experiment were still available so that measurements could be rechecked. The problem, as it turned out, was the steel clad. Rather than 0.1 cm thick, new measurements found its thickness to be 0.119 cm. And its density, which had first been reported at as 8.0 ± 0.2 g/cc, was remeasured and found to be 7.86 ± 0.07 g/cc. The new clad dimensions and density, when used for all fuel rods, gave good calculated results for both pitch values (within 0.3% of the expected k_{eff} values). Uncertainties of clad OD, thickness, and density, even though they had bracketed the revised values, had not predicted the possibility of the high k_{eff} values from systematic changes because the uncertainties had been combined as if they were random.

3.7 UNREPORTED MEASUREMENT ERROR

In cases where no error limits are reported for some parameters, the evaluator should estimate an uncertainty. The assumed error can be 5 in the next digit to the right of the least significant digit of the reported parameter value, or the evaluator can estimate a typical or likely uncertainty for that parameter (perhaps from another ICSBEP evaluation). Then this estimated parameter uncertainty can be used to calculate the k_{eff} uncertainty due to possible error in that parameter. This should be done, not only because all measured parameters have associated uncertainties, whether reported or not, but also to demonstrate that the experiment is not highly sensitive to small variations in the particular parameter. The contrived uncertainty should be clearly described in Section 2 of the evaluation so that the reader can interpret the overall, or total, uncertainty of the benchmark-model k_{eff} correctly.

Users of ICSBEP benchmark models are advised to read Sections 1 and 2 carefully to learn what is known about uncertainties and whether any uncertainties are not included in the total uncertainty of the benchmark-model k_{eff}. Sometimes evaluators and reviewers may mistakenly overlook a significant uncertainty or judge its effect to be negligible relative to the total uncertainty. The definition of 'negligible uncertainty in k_{eff}' varies among evaluations; negligible uncertainty may mean ≤0.0005, ≤0.0001, or so small that it does not increase the cumulated uncertainty. Evaluators should define 'negligible uncertainty', or its definition should be obvious. Sometimes many small uncertainties are dismissed as negligible, even though their quadratically combined effect is not negligible. At times some uncertainties are not calculated because the effects of other uncertainties are so large that it is expected that the additional uncertainty will not increase the value of the overall uncertainty. Although the working group tries to ensure that all significant uncertainties have been included, this goal may not have always been achieved.

4. BENCHMARK-MODEL SPECIFICATIONS

4.1 DETAILED MODEL VERSUS SIMPLE MODEL

Section 3 begins, optionally, with an overview of the model, then describes how and why the geometry and materials of the actual experiment are simplified to obtain the benchmark model. Some users prefer a highly detailed model that duplicates the experiment as nearly as possible. These users may be concerned especially with accuracy of cross sections or with unintended effects of simplifications, or may want to verify a code's geometry-handling abilities. Other users want a simple model that is easy to convert to code input, easy to understand and check for errors, easy to modify for parameter-effect studies, and faster running on a computer. Ideally, an evaluator will provide both a detailed and a simplified model, to satisfy the needs of both types of user, though providing two models is not required.

4.2 SIMPLIFICATIONS AND SUBSTITUTIONS

When creating the benchmark model, the evaluator should conform as closely as possible to actual geometry and materials. It is strongly preferred not to substitute one material for another. The main reason for this is that any substitute will change the neutron spectrum, which affects how all components contribute to k_{eff}. Another reason is that using a substitute interferes with testing the code/cross-section package's ability to correctly calculate the actual experiment by introducing something that was not present in the experiment. The calculation then becomes as much a test of one's choice of substitute, as a test of how well the code calculates what was actually present.

One ICSBEP participant [3] has suggested that a substitution may be better than omission of poorly known details whose effect is small, if the substitution is 'as near as possible' to the neglected details. An appropriately chosen substitution can preserve the essential characteristics of the experiment better than omitting details. For example, suppose the experiment is a bare plutonium sphere with the references giving only a vague description of its surroundings. Rather than substitute an increase in plutonium density for room return (which alters the spectrum by substituting fission neutrons for slower room-return neutrons), or calculate a bias using a rough approximation to the room (as I would propose), it was suggested to include a 1-dimensional approximation to the room (i.e., spherical concrete shell) in the benchmark model. (Uncertainty due to departure from verity would also be added.) This would preserve the approximate neutron spectrum of the actual experiment. In fact, completely omitting the room and introducing a bias is guaranteed to alter the spectrum somewhat. In this particular example, the proposed substitution also preserves the simplicity of a 1-dimensional model. Including an approximation to the room also preserves an expected k_{eff} of 1 for the benchmark model, which is convenient to users performing a systematic analysis of critical experiments.

It is arguable whether a shell of concrete is a substitution for room return or a simplification of the room. Either calculating a room-return bias or including a simplified room, assuming the effect on k_{eff} is small and its uncertainty is included, would provide an acceptable benchmark model.

As is true in the room-return example above, sometimes the model is simplified because the exact configuration is not known. In the example, only the concrete room is modeled around a bare core, even when it is known that other apparatus was present in the room. As long as the effects are known to be small, the evaluator may choose not to model the apparatus, perhaps because it was unreported or was thought to be too complicated for a benchmark model. Since details farther from the core have less neutron importance than details in the core or in its closest reflector, such apparatus, supporting structure, or measuring devices outside the assembly are often simply judged to have negligible effect and are not included in the benchmark model. Evidence for a judgment such as this in the form of a calculated result is encouraged.

In cases where a structure is close to or within the core but is judged by the evaluator to be too complicated to be modeled in detail, the evaluator might choose to create a super-detailed model whose calculated k_{eff} is compared with that of the benchmark model. The benchmark model might or might not include a simplification of the structure. The comparison either verifies that the simplification or omission has negligible effect, or it provides a value of a bias which is added to or subtracted from 1, as appropriate, to give the benchmark-model k_{eff}. If details of the structure are not known, then additional uncertainty to account for the unknown details is estimated and included in the total uncertainty of the benchmark-model k_{eff}. At the least, uncertainty equal to the standard deviation of the comparison should be included. This is because the bias or negligible effect is only known to within this uncertainty. It is also recommended that the evaluator who chooses to omit or simplify objects within the core check that the omission or simplification does not distort the neutron spectrum severely.

The recommendation against substitute materials also applies to the common practice of providing an equivalent amount of boron in place of impurities in the fuel of low-enriched fuel rods. The purpose of providing this "boron equivalent" is so that calculations will produce the experimentally determined k_{eff} for the rods in the typical thermal spectrum in which such rods are used. This is especially useful for engineers analyzing operations of power reactors that use the rods so that they can more accurately predict criticality. But the boron probably does not well-represent the spectral effects of the impurities it replaces. It is possible that the boron equivalent also substitutes for differences between model and actual fuel other than presence of impurities, such as incorrect fuel density or incorrect proportion of oxygen in the uranium oxide. Or the boron equivalent might make up for deficiencies in the particular cross-section set used in the calculations.

What, then, is the best benchmark model of an experiment with low-enriched fuel rods for which a boron equivalent is given for impurities? In my opinion, the best outcome would be to chemically analyze the fuel for its impurities and include them in the model, rather than boron. Perhaps second-best would be to use measured impurities for similar fuel. In lieu of either of these, my opinion is that it is preferable to include the calculated effect of the boron equivalent as either a bias or uncertainty in the benchmark-model k_{eff}, rather than to include the boron in the benchmark model.

It has generally been the policy of the working group to omit poorly known details if their effect is small, and calculate and include a bias (and its uncertainty) in the benchmark-model k_{eff}, rather than make substitutions. One advantage of adding a bias instead of substituting is that models

whose k_{eff}'s are different from 1 may be quickly singled out by the user as either not critical or modified in some way. This should lead the user to read the evaluation carefully to better judge the reliability of the benchmark model whose k_{eff} is not exactly equal to 1. Another possible advantage of using a bias rather than substitutions is that evaluators are perhaps less tempted to make large substitutions which would take the model farther away from the actual experiment.

Both users and evaluators should be aware of the possibility that the k_{eff} effects of several substitutions or simplifications or biases might cancel each other so that the cumulative effect on k_{eff} is small, but the effect on the spectrum could, at the same time, be very large. Consequently, the model would no longer well represent the experiment. To extrapolate this practice to its logical (but unreasonable) extreme, any critical experiment with highly enriched uranium could eventually, with enough simplifications and substitutions, be transformed into Godiva!

It is also recognized that the value of the calculated bias is itself, to an unknown extent, dependent on the code and cross sections used to calculate it. This is another reason why it is preferred not to simplify to such a degree that a bias is required. If a bias is required, it should be small and it should contribute additional uncertainty, partly for this reason. Generally, any biases to k_{eff} of the benchmark model should be on the order of ½ % or less.

4.3 NUMBER OF SIGNIFICANT DIGITS OF DERIVED QUANTITIES

The values resulting from any unit conversions (such as from inches to centimeters or from temperature to material density) or values derived from other quantities (such as atom densities derived from masses, volumes and atomic weights) are provided to at least 5 significant digits. Five digits has turned out to be a compromise between evaluators who believe that converted dimensions or derived quantities must be rounded to the same number of significant digits as what was measured and evaluators concerned that rounding of a value derived from measured quantities is a systematic error away from the experiment as defined by its measurements. [7]

Values given in Section 3 are values to be used in code inputs to calculate the benchmark model to test the code and cross sections. A justification for keeping at least 5 significant digits for derived values is to consider them as intermediate values in the calculation of k_{eff} from measured quantities. (The practice of keeping many extra digits while calculating is employed by computers and calculators in order not to introduce additional error from rounding.) The effect of the error from rounding is difficult to estimate because of the complicated computations performed by a neutron-transport code to obtain a value of k_{eff}. The effect of rounding is variable, depending on the algorithms in which the value is used. Keeping extra digits obviates the need to verify that the effect of rounding is indeed small. Keeping 5 or 6 digits in quantities that describe the benchmark model contributes to the goal of creating 'the best' benchmark model derivable from the reported data (rather than creating merely 'any' model of the experiment with dimensions within experimental uncertainties). Minimal rounding of quantities that define the benchmark model ensures that they reproduce the measured experiment and that specifications are at the middle of the uncertainty ranges used to calculate the benchmark-model total k_{eff} uncertainty.

4.4 COMPLETE DESCRIPTION OF THE MODEL

Section 3.2 is a complete description of the model. It is easy to mistakenly omit a dimension or otherwise incompletely describe a configuration. Therefore, one requirement of the independent review is to create a computer input from the model described in Section 3 of the evaluation, to verify that the model description is complete. It is preferred that all information necessary to create the computer input from the model is given concisely in Section 3, rather than by reference to previous sections, for ease of use of the benchmark-model description.

Users prefer that configurations be completely shown in figures. Therefore, figures illustrating the complete benchmark-model geometry are highly recommended.

Section 3.3 lists the materials and their atom number densities in units of atoms/barn-cm to 5 significant digits. If the derivation of the atom density of a particular element is not straightforward (i.e., N = weight fraction×density×Avogadro's number/atomic weight), the method of obtaining the atom densities from data in Section 1 should be explained. It should be clear to the user how all pieces of the benchmark model are obtained from measured data.

Section 3.4 repeats the experimental temperatures from Section 1 (or gives them by reference) and states the temperature of the benchmark model. Any bias or uncertainty in the benchmark-model k_{eff} resulting from uncertainty in the experimental temperature or from not using the given experimental temperature in the model should also be mentioned here.

4.5 THE BENCHMARK-MODEL k_{eff} AND ITS UNCERTAINTY

Section 3.5 summarizes the experimental k_{eff} value and the value of the *benchmark-model k_{eff}* (the desired result of a calculation of the benchmark model) and its "total uncertainty". The *total benchmark-model k_{eff} uncertainty* includes the experimental-k_{eff} measurement uncertainties from Section 1, effects of parameter uncertainties from Section 2, and uncertainties from model simplifications from Section 3. Providing the total uncertainty in the benchmark-model k_{eff} is the ICSBEP method of indicating model accuracy (rather than by the number of digits in model dimension values or atom densities). The total uncertainty of the benchmark-model k_{eff} indicates the expected and acceptable range of calculated k_{eff} results by the code/cross-section package to be validated.

Correctly characterizing the k_{eff} uncertainty is important. As an example, consider users who apply the benchmark models as integral tests of cross sections. If the benchmark-model k_{eff} uncertainty is underestimated, then a calculated result that is outside the uncertainty will erroneously indicate a problem with the code or cross sections. If the benchmark-model k_{eff} uncertainty is overestimated, then the benchmark model will be less useful for finding cross-section errors, since the source of poor results will not be recognized as the cross sections or code.

5. SAMPLE CALCULATION RESULTS

5.1 USE MODELS, NOT SAMPLE INPUTS

Results of sample calculations of the benchmark model described in Section 3 using various code and cross-section packages chosen by the evaluator are presented in Section 4. These calculated results are intended as an indication that the experiments were indeed critical experiments and may also indicate that certain codes and cross section sets can or cannot calculate them well. However, the user is cautioned that the results in Section 4 are not to be used for validation purposes. This is because validation of a neutron-transport code and cross-section set includes the specific computer on which it runs as well as the particular input options chosen by the user in his or her implementation of the benchmark model. Another reason not to regard the sample calculation results too highly is that the sample inputs are not as carefully reviewed as the model description in Section 3. Therefore the sample inputs may contain errors.

There are two reasons why ICSBEP reviewers do not check input listings as carefully as other parts of an evaluation. First, as just mentioned, providing inputs to validate a calculation method (i.e., use of a particular code package on a particular computer by a particular safety analyst) should include, within the validation, the safety analyst's particular choices of code input options. For example, MCNP problems can be set up as universes within other universes, or without defining universes. Another example is CSAS input for processing of cross sections may include specifications of shapes and sizes of structural material in MORE DATA, or it may not. In each case, the safety analyst who creates code input from the benchmark model might reasonably use either option. (It is assumed that the safety analyst will use the same input options in calculations of the configuration whose safety is being analyzed.) Ideally, the input options chosen by the safety analyst and those chosen by the evaluator both give the same result, either because the sets of options are essentially equivalent or because both choose the same best set. But verifying input options is beyond the scope of the ICSBEP review.

A second reason for not checking input listings carefully is that such detailed checking is very time consuming. The main objectives of the ICSBEP are to identify, evaluate, verify, and preserve benchmark-quality data and derive verified benchmark models that can easily be used by criticality safety analysts for validation purposes. Detailed checking of sample input listings does not contribute significantly to these objectives.

For these two reasons, safety analysts are responsible for the inputs to their validation calculations.

5.2 SAMPLE CALCULATED RESULTS

In Section 4, results of the sample calculations of the benchmark model that differ by more than 1% from the benchmark-model k_{eff} are so noted, along with any (optional) opinions or observations about correlations or about possible causes of unexpectedly high or low results. The working group, early in the project, reached consensus that poor calculated sample results should not be the sole reason for rejection of experiments that otherwise seem to be acceptable. This is because a purpose of the ICSBEP is to discover any types of configurations for which the

calculation methods and cross-section sets are insufficient. If the research and evaluation of the experiment data uncover no large uncertainties that could possibly explain poor calculated results, then a reasonable conclusion is that poor calculated results may, indeed, be due to the software package's poor representations of cross sections or neutron transport. (This, of course, assumes proper choices of code input options.)

Appendix A contains the sample input listings of the benchmark model for particular codes. Preferably all inputs whose results are tabulated in Section 4 are included in Appendix A. Any minor discrepancies with the model description in Section 3 are noted in the paragraphs preceding the listings, along with any unusual input options or other notices that the evaluator would like to mention. The usefulness of the input listings might be for learning about various input options, or for running, with or without a modification, for comparison with the reported sample results in Section 4, or as a starting point for creating an input. But, for reasons previously mentioned, they should not be blindly used as inputs for safety-analysis validations of codes and cross sections.

6. CONCLUSIONS

The above is one working-group member's summary of the ICSBEP's experiment-evaluation process. The discussion is intended to be helpful to both users and contributors of ICSBEP evaluations of benchmark models and experiments.

Experiment evaluators are encouraged to thoroughly research all sources of data about the experiment and to include all relevant data in the evaluation. Evaluators are reminded of the importance of the benchmark models for code and cross-section validation. They are encouraged to develop the models carefully and fully analyze and explain their limitations. Users of benchmark models are admonished to read the entire evaluation carefully to become aware of the uncertainties of the model, and to create their own inputs for code validation.

ACKNOWLEDGMENTS

The author wishes to thank her colleagues in the ICSBEP working group for their contributions to her understanding of the benchmark-experiment evaluation process. She would especially like to acknowledge the benefit of email discussions with Dr. Tom G. Williamson, Westinghouse Savannah River, and with Dr. Reuven L. Perel, Racah Institute of Physics, Hebrew University of Jerusalem.

REFERENCES

1. *International Handbook of Evaluated Criticality Safety Benchmark Experiments*, NEA Nuclear Science Committee, Nuclear Energy Agency, Organization for Economic Co-operation and Development, Paris, 1999.

2. T. Ivanova, M. Nikolaev, Y. Rozhikhin, M. Semenov, A. Tsiboulia, "Validation Of The KENO/ABBN-93 Package Based On Data From The International Handbook Of Evaluated Criticality Safety Benchmark Experiments", Proceedings of ICNC'99, The Sixth International Conference on Nuclear Criticality Safety, Versailles, France, Vol. II, pp. 722-730 (September 20-24, 1999).

3. personal communication, Dr. Reuven L. Perel, Racah Institute of Physics, Hebrew University of Jerusalem.

4. Baird, D. C., *Experimentation: An Introduction to Measurement Theory and Experiment Design*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1962).

5. Szatmáry, Z., "The VVER Experiments: Regular And Perturbed Hexagonal Lattices Of Low-Enriched UO_2 Fuel Rods In Light Water, Part 2", LEU-COMP-THERM-036, *International Handbook of Evaluated Criticality Safety Benchmark Experiments*, Vol. IV (1999).

6. Bykov, A. A., A. Yu Gagarinski, V. D. Pavlov, O. E. Zhukov, L. M. Montierth, M. N. Neeley, "Experiments With Clusters Of Square-Pitched Lattices of Highly Enriched (~80% U235) Stainless-Steel-Clad Fuel Rods", HEU-COMP-THERM-011, -012, -013, and -014**,** *International Handbook of Evaluated Criticality Safety Benchmark Experiments*, Vol. II (1999).

7. Atkinson, Kendall E., *An Introduction to Numerical Analysis*, John Wiley & Sons, New York, 1978 (p. 24).