SUGGESTED SPECIFICATIONS FOR A STANDARDIZED ADDICTION SEVERITY INDEX DATABASE

By

***Deni Carise, PhD[1], A. Thomas McLellan, PhD[1, 2], John Cacciola, PhD [2], Meghan Love[1], Terry Cook[2], Greg Bovasso, Ph.D.[2], & Van Lam[1]**

From
The Treatment Research Institute[1] and
The Center for Studies of Addiction[2]
at
The University of Pennsylvania,
Philadelphia, PA

*Correspondence to
Deni Carise, PhD
Treatment Research Institute
600 Public Ledger Building
150 South Independence Mall West
Philadelphia, PA 19106
USA
dcarise@tresearch.org

Acknowledgments:
Supported by grants from the Office of National Drug Control Policy (ONDCP), the National Institute on Drug Abuse (NIDA), and the Center for Substance Abuse Treatment (CSAT).

Suggested Specifications for a Standardized Addiction Severity Index Database

Abstract

The Addiction Severity Index (ASI) has become one of the most widely used instruments in the addictions field. As a result of its wide popularity, there are multiple versions of the instrument in use, and a wide range of computer systems used to collect and/or store ASI data. Thus, it has been difficult for different users and systems to share ASI data. This difficulty significantly reduces the value of the information for treatment providers, policy makers, and researchers.

This article provides operational definitions and specifications for a **"**Standard ASI Database." Descriptions for standard variable names, data types, field lengths, value labels, range checks, and programming notes for all items in the fifth edition of the ASI are available electronically from the senior. Examples from the full protocol and the rationale for producing the Standard ASI Database elements are illustrated here. It is hoped that the format suggested will become the "industry standard" for ASI data storage among all users of the ASI and that, regardless of the software used or the method of data collection, there will be a single, standard format for all ASI databases. The potential applications from such a database would benefit treatment providers/clinicians and researchers as well as payers and policy makers.

1. Introduction

The Addiction Severity Index or ASI (McLellan, Luborsky, O'Brien, & Woody, 1980) has been used in clinical research studies for over 20 years (McLellan et al., 1992). It has been shown to be reliable and valid, when correctly administered, in a wide range of clinical populations and treatment settings (McLellan et al., 1980; 1985; Kosten, Rounsaville, & Kleber, 1985; Leonhard, Mulvey, Gastfriend, & Schwartz, 2000). Perhaps for these reasons, there has been increased use of the ASI in clinical decision-making, research studies, and treatment outcome evaluations. Moreover, the ASI is being used with many and more varied populations. These include women in welfare-to-work programs, psychiatrically severe substance abusers, parole and probation candidates, drug courts, and Employee Assistance Program populations, as well as those entering traditional addiction treatment programs. The ASI has been used in performance-based contracting of treatment programs by cities and states; as part of the pre-trial evaluations of drug-related offenders in drug courts; as part of the information used for sentencing, and for probation and parole decisions; in the screening and treatment planning of primary care patients thought to have addiction problems; and for comparing outcomes of patient samples treated by specialty programs in state-wide outcome evaluations.

It would seem that this increased use of a "standardized instrument" would result in more, and better information being shared by those in the substance abuse treatment field, especially given the remarkable advances in the science of information collection, transmission, and management. Unfortunately, this is not the case for several reasons. First, although the ASI is probably the most widely used instrument for collecting treatment information, even it was used by fewer than 50% of treatment programs in 1998 (McCarty, 1998). In addition, the substance abuse treatment field has lagged substantially in embracing information management systems (Lamb, Greenlick, & McCarty, 1998). Many treatment programs still do

not have access to computers and lack updated information systems (Carise, 1999). Among those facilities with access to computers and information management software, there is no standard program or system for collecting, storing, or reporting information on patients entering addiction treatments. Indeed, there are now so many different computer systems, employing a still wider range of software, it is not likely that any single software system will become an "industry standard."

We believe the next best alternative to a standard software/hardware system would be a uniform set of conventions for coding and storing ASI information in databases. Such conventions would enable programmers or database managers to convert their existing databases into "standard" ASI databases. This would provide immediate benefits in the form of expanded availability and comparability of information that could be used by researchers, policy makers, and treatment providers alike. For example, such a large database would be valuable to researchers in calculating effect sizes for statistical power estimates, and in meta-analyses. Additionally, use of combined datasets would allow for comparisons of specific sup-groups of treatment seekers such as dually diagnosed individuals, those entering treatment through the criminal justice system, etc. It also has the longer term potential for creating an attractive market for software developers to create useful routines and tools to make the collected information even more helpful for informing clinical, research, and policy questions. Finally, standardization should also assure the buyers of existing software products that data collected by or stored in those products will be compatible with other data systems as well as large normative or reference databases. Importantly, this strategy would allow the benefits of standardized data without forcing conformity in the hardware or software used to collect that data.

With this as background rationale, this article suggests uniform variable names and field specifications for storage of all items in the 5th edition of the ASI. The result is operationally defined as a "Standard ASI5 Database." Note: Space here does not permit a complete listing of all the specifications, and conventions that are suggested for preparing a full standard ASI5 database. The full set of specifications is

available at no charge through e-mail from the senior author at dcarise@tresearch.org or directly from the Treatment Research Institute web site (www.tresearch.org). Through this same medium, we also provide our own Standard ASI5 Database consisting of approximately 17,000 admission records (without identifying information) and a set of additional variables (treatment modality, city/program location, etc) designed for use with the standard database format in SPSS format.

We are aware that there are already many software vendors who have created products to collect ASI information to simplify reporting and to enhance ease of clinical decision-making. The authors of this article have never competed with these companies nor received royalties from any commercial vendor of ASI software. Indeed, we encourage the development of scientifically valid software and technology transfer efforts and hope that these standards will increase these efforts - as has been recommended in the recent Institute of Medicine report on "Bridging the Gap" between research and practice (Lamb et al., 1998).

2. Methods

2.1 Elements defining the "Standard ASI5 Database"

The Standard ASI5 Database format is specified for each question or item in the fifth edition of the ASI (with the exception of the Family History grid), and is defined by 6 elements:

1. Variable/Field Name: ASI Section Initial and Question Number.
2. Description/Variable Attribute: ASI Question Description.
3. Data type: Text/Char, Numeric, Long Date, etc.
4. Field Width/Length.
5. Valid Values, Labels, and Range.
6. Coding Notes: Missing data protocol and programming notes.

These specifications allow uniformity in the storage of ASI data, as well as some standardized basic range and consistency checks to increase the validity of the data. Table 1 lists a *sample* of these elements for a subset of the ASI questions. As previously indicated, the complete data dictionary for the ASI 5th edition is available in electronic format at the Treatment Research Institute web site (www.tresearch.org).

It is expected that each system or program will add their own format for storage of "header items" including such data as name, address, patient identifier, etc. However, in the full ASI5 Data Dictionary we do provide our format for storage of any item on the ASI 5th Edition.

[Insert Table 1 here]

2.2 Rationale for variable names in the "Standard ASI5 Database"

Those familiar with the ASI will notice that we utilized ASI section initials and question numbers to create the variable names - broken into subparts (characters) - to designate each ASI variable. For example, the Standard ASI5 Database v*ariable name* for the ASI question "How many days in the past 30 have you used any cocaine?" is "D8a" on both the ASI instrument, and now the Standard ASI5 database. The first character, "D" designates the ASI problem area from which the question comes ("D" stands for drug/alcohol problem area). The second character, "8," is the specific question number in that problem area, while the third character, "a," identifies the subpart of question 8. Use of cocaine in the past 30 days (D8a) is the first part of the cocaine set of questions, lifetime use of cocaine (D8b) is the second part, and D8c designates the variable name for the usual route of administration for cocaine – the third part of the cocaine set of questions. This coding allows programmers and data analysts immediate knowledge that the three questions (D8a, D8b,

and D8c) are related. Similar conventions are used in other problem areas, though space constraints do not permit detailed description here.

For experienced users of the ASI who are familiar with the content of the ASI questions, a more descriptive name such as “Cocaine30” for “How many days in the past 30 have you used any cocaine?” may have more appeal than “D8a”. However, while the intended audience for these standards includes treatment providers/clinicians, researchers, and policy makers, the direct users and those developing the database will be data analysts, database managers and programmers. These individuals are not likely to be familiar with the substance abuse field, or the Addiction Severity Index. Therefore, we felt it would be most efficient to standardize the ASI database variables by question number, as is more common in programming. We expect that programmers will use variable labels to translate the standard variable names into more descriptive names when providing database reports.

2.3 Storing datasets that have additional items with the Standard ASI-5 variables

Many users of the ASI add extra items to satisfy local needs. Just as the Standard ASI5 Database has not designated a format for storing “header” data, we do not suggest a specified format for storing additional variables. We have decided for the sake of simplicity, not to proscribe a position or mode of storage for these extra items, except that if these non-standard items are stored in the same database, they should be included at the end of, and apart from, the standard ASI items. This will assure that they do not become mistakenly transmitted in any effort to share standard ASI data between users. In addition, the separate storage location of additional items will prevent them from affecting any ASI software programs that may be designed to operate utilizing the Standard ASI5 Database.

2.4 Storage of all types of missing or omitted data

Various codes have been used for storage of "missing" data. Data may be missing from an ASI database for several reasons. First, some users of the ASI may systematically exclude specific questions or sections in the ASI. For example, the director of an educational program for drunk drivers may feel collection of ASI information from the medical, family, or psychiatric sections is outside the scope of their program and may decide to exclude these variables. Second, an item on the ASI simply may not be applicable. This instance always occurs with a follow-up from a previous item, such as consecutive questions asking, "How many months were you incarcerated in your life?" and the follow-up item "How long was your last incarceration?" In this case, if the patient reports never having been incarcerated, the follow-up item: "How long was your last incarceration?" would not be applicable.

Finally, data may be missing because a client declined to answer a question or could not remember the answer to a question. This is not common as most patients answer the majority of the questions on the ASI; however, to further insure collection of valid data, patients are informed that they always have the right to skip any questions they do not want to answer, and the interviewer will simply move on to the next question.

The rate at which an individual client declines to answer questions, or the frequency with which a specific question is skipped by various clients (or interviewers) is interesting and potentially important data and should not be confused with items excluded from an ASI interview or items not applicable. For this reason, we do not want to enter simply a universal "system missing" code for data from questions that are either excluded by the program, not answered by patients, or not applicable to particular patients. Information on why data is missing should be maintained and made available for analysis.

Therefore, the following are examples and suggestions on how to code these items both in collecting and storing the data and how to transform the data for further statistical analysis. We understand that no

single method of collecting and storing missing data will be the best choice for everyone. For example, the use of alphanumeric characters ("N" for not applicable, etc.) would seem most helpful for data entry persons, some statisticians, and would be most similar to current storage procedures, whereas numeric assignment (not applicable = "-98") would seem easier for programmers and database managers. What we have tried to accomplish here is to select the method that would be the most helpful to the majority of users as well as the method least likely to result in mistakes during analysis, particularly with less experienced users.

2.4.1 Storing datasets that do not include all ASI-5 variables

As mentioned above, many users of the ASI do not ask all the ASI questions or include all sections on the ASI. In fact, to accommodate the Veterans Administration (VA), we developed a self-administered version of the ASI that has substantially fewer items than the ASI 5th edition. Parenthetically, this shortened ASI was compared with the full version, and high correlations (.59 to .87) were shown between versions on the alcohol, drug, psychiatric, family, legal, and employment section composite scores (see Rosen, Henson, Finney, & Moos, 2000). Regardless of the number of items excluded from the ASI, the specifications suggested here call for a standard database that includes fields for all ASI 5th edition variables, but with separate codes to identify those variables or questions systematically excluded by the system or at a particular program.

Thus, in a system or program where the 11 questions in the medical section are excluded from ASI administration, the database should not simply eliminate the fields (space) allocated for these variables. Instead, skipping the Medical section should result in codes of "E" in all medical section fields, signifying that the item was excluded and all data pertaining to this item should be excluded from analysis or reports.

The use of an alphanumeric character in the database has several advantages. First, in any field that could maintain a 1-character width for storage (i.e.: common in the ASI; scales of 0-9, 0-5, or yes/no

answers), the use of an alphanumeric such as “E” would allow the maintenance of a 1-character field. This is particularly helpful for programs completing the ASI in paper-and-pencil format and entering the data manually into a database. Having all possible answers share the same “width” or number of characters in the field significantly decreases the likelihood of a mistake in data entry. In the medical section, this format applies to 9 of the 11 questions. Secondly, the entry of a code such as “E” when the question is excluded is more intuitive for data entry and less likely to be confused. Finally, the use of an alphanumeric character decreases the likelihood that less experienced analysts may calculate means or perform other statistical manipulations without accounting for numeric values used to signify the excluded questions.

However, it is clear that for programmers and some analysts and database managers, an obvious numeric code is easier to manipulate. Therefore, when a numeric code is desired, we suggest transformation from the code of “E” in the database to the universal code “-97” to designate those variables excluded from collection by the system. Additionally, if entire sections of the ASI are excluded from collection, a syntax program can be used to automate the entry of the “-97” code in all variable fields for unasked questions (just as the database can be programmed to copy the “E” code into fields for all unasked questions prior to individual data entry). This can be done regardless of the software program used for database storage (e.g., SPSS, SAS, Excel, etc.). This maintains the format of the Standard ASI5 Database, allows easy sharing of information across users, and assures that it will not be necessary for users to ask patients all the questions in the 5th edition of the ASI in order to develop a Standard ASI5 Database.

2.4.2 Storing data when a question is not applicable

As indicated in the *ASI User’s Guide* (Fureman, 1990), one possible reason for missing data is that the question may simply be “not applicable” based on a previous question. This answer is coded with an “N.” For example, on the paper-and-pencil version of the ASI, if an item asking about the route of

administration of cocaine in the Drug Section were skipped because the patient had never used cocaine, the interviewer would enter an “N” for route of administration. The question about route of administration is “not applicable" if the patient just reported that the drug was never used. In an automated version of the ASI interview, the route of administration field would default to “N.”

Again, for programmers and some analysts and database managers, an obvious numeric code may be easier to manipulate. Therefore, when a numeric code is desired, we suggest transformation from the code of “N” in the database to the universal code “-98” to designate those questions that were not applicable.

2.4.3 Storing data when a patient declines to answer or cannot remember

A final reason for missing data occurs when a patient states that he/she does not want to answer the question or doesn’t know or cannot remember the answer to a specific question. The code “X” is used when a patient states they “don’t know,” or “don’t want to answer the question.” In this case, if the code for the route of administration of cocaine in the Drug Section were missing because the client did not want to answer the question, the interviewer would enter an “X.” When a numeric code is desired, we suggest transformation from the code of “X” in the database to the universal code “-99” to designate those questions asked of, but not answered by the patient.

3. Discussion

There are good reasons to utilize a standardized database format with universal variable names, data types, field lengths, value labels, and range checks. Such standardization of the ASI5 database format will promote consistency, data integrity, comparability, and sharing of information. It will also allow many different types of data collection systems and software programs to access and analyze the same information. Thus, our approach has not been to standardize data collection using a particular type of hardware or

software – only a single set of naming conventions, and a single storage format to produce a common database that will be accessible regardless of programming language, server, or software system.

In the spirit of collaboration and sharing, we are happy to offer interested readers copies of our own ASI database in the format specified above (see www.tresearch.org, Addiction Severity Index). In the spirit of standardization, we do not offer it in any other format. Our future plans include making available standardized analytic programs, data checking and cleaning protocols, and norm sets utilizing the Standard ASI Database.

We also hope this article and the standardization efforts included here serve additional functions such as reducing the effort necessary to customize software programs and data analysis protocols. We also hope the creation of a Standard ASI5 Database will increase the already large market for addiction treatment information. In turn, this may instill motivation for the creation of new automated clinical, evaluation, and administrative reports, as well as other software tools that are much needed in the addictions treatment field.

Table 1 – Examples from the ASI Data Dictionary

Variable /Field Name	Description/ Variable Attribute	Data Type	Field Width/ Length	Valid Values, Labels, Range	Coding Notes Missing data & Programming notes
D1a	Days used alcohol past 30 days	Alpha Numeric	3	Positive integer (between 0 and 30) E=Excluded from collection X = Not answered (N not an acceptable code)	If transforming to all numeric database: Store “E” as –97 Store "X" as -99
D1b	Years used alcohol lifetime?	Alpha Numeric	3	Positive integer (0 to 99) calculated age based on DOB E=Excluded from collection X = Not answered (N not an acceptable code)	If transforming to all numeric database: Store “E” as –97 Store "X" as -99
D1c	Route of alcohol administration	Alpha Numeric	3	1 = Oral 2 = Nasal 3 = Smoking 4 = Non-IV injection 5 = IV E=Excluded from collection N = Not applicable X = Not answered	If D1A and D1B both equal 0 then skip question D1C and default code to "N" If transforming to all numeric database: Store “E” as –97 Store “N" as -98 Store "X" as -99
D20	# Times in life treated for drug abuse	Alpha Numeric	3	Positive integer (0 to 99) E=Excluded from collection X = Not answered (N not an acceptable code)	Store "X" as -99 in database Store “E” as –97 Store "X" as -99
D22	# Of treatments for drug detox only	Alpha Numeric	3	Positive integer (less than or equal to answer from D20) E=Excluded from collection X = Not answered N = Not applicable	If D20 equal 0 or "X" then D22 = "N" If transforming to all numeric database: Store “E” as –97 Store “N" as -98 Store "X" as -99

REFERENCES

Carise, D., McLellan, A.T., Gifford, L., & Kleber, H.D. (1999). Developing a national addiction treatment information system: An introduction to The Drug Evaluation Network Study. The Journal of Substance Abuse Treatment, 7(1-2), 67-77.

Fureman, I. (Ed.). (1990). Addiction Severity Index User's Guide: A compilation of instructions, conventions, and suggestions for each item on the Addiction Severity Index. Unpublished Manuscript.

Kosten, T.R., Rounsaville, B.J., & Kleber, H.D. (1985). Concurrent validity of the Addiction Severity Index. Journal of Nervous and Mental Disorders, 171:606-610.

Lamb, S., Greenlick, M. R., & McCarty, D. (Eds.). (1998). Bridging the Gap between Practice and Research. Washington, DC: National Academy Press.

Leonhard, C., Mulvey, K., Gastfriend, D., & Shwartz, M. (2000). The Addiction Severity Index: A field study of internal consistency and validity. Journal of Substance Abuse Treatment, 18, 129-135.

McCarty, D., McGuire, T.G., Harwood, H.J., & Field, T. (1998). Using state information systems for drug abuse services research. American Behavioral Scientist, Vol 41(8): 1090-1106.

McLellan, A.T., Cacciola, J. Kushner, H., Peters, R., Smith, I., & Pettinati, H. (1992) The Fifth Edition of the Addiction Severity Index: Cautions, additions and normative data. Journal of Substance Abuse Treatment, 9(5): 461-480.

McLellan, A.T., Luborsky, L., Cacciola, J., Griffith, J.E., Evans, F., Barr, H., & O'Brien, C.P. (1985). New data from the Addiction Severity Index: Reliability and validity in three centers. Journal of Nervous and Mental Disorders, 173: 412-423.

McLellan A.T., Luborsky, L., O'Brien, C.P., & Woody, G.E. (1980). An improved evaluation instrument for substance abuse patients: The Addiction Severity Index. Journal of Nervous and Mental Disorders, 168: 26-33.

Rosen, C. S., Henson, B. R., Finney, J. W., & Moos, R. H. (2000). Consistency of self-administered and interview-based Addiction Severity Index composite scores. Addiction, 95(3), 419-425.