DYNAMIC SIMULATION MODELS: HOW VALID ARE THEY?

Raymond C. Shreckengost

VALIDITY, PURPOSE, AND CONFIDENCE

In a strict sense, the subject of dynamic simulation model validity can be treated thoroughly and quickly: there are no fully valid models because all models are something less than the object, or system, being modeled. For example, millions of people have a conceptual model of the President, but, like fingerprints, no two of these models is exactly the same. Further, none matches precisely every detail of the real system. The same reasoning applies to all types and kinds of models of drug abuse.

In a practical sense, we are concerned with usefulness rather than validity. Does the model serve the purpose for which it was intended? Is it helpful? Thus, the developer's or user's purposes must be kept in mind in evaluating a model's usefulness, or validity. Criticisms of models also should reflect this perspective.

Much depends on the purpose for which the model is developed -- for example, the choice of the level of detail used in the model. Just as micrometers are not used to measure inter-city distance a tenth of a mile, explicit modeling of each household would be equally absurd in a model treating the gross behavior of drug abuse systems. The selection of an appropriate level of detail, problem boundaries, and similar considerations constitute the "art" aspect of dynamic simulation model development.

Validity, or usefulness, lies in the subjective view of the user. We think of models as valid when they can be used with confidence. So, this paper focuses on how we can gain confidence in dynamic simulation models. In particular, it considers confidence or validity tests as they relate to a particular dynamic simulation method, System Dynamics (Forrester 1961, 1975; Robbers 1978).
These tests, however, are equally valid for other simulation techniques. They have evolved from nearly 30 years' experience of the inventor of System Dynamics, Jay W. Forrester (Bell and Senge 1980 Forrester and Senge 1980).

MODEL STRUCTURE TESTS

Because the foundation for model behavior is the model's structure, the first test in validating a model is whether the structure of the model matches the structure of the system being modeled. Every element of the model should have a real-world counterpart, and every important factor in the real system should be reflected in the model. Although this may seem like a simple, obvious test, it may not be so. For example, descriptions of how all of the structural parts of real systems are tied together rarely exist. More often than not, such descriptions must be based on the concepts, or mental models, of people familiar with the system. Further, important parts of some systems may lie unrecognized prior to modeling. During the development of a model dealing with the effects of heroin imports into the United States, for example, the key factor in the system, the relative abundance of heroin, was not immediately identified (Gardener and Shreckengost this volume). Thus, the art of model building may, at times, entail discovery and invention.

This approach differs strongly from "Let's collect lots of data and then see what they tell us." Structure, like many other System Dynamics model elements, exploits judgment, experience, and intuition. Data plays a secondary role.

Model Parameter Tests

The model's parameter values are a specific area for testing. Parameter values in a model often may be tested in a straightforward manner, e.g., against historical data. However, in dynamic simulation models of social systems the desired data may be unavailable, in an inappropriate form, or incorrect. There may be elements that are not usually quantified, but that are critical to the system being modeled. These elements must be included in the model. If prejudice, for example, is an important element, it must be included in the model, and its relationship to other pertinent parts of the system must be specified quantitatively. Many required parameter values may not exist and must be developed. In the heroin model, data and descriptions relating to heroin's relative abundance were initially absent. On the other hand, some available, apparently reasonable and acceptable, data on heroin imports turned out to be unreasonable and unacceptable when employed in the model. The point is that dynamic simulation model parameter values, from whatever source they may be derived, are subject to a rigorous and demanding environment. These

values contribute significantly to confidence in the model when the specified parameter values are reasonable and consistent with whatever supporting data might exist.

Boundary Adequacy Test

If a model is focused on the heroin system in New York City, it will not generate national behavior. Conversely, a national heroin model is not likely to replicate the behavior of local systems. Model boundaries must match the purpose for which the model is designed, if the model is to be used with confidence: that is, the model must include all of the important factors affecting the behavior of interest. In practice, boundaries tend to shift as the developers' and users' understanding of a problem evolves with the model's development. As model purpose shifts, changes in the model's boundaries may be required.

In many problems, a simple model with limited boundaries may be expanded, or disaggregated, from time to time, as the model is used to address problems in greater detail. When this occurs, careful attention must be given to indirect effects, which may not be obvious. Suppose, for example, a model treating United States heroin users as a homogeneous group is disaggregated to identify users of small, medium, and large amounts. This will change the user boundaries, and associated changes will be needed in the consumption boundaries.

If the model boundaries are improper, or inadequate, the model's validity is degraded. However, criticism of dynamic simulation models aimed at boundary issues frequently reflects different notions about the model's intended use or purpose. For example, criticism of the user boundary in a model treating users as a homogenous group may ignore the fact that the grouping is consistent with the purpose of the model. But, as explained above, if the purpose is to account for different classes of users, a boundary change is required to account for the change in purpose.

Extreme Conditions Test

A less obvious test relating to model structure involves the effects of extreme conditions. The ability of a model to function properly under extreme conditions contributes to its utility as a policy evaluation tool as well as user confidence. Testing to extreme conditions may easily be overlooked or brushed aside in the hectic environment of early model

development. Subsequently, this oversight may degrade model performance: subtly under normal conditions and significantly when the model is used to answer "What if?" questions that fall outside the operating regions emphasized in early development.

Again, the heroin import model provides a good example. In the past, heroin imports have been, roughly, 5 metric tons per year. They have not fallen to zero, nor have they soared to 10 or 20 tons. Consequently, during the model's development parameter values covered the range of import variations that were of immediate interest, say, 3 to 7 tons. If these initial values only were retained in later versions of the model, the model would show a residual, sizable user population even if imports were reduced to zero. At the other extreme, the number of users would reach an understated upper limit in the presence of a very large heroin supply surplus. The point is that model validity is enhanced if the region within which the model was originally designed to operate is extended so the model generates plausible behavior conditions outside the initial region. For example, the user population should be zero when imports are zero.

Tests under extreme conditions may also expose structural faults or inadequacies and incomplete or erroneous parameter values.

MODEL BEHAVIOR TESTS

Behavior Replication Test

The tests relating to model behavior are less technical and, for many users, more appealing and convincing than the structural tests. Foremost among these tests is the comparison of model behavior with the behavior of the system being modeled. A model whose behavior has little, or nothing, in common with that of the system of interest generates little, or no, confidence.

Where historical time series data are available, the model must be capable of producing similar data. That is, if the model's initial conditions are matched to the state of the system being modeled at some time in the past, the model's behavior should parallel the historical data from that time to the present. In this test, it is again important to keep in mind the purpose of the model -- including the time span of the areas of behavior that are of interest. Further, judgment must be exercised about how closely the model's behavior should match the historical data, since historical data are less than perfect and, sometimes, far from

perfect. It is not at all uncommon for; models to illuminate erroneous data. Where
historical data are very poor or nonexistent, the test may be one of reasonableness.

Given the imports over 10 years or so, the heroin model (Gardiner and Shreckengost, this
volume) generates heroin purity and price values that match well with the historical data for
these parameters. Further, it also produces heroin-related death figures that match the
historical data closely. The closeness of the model-historical correspondence is quite
surprising, given the difficulties inherent in collecting and processing the data that the
historical time series represents.

Purity, price, and deaths can be defined and measured with relative ease compared to the
heroin user population. The model generates user population values against a strict, limited
definition of a heroin user. Here, no parallel historical data exist, and the test becomes one
of reasonableness considering the purpose for which the model was developed.
Subsequent to its initial development, this sector of the model has been detailed to
accommodate users with different consumption habits and varying responses to the
abundance of heroin. Although this is intuitively more satisfying, there is still no
opportunity for a confirming historical test.

Anomalous Behavior Test

When model behavior does not replicate the behavior of the real system, model structure,
parameter values, boundaries, or similar factors are suspect. Something may have been
omitted, improperly specified, or assigned incorrect values. In addition to being a
powerful tool during model development, tests of anomalous behavior may contribute
convincingly to model validity. For example, if a model behaves well except for, say, a
limited period of time, and no faults can be found in the model, the error may lie in the data
with which the model behavior is being compared. Or, matching the real system's
purported behavior may require the inclusion of implausible structure, or parameter values,
in the model. In the heroin model, the import data for 1 year were revised downward,
because the consumption required to match that import level could be achieved only by an
unrealistic increase in heroin user population. Whether due to faults in the model or in the
real system, the resolution of the discrepancies found through the anomalous behavior test
bolsters confidence and validity.

Behavior Sensitivity Test

Most, but certainly not all, social systems are stable -- bureaucracies, in particular, are frequently lampooned for their very, very stable behavior. Small, reasonable changes in a model's parameter values, then, should normally not produce radical behavior changes. If the model's behavior is not seriously affected by plausible parameter variations, confidence in the model is increased. On the other hand, dynamic simulation models are often used to search for parameters that can effect behavior changes. The criterion in the sensitivity test is that any sensitivity exhibited by the model should not only be plausible, but also consistent with observed, or likely, behavior in the real system.

Behavior Prediction Test

Dynamic simulation models are especially useful in predicting how a system would behave if various policies of interest were implemented. Dynamic simulation models offer significant advantages when used in this role; they provide a consistent basis for the predictions. This basis is a consolidation of judgment, experience, and intuition that has been tested against historical evidence, and the predicted effects of implementing alternative policies are promptly available. Confidence in the model is reinforced if the model not only replicates long-term historical behavior, but also responds similarly to existing systems in which various policies have been implemented. For example, over the years many treatment policies have been followed in drug abuse treatment centers. A generic model of such a system, tailored to match any particular center of interest, should replicate the effects produced by the policies implemented in that center.

Family Member Test

Dynamic simulation models acquire added value and confidence when they are generic, i.e., applicable to a family of similar situations, as in the case of treatment centers mentioned above. Drug abuse treatment centers have common basic features, so any one facility may be thought of as a particular case of the basic model embodying these common features. The same is true of payroll, retirement, university, village, city, region, and many other social systems or organizations.

Under these conditions, confidence is enhanced not only because the complementary systems can contribute to the robustness of the model developed for a particular member of

the family, but also because the differences among the members can be explicitly identified and defined.

Some family member applications of the heroin model, for example, are readily apparent. The structure is equally applicable to subdivisions of the United States, such as regions or cities. Further, it appears that it is also directly applicable for cocaine, and, possibly, other illegal drug systems.

Behavioral Boundary Test

Exploiting generic models, behavior prediction, and tests of extreme policies may impinge on the model boundary. Is the boundary still adequate for excursions that may extend beyond the region of operation initially envisioned for the model? In prediction, for example, the basic model may have to be revised, so that policy alternatives, or events, such as the impact of discoveries in research programs, can be introduced. In drug abuse models, the inclusion of social trends, or new domestic or foreign policies, may require boundary modifications. The behavioral boundary test is an important step in determining whether the model includes the necessary modifications.

OTHER TESTS

A third class of test -- policy implication tests -- which includes system improvement, changed behavior prediction, boundary adequacy, and policy sensitivity tests, deals with whether a real system's response to a policy change would replicate the response to the policy change predicted by a model. These tests reflect a different perspective in the application of some of the tests discussed earlier. For example, if real system behavior improves as predicted when tested in a model, was the policy change responsible for the improvement, or were other factors responsible?

This test builds confidence only after numerous real life tests have been completed. The boundary question is inverted: how would boundary changes alter the evaluation of policies and the selection of policies for implementation? These tests tend to be long term and to contribute to confidence and validity most importantly by enlarging the scope of congruence between dynamic simulation models and the systems they represent.

Checking the dimensional consistency of model equations is an additional structural test that may be ignored as trivial, or obvious, but at some peril. For example, if a model contains an equation with heroin expressed in grams on the left side of the equal sign, heroin in grams, and heroin in grams only, must fall out from the right side of the equation. Errors in dimensional consistency can easily creep into model equations during model development and, subsequently, during revisions.

Additional behavioral tests, surprise behavior, and extreme policy behavior can also contribute to confidence and validity. Surprise behavior relates to the recognition of behavior in the real system that was there all along, but not noticed until the system was modeled. Because of its emphasis on identifying the causes underlying observed behavior, System Dynamics readily leads to such discoveries. For example, in the heroin system model, such a surprise was the identification of the relative abundance of heroin as a key parameter influencing the purity, price, and heroin-related deaths that occurred in the real system. In retrospect, like many inventions and discoveries, the relationships may seem very obvious. Such new-found perspectives, of course, contribute significantly to confidence in the model. Extreme policy tests introduce radical policies into the model to see if the behavior of the model is consistent with what would be expected under these conditions. This helps affirm the model's robustness.

COMMON TESTS NOT USED

Paralleling the development of the tests described above has been a growing body of evidence and opinion that many tests commonly associated with model testing are inappropriate, inadequate, or even dysfunctional. In part, these changes derive from the philosophy underlying the System Dynamics method of dynamic simulation modeling, particularly, the notion that all important factors in the real system exerting an influence on the behavior of the system must appear in the model -- whether these factors are normally modeled or not. Further, all factors in the model must have a counterpart in the real system. Together with the dynamic, rather than static, nature of the simulation, these characteristics have shifted emphasis from more traditional, statistical tests to the kinds of tests described in this paper -- whole model tests that engage all the model variables and their relationships in the testing process.

The t-test, for example, has been shown to be of little use, and possibly misleading, in several studies (Johnson 1980, Mass and Senge 1978).

Briefly, the tests can lead to the exclusion of factors that are important to a model's behavior. Although the tests may be helpful in detecting structural flaws, they are insufficient in the absence of whole model tests. Recently, statistical tests employing Kalman filtering principles have been developed. These tests may be more useful in the development of dynamic simulation models (Peterson 1979). The greater power of these tests stems from their ability to eliminate the effects of measurement error in hypothesis testing.

REFERENCES[1]

Bell, J.A., and Senge, P.M. Enhancing refutability. In: Legasto, A.A. Jr.; Forrester, M.W.; and Lyneis, T.M., eds. System Dynamics. New York: Elsevier North-Holland, 1980. pp. 61-73.

Forrester, Jay W., 1961, Industrial Dynamics. 464 pages, ISBN 0-915299-88-7 Item #XINDDY-315, $50.00 Paper, Productivity Press Inc., PO Box 13390, Portland, Oregon 97213-0390, Phone: 1/800/394-6868 or 503-235-0600.

Forrester, Jay W., 1975, Collected Papers of Jay W. Forrester. 284 pages, lSBN 0-262-06065-5 Item #XJFORR-412. $65.00, Productivity Press Inc., PO Box 13390, Portland, Oregon 97213-0390, Phone: 1/800/394-6868 or 503-235-0600.

Forrester, J.W., and Senge, P.M., Tests for building confidence in System Dynamics models. In: Legasto, A.A. Jr.; Forrester, J.W. and Lyneis, T.M. eds. System Dynamics. New York: Elsevier North-Holland, 1980. pp. 209-228.

Johnson, C.B., Some effects of data error on econometric models. In: Legasto, A.A. Jr.; Forrester, J.W. and Lyneis, T.M. eds. System Dynamics. New York: Elsevier North-Holland, 1980. pp. 143-159.

Mass, N.J., and Senge, P.M. Alternative tests for the selection of model variables. IEE Systems, Man and Cybernetics, 8 (6):450--459, 1978.

Peterson, D.W. statistical tools for System Dynamics. IN: Randers, J., ed. Element of the System Dynamics Method. 1979, pp. 344,1 SBN 0-915299-39-9. Item #XELEM-

---

[1] Some addresses in this reference section were updated by Nan Lux, System Dynamics Group, 10/94.

412. $50.00 paper, Productivity Press Inc., PO Box 13390, Portland, Oregon
97213-0390, Phone: 1/800/394-6868 or 503-235-0600.

Roberts, E.B., ed. <u>Managerial Applications of System Dynamics</u>. pp. 562, ISBN 0-
915299-59-3. Item #XMNGAP-412. $25.00 paper, Productivity Press Inc., PO
Box 13390, Portland, Oregon 97213-0390, Phone: 1/800/394-6868 or 503-235-
0600.

ACKNOWLEDGMENT

COPYRIGHT STATUS

Self-Report Methods of Estimating Drug Use:
Meeting Current Challenges to Validity

Editors:
Beatrice A. Rouse, Ph.D.
Nicholas J. Kozel, M.S.
Louise G. Richards, Ph.D.
Divislon of Epidemiology and Statistical Analysis
National Institute on Drug Abuse

N I DA Research Monograph 57
1985

DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
Alcohol Drug Abuse and Mental Health Administration