# Can the Short Index of Problems (SIP) Be Improved? Validity and Reliability of the Three-Month SIP in an Emergency Department Sample*

GEORGE A. KENNA, PH.D., R.PH.,† RICHARD LONGABAUGH, ED.D., ARUNA GOGINENI, PH.D., ROBERT H. WOOLARD, M.D.,† TED D. NIRENBERG, PH.D.,† BRUCE BECKER, M.D., M.P.H.,† P. ALLISON MINUGH, PH.D.,† KATHLEEN CARTY, PH.D.,† PATRICK R. CLIFFORD, PH.D.,† AND KATHRYN KAROLCZUK, B.A.

*Center for Alcohol and Addiction Studies, Brown University, Box G-BH, Providence, Rhode Island 02908*

ABSTRACT. **Objective:** Although the Short Index of Problems (SIP) is often used, little is known about the psychometric properties of the SIP in special populations. The present study seeks to determine the following: (1) whether it is possible to substitute items to enhance the psychometric properties of the SIP and (2) whether the SIP, or improved scale, is as sensitive as the Drinker Inventory of Consequences (DrInC) to assess intervention effectiveness. **Method:** The sample consisted of 404 injured patients who were treated in the Emergency Department (ED) of a major teaching hospital that serves southern New England. Three approaches were used to guide development of the 3-month SIP-R, the potential alternative to the SIP. Cronbach's alpha assessed intrascale reliability; hierarchical multiple regression assessed construct validity; performance of the scales assessing intervention change were compared to the total 3-month DrInC as a function of intervention using analysis of covariance (ANCOVA). **Results:** There was no evidence that changing the current SIP items will significantly improve performance. The 3-month SIP performed as well as the 3-month DrInC-2R in predicting 12-month DrInC scores and in determining intervention change at 12 months. Of the 45 DrInC items, 31 also predicted a difference across intervention groups. **Conclusions:** These results suggest there is no advantage to changing the current SIP items. The 3-month SIP is a psychometrically sound measure for assessing consequences of alcohol consumption in an ED sample and is almost as sensitive to intervention change as the full DrInC. (*J. Stud. Alcohol* 66: 433-437, 2005)

MILLER ET AL. (1995) developed the Drinker Inventory of Consequences (DrInC) to assess alcohol-related problems independent of alcohol consumption and dependence. The 50-item DrInC consists of 45 items that measure self-reported alcohol-related consequences in five domains: interpersonal, intrapersonal, physical, impulse control and social subscales, and a 5-item subscale to check for response acquiescence. The Short Index of Problems (SIP), a 15-item version of the DrInC, was developed for situations in which time was limited (Miller et al., 1995). Although these measures are increasingly used, the SIP remains untested in many special populations (Feinn et al., 2003). Moreover, few, if any, studies have sought to improve the reliability or validity of the SIP or compare the performance of the SIP as an outcome measure to the DrInC. Therefore the aims of this study were to: (1) test the psychometric properties of the 3-month SIP by comparing them with a competing measure, the SIP-R; (2) examine the psychometric properties of both the 3-month SIP and SIP-R in an emergency department (ED) sample and (3) determine whether the SIP and SIP-R are as sensitive to intervention change as the full DrInC.

The original SIP was developed by choosing three items from each of the five theoretical DrInC domains based on the strongest item to subscale correlations (Miller et al., 1995). Recently Blanchard et al. (2003) proposed that the preferred method to shorten the Inventory of Drug Use Consequences (InDUC-2R; Tonigan and Miller, 2002) was to use the 15 items with the strongest item-total correlations, rather than item-subscale correlations. This resulted in the SIP-AD (Blanchard et al., 2003) with at least two items from each of the five subscales overlapping with 12 of the 15 items from the SIP.

Noting the lack of studies on the SIP and the alternate methods used to shorten the longer scales, further research of the SIP's psychometric properties in various samples is vital to assessing its stability across populations (Feinn et al., 2003) as well as determining its sensitivity to treatment.

Received: October 29, 2004. Revision: February 15, 2005.

*This research was supported by National Institute of Alcohol Abuse and Alcoholism grants R01 AA09835-01A2 and R01 AA13709-01.

†Correspondence may be addressed to George A. Kenna at the above address, or via email at: George_Kenna@brown.edu. Robert H. Woolard and Bruce Becker are with the Department of Emergency Medicine, Brown Medical School, Providence, RI. Robert H. Woolard and Ted D. Nirenberg are also with the Injury Prevention Center, Rhode Island Hospital, Providence, RI. Ted D. Nirenberg is also with the Department of Psychiatry and Human Behavior, Brown Medical School, Providence, RI. P. Allison Minugh is also with Datacorps, Providence, RI. Kathleen Carty is with the Boston College Graduate School of Social Work, Boston, MA. Patrick R. Clifford is with the University of Medicine and Dentistry of New Jersey, School of Public Health, Piscataway, NJ.

We used data from a previously published study (Longabaugh et al., 2001) performed in an ED to assess the psychometric properties of these scales. Annually, there are millions of ED visits made, with a substantial proportion known to be associated with alcohol use (Cherpitel, 1994; Cherpitel et al., 2003; Davidson et al., 1997). Assessing alcohol-related consequences is an important outcome measure of ED interventions (Longabaugh et al., 2001).

## Method

The complete sample consisted of 539 injured patients treated in the ED of a major teaching hospital that serves southern New England. A thorough description of the methods, recruitment and interventions used in the study has been published (Longabaugh et al., 2001).

Inclusion criteria consisted of a score of ≥8 on the Alcohol Use Disorders Identification Test (Saunders et al., 1993), self-reported alcohol consumption during the 6 hours prior to injury or testing positive for alcohol at the time of their ED visit. In addition to other measures, patients were administered the Lifetime DrInC at baseline. At 3- and 12-month follow-ups the DrInC-2R assessed past 3- and 12-month alcohol-related consequences, respectively.

Patients were assigned to one of three treatment conditions to be reassessed at a 1-year follow-up: Standard care (SC) consisted of customary care normally expected in an ED, brief intervention (BI), adapted to the ED setting (Nirenberg et al., 1996), and BI with a booster session (BIB) given 7-10 days after the BI. The original study demonstrated that the BIB group reported significantly fewer alcohol-related consequences and alcohol injuries at the 1-year follow-up than the SC group (Longabaugh et al., 2001).

Four approaches were used to guide development of the SIP-R, a proposed alternative brief version of the DrInC for measuring alcohol-related consequences. All DrInC items used were standardized ($z$ scores) on the sample. First, $R^2$ regression was used to find the 15 most robust 3-month DrInC items that predict the 12-month total DrInC score. We hypothesized that these items would represent the most variance possible from 15 items measuring DrInC consequences. Second, using $t$ tests, we examined whether any of the 12-month DrInC items at the 1-year follow-up were significantly different between the SC and BIB groups. We considered these items to be the best outcome measures of pre/post-intervention change. Third, consistent with Miller et al. (1995), we identified the 15 strongest item-subscale correlations from the full DrInC at 3 months. Last, consistent with Blanchard et al. (2003), we examined the 15 items that most correlated with the item-total correlations with the 1-year DrInC. As there is no single best method for item analysis, we utilized a combination of methods, incorporating both statistical and judgmental procedures (Anastasi and Urbina, 1997). Conservatively, items for the SIP-R should be significant across at least two of the four analyses and not be significant for an original SIP item that would take precedence over any alternative item. Intrascale reliability was assessed by Cronbach's alpha. Validity was assessed using confirmatory factor analysis (CFA) on the SIP and SIP-R. Hierarchical multiple regression was used to assess construct validity to compare the associations of the total 45-item 3-month DrInC-2R, and both 3-month SIP and SIP-R on the 12-month DrInC total score. The performance of the 3-month SIP and SIP-R were compared to the 3-month DrInC as a function of intervention by using analysis of covariance (ANCOVA) controlling for baseline scores of these measures.

## Results

The analyses only include cases with complete data at the 3- and 12-month data points ($N = 404$). The majority of the sample were men (76.2%), in their mid to late twenties (mean [SD] = 26.8 [8.6] years), and mostly white (73.5%). This sample did not differ significantly from the original sample (78% men; mean age: 27 [9] years; 72% white). Sample sizes for each group were $n = 142$ (SC), $n = 137$ (BI) and $n = 125$ (BIB). Table 1 displays the SIP items assessing past 3-month alcohol-related consequences (i.e., 3-month SIP), with three of the four different methods used to analyze the DrInC-2R and SIP-R for the same time period. At the 1-year follow-up, no items between SC and BIB groups were significantly different at a meaningful level ($p < .05$); therefore this column was deleted from Table 1. The 31 DrInC items in Table 1 reflect items involved in the remaining methods used to assess potential SIP items. Therefore, 14 DrInC items did not appear as a result of the assessment method.

The forward-regression $R^2$ method ($R^2 = .769$) used the 3-month DrInC items to predict the 12-month DrInC score. This determined that 5 of the 15 items were social subscale items and were significantly associated with the 12-month DrInC score. Each of the remaining domains contained at least two items, though a majority of the items did not overlap with the SIP items. The DrInC items in the next two columns show correlations loading most highly on either the subscale items or total 3-month DrInC items. Twelve of the items constrained to each of the five domain subscales overlap with the original SIP items, with correlations ranging from $r = .64$ to $r = .85$. Nine of 15 items derived from item-total correlations method were consistent with the SIP. The last column displays items chosen for the SIP-R. To decide which items should be incorporated in the SIP-R, we chose items represented across multiple methods of analysis. Considering each of these methods, the SIP-R is unchanged from the SIP except for one item: "I have broken things while drinking or intoxicated," which

TABLE 1. Comparison of methods used to examine the psychometrics of the SIP

| Subscale | DrInC Item | SIP items | $R^2$ method items | Items with highest item-subscale corr. | Items with highest item-total corr. | Proposed SIP-R 5 items |
|---|---|---|---|---|---|---|
| Physical | Sick and vomited after drinking | | Y | | | |
| | I have not eaten properly | Y | | .66 | | Y |
| | Physical health harmed by drinking | Y | Y | .67 | | Y |
| | Physical appearance harmed by drinking | Y | Y | .64 | | Y |
| | I have been hurt, injured or burned | | | | | |
| Intrapersonal | I have been unhappy | Y | | | .77 | Y |
| | I have felt bad about myself | | | | .74 | |
| | I have felt guilty or ashamed | Y | | .82 | .77 | Y |
| | Spiritual or moral life has been harmed | | Y | | | |
| | Not had the kind of life I want | | | .85 | .79 | |
| | Gotten in the way of my growth | Y | Y | .84 | .79 | Y |
| | My personality has changed for the worse | | Y | | .77 | |
| | Lost interest in activities and hobbies | | | | .74 | |
| Interpersonal | Ability to be a good parent harmed | | Y | | | |
| | Marriage or love relationship harmed | | | .78 | .73 | |
| | My family has been hurt by my drinking | Y | | .73 | .76 | Y |
| | Friendship or close relationship damaged | Y | | .80 | .78 | Y |
| | Said harsh or cruel things to someone | | | | .73 | |
| | Damaged social life, popularity or reputation | Y | | | .75 | Y |
| Social | Quality of work has suffered | | Y | | | |
| | Failed to do what is expected of me | Y | Y | .74 | .76 | Y |
| | Money problems because of my drinking | Y | | .76 | .81 | Y |
| | Gotten into trouble because of my drinking | | Y | | | |
| | Spent too much or lost a lot of money | Y | Y | .66 | .73 | Y |
| | Suspended, fired from or left a job or school | | Y | | | |
| Impulse control | Caused me to use other drugs more | | Y | | | |
| | Taken foolish risks | Y | | .66 | | Y |
| | Done impulsive things regretted later | Y | Y | .67 | | Y |
| | Gotten into a physical fight | | | | | |
| | Had an accident while drinking or intoxicated | Y | | | | |
| | Broken things while drinking or intoxicated | | Y | .68 | | Y |

*Notes*: SIP = Short Index of Problems; DrInC = Drinker Inventory of Consequences; corr. = correlations; Y= denotes either an item on the original 3-month SIP or 1 of the top 15 items considered for inclusion in the alternative 3-month SIP-R measure. Items are shortened to fit table format but are consequences of alcohol consumption.

replaced the item "I have had an accident while drinking or intoxicated." Similar analyses were conducted on both the 3-month SIP and SIP-R for the purpose of comparison, noting in advance, however, that there is no evidence to suggest changing the current SIP.

### *Reliability*

Cronbach's coefficient alphas were computed for the 3-month SIP and SIP-R and for each subscale. Internal consistency for both scales was very high (alpha = .95), suggesting some degree of unidimensionality (Cortina, 1993). All of the SIP subscale alphas were in an acceptable range: physical (alpha = .79), interpersonal (alpha = .86), intrapersonal (alpha = .89), social (alpha = .85). The only difference between the SIP and SIP-R was the impulse control subscale in the SIP (alpha = .73) and the SIP-R (alpha = .80) that contained the different item. The average inter-item correlations for both measures was high, $r = .56$ (range: .38-.74).

### *Validity*

Chi-square tests for both five-factor correlated models of the SIP ($\chi^2 = 417.4$, 80 df, $p < .001$) and SIP-R ($\chi^2 = 373.7$, 80 df, $p < .001$) were significant, though nonsignificance is rarely obtained (Schumacker and Lomax, 1996). The normed chi-square (NC) was slightly improved for the SIP-R (4.48) versus the SIP (5.22). Both the SIP (.88) and SIP-R (.89) fell just short of reaching the minimum recommended goodness-of-fit (GFI) cutoff of > 0.90 (Maruyama, 1999), but the confirmatory fit indices (CFI) for both the SIP (.92) and SIP-R (.93) exceeded .90, reflecting good fit of the models to the data (Bentler, 1990). The root mean squared error of approximation (RMSEA) for the SIP (.102) and the SIP-R (.095) both approached a criterion of >.10, indicating a poor fit (MacCallum et al., 1996). There were relatively minor differences between the one-factor models of the SIP ($\chi^2 = 487$, 90 df, $p < .001$; NC = 5.41; GFI = .86; CFI = .91; RMSEA = .096) and SIP-R ($\chi^2 = 488.9$, 90 df, $p < .001$; NC = 5.43; GFI = .86; CFI = .91; RMSEA =

.105). In all, the fit criteria displayed for the five-factor-correlated SIP and SIP-R models showed a slightly improved fit over the one-factor models.

The correlations between the total 3-month SIP ($r = .97$) and SIP-R ($r = .972$) scores with the 3-month DrInC were high, suggesting that these short forms satisfactorily represent the total DrInC score. The total 45-item 3-month DrInC significantly predicted the 12-month DrInC ($F = 461.86$, 1/402 df, $p < .001$; $R^2 = .534$). The 15-item SIP predicted the 12-month DrInC ($F = 418.17$, 1/402 df, $p < .001$; $R^2 = .509$), while the 15-item scale SIP-R also predicted the 12-month DrInC ($F = 426.13$, 1/402 df, $p < .001$; $R^2 = .513$). The SIP-R did not add significantly to the prediction of the DrInC over and above the SIP. The fourteen 3-month DrInC items not included in Table 1 and least associated with the DrInC still predicted the 12-month DrInC total ($F = 326.81$, 1/402 df, $p < .0001$), although they accounted for less variance ($R^2 = .447$).

When assessing sensitivity of the 3-month DrInC across the three intervention groups after controlling for the baseline DrInC score, we found a significant difference between groups ($F = 4.30$, 2/399 df, $p < .02$). Similar analyses found the 3-month SIP (using the baseline SIP score as a covariate) ($F = 3.12$, 2/399 df, $p < .05$) and SIP-R scores ($F = 3.34$, 2/399 df, $p < .04$) were also significant.

## Discussion

Consistent with our study aims, we first sought to improve the SIP by developing a competing scale. Given that the psychometric properties of both the SIP and SIP-R were essentially identical, there was no advantage to further exploring the use of the SIP-R. Despite evidence that the SIP-R explained additional variance over and above the SIP items at the 1-year assessment, other measures of reliability and validity suggest that the SIP performed as well as the SIP-R in this sample. Furthermore the SIP accounted for almost as much variance as the full 45-item DrInC. In comparing the SIP to the SIP-R, ultimately only one item could be considered to be a candidate for replacing any of the original 3-month SIP items. The similarity of the two items in question provided no justification for changing the SIP because of the performance of this specific item.

Second, we examined the psychometric properties of the SIP and the SIP-R in an ED sample. In an adequate sample using CFA, the SIP confirmatory fit indices suggest the scale to be a modestly valid and reliable instrument for measuring alcohol-related problems in an ED setting. A recent study using CFA to validate the psychometrics of the SIP in a treatment sample of 153 problem drinkers reported poor-fit indices (e.g., GFI = .831; CFI = .68; RMSEA = .065) for the five-factor-correlated model examined (Feinn et al., 2003). However, using a larger sample size (Guadagnoli and Velicer, 1988) may have enhanced the CFA fit indices of the SIP reported in that study. Although CFA demonstrated that the five-factor scale did fit the data slightly better than the one-factor scale, there was no significant improvement. Additionally, though Feinn et al. (2003) found evidence of possible higher-order factors when assessing the 3-month SIP, a one-factor model is supported by studies of drug use consequences (e.g., Blanchard et al., 2003; Tonigan and Miller, 2002) and might be preferred as a more parsimonious representation of the data (Harlow and Rose, 1994; Noar, 2003).

Using an ED sample, the 3-month SIP remained a sound measure for assessing consequences of alcohol consumption. Despite utilizing multiple formats to assist in the development of the SIP-R, 14 of the original 15 SIP items remained constant. The high alpha and inter-item correlations suggest that though reliable, the SIP may be unidimensional (Clark and Watson, 1995). Furthermore, an alpha approaching 1.0 suggests the SIP may still be reliable with fewer items (DeVellis, 1991).

The DrInC is obviously a more comprehensive assessment than the SIP, relating consequences of alcohol consumption to amount of alcohol use by patients, and presents a broader opportunity for the therapist to discuss reasons why patients might consider reducing hazardous alcohol consumption. Nonetheless, only 31 of the original 45 DrInC consequences met the criterion that was being tested for across one or more assessment methods, suggesting that not all of the DrInC items may be essential to predict subsequent outcomes. As done with the Alcohol Dependence Scale (Kahler et al., 2003) perhaps this exemplifies the potential for applying a less traditional psychometric method such as Item Response Theory to the DrInC (Lord, 1980). Such an approach may allow exploration of the level of alcohol-related consequences at which specific item responses are likely to be endorsed and how well item responses can discriminate among cases, given their stage of alcohol-related consequences.

Our last aim examined whether the SIP was as sensitive to intervention change as the DrInC. The high correlation and regression between the 3-month SIP and 12-month DrInC suggest that the 3-month SIP is capturing a substantial amount of overlapping variance with the longer measure. Importantly, the 3-month SIP predicted the 12-month DrInC scores about as well as the 3-month DrInC. The SIP and SIP-R also detected differences across intervention groups at the 3-month follow-up, though not as robust as that found with the DrInC.

A major limitation of this study affecting its generalizability is the inclusion of primarily white men from a single site in the sample. In addition, the SIP items were derived from the 3-month DrInC items which do not include the intensity response dimension for 9 of the 15 items. Recommendations from these analyses suggest that more studies aimed at shortening these scales would aid the

alcohol-assessment field but that the sensitivity of DrInC and SIP should be maintained. Finally, continued research is needed on the psychometric properties of both the DrInC and the SIP in diverse and under-represented populations.

## References


ANASTASI, A. AND URBINA, S. Psychological Testing, Seventh Edition, Upper Saddle River, NJ: Prentice Hall, 1997.

BENTLER, P.M. Comparative fit indices in structural models. Psychol. Bull. 107: 238-246, 1990.

BLANCHARD, K.A., MORGENSTERN, J., MORGAN, T.J., LABOUVIE, E.W. AND BUX, D.A. Assessing consequences of substance use: Psychometric properties of the inventory of drug use consequences. Psychol. Addict. Behav. 17: 328-331, 2003.

CLARK, L.A. AND WATSON, D. Constructing validity: Basic issues in objective scale development. Psychol. Assess. 7: 309-319, 1995.

CHERPITEL, C.J. Alcohol consumption and injury in the general population: From a national sample. Drug Alcohol Depend. 34: 217-224, 1994.

CHERPITEL, C.J., BOND, J., YE, Y., BORGES, G., MACDONALD, S., STOCKWELL, T., GIESBRECHT, N. AND CREMONTE, M. Alcohol-related injury in the ER: A cross-national meta-analysis from the Emergency Room Collaborative Alcohol Analysis Project (ERCAAP). J. Stud. Alcohol 64: 641-649, 2003.

CORTINA, J.M. What is coefficient alpha? An examination of theory and applications. J. Appl. Psychol. 78: 98-104, 1993.

DAVIDSON, P., KOZIOL-MCLAIN, J., HARRISON, L., TIMKEN, D. AND LOWENSTEIN, S.R. Intoxicated ED patients: A 5-year follow-up of morbidity and mortality. Ann. Emer. Med. 30: 593-597, 1997.

DEVELLIS, R.F. Scale Development: Theory and Applications, Thousand Oaks, CA: Sage, 1991.

FEINN, R., TENNEN, H. AND KRANZLER, H.R. Psychometric properties of the Short-Index of Problems as a measure of recent alcohol-related problems. Alcsm Clin. Exp. Res. 27: 1436-1441, 2003.

GUADAGNOLI, E. AND VELICER, W.F. Relation of sample size to the stability of component patterns. Psychol. Bull. 103: 265-275, 1988.

HARLOW, L.L. AND ROSE, J.S. Prediction models: Optimal conditions and fit assessment. Paper presented at the annual meeting of the Society for Multivariate Experimental Psychology, Princeton, NJ, 1994.

KAHLER, C.W., STRONG, D.R., HAYAKI, J., RAMSEY, S.E. AND BROWN, R.A. An item response analysis of the Alcohol Dependence Scale in treatment-seeking alcoholics. J. Stud. Alcohol 64: 127- 136, 2003.

LONGABAUGH, R., WOOLARD, R.F., NIRENBERG, T.D., MINUGH, A.P., BECKER, B., CLIFFORD, P.R., CARTY, K,. SPARADEO, F. AND GOGINENI, A. Evaluating the effects of a brief motivational intervention for injured drinkers in the emergency department. J. Stud. Alcohol 62: 806-816, 2001.

LORD, F.M. Applications of Item Response Theory to Practical Testing Problems, Mahwah, NJ: Lawrence Erlbaum, 1980.

MACCALLUM, R.C., BROWNE, M.W. AND SUGAWARA, H.M. Power analysis and determination of sample size for covariance structure modeling. Psychol. Meth. 1: 130-149, 1996.

MARUYAMA, G.M. Basics of Structural Equation Modeling, Thousand Oaks, CA: Sage, 1999.

MILLER, W.R., TONIGAN, J.S. AND LONGABAUGH, R. The Drinker Inventory of Consequences (DrInC): An instrument for assessing adverse consequences of alcohol abuse (Test Manual). NIAAA Project MATCH Monograph Series, Vol. 4, NIH Publication No. 95-3911, Bethesda, MD: Department of Health and Human Services, 1995.

NIRENBERG, T., SPARADEO, F. AND LONGABAUGH, R. Brief Motivational Interventions for Drinking Injured Emergency Department Patients, Providence, RI: Center for Alcohol and Addiction Studies, Brown Medical School, 1996, unpublished manual.

NOAR, S.M. The role of structural equation modeling in scale development. Struct. Equat. Model. 10: 622-647, 2003.

SAUNDERS, J.B., AASLAND, O.G., BABOR, T.F., DE LA FUENTE, J.R. AND GRANT, M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. Addiction 88: 791-804, 1993.

SCHUMACKER, R.E. AND LOMAX, R.G. A Beginner's Guide to Structural Equation Modeling, Mahwah, NJ: Lawrence Erlbaum, 1996.

TONIGAN, J.S. AND MILLER, W.R. The Inventory of Drug Use Consequences (InDUC): Test-retest stability and sensitivity to detect change. Psychol. Addict. Behav. 16: 165-168, 2002.

COPYRIGHT INFORMATION

TITLE: Can the Short Index of Problems (SIP) Be Improved? Validity and Reliability of the Three-Month SIP in an Emergency Department Sample
SOURCE: J Stud Alcohol 66 no3 My 2005
WN: 0512102128016

The magazine publisher is the copyright holder of this article and it is reproduced with permission. Further reproduction of this article in violation of the copyright is prohibited.

Copyright 1982-2005 The H.W. Wilson Company. All rights reserved.