Published in final edited form as:

J Nerv Ment Dis. 2013 January ; 201(1): 56–59. doi:10.1097/NMD.0b013e31827ab210.

Timeline Historical Review of Income and Financial Transactions (THRIFT): A reliable assessment of personal finances

Anne C. Black, Ph.D.¹, Kristin L. Serowik, M.A.¹, Karen M. Ablondi, M.P.H.¹, and Marc I. Rosen, M.D.¹

¹Yale University, New Haven, CT

Abstract

The need for accurate and reliable information about income and resources available to individuals with psychiatric disabilities is critical for assessment of need and evaluation of programs designed to alleviate financial hardship or affect finance allocation. Measurement of finances is ubiquitous in studies of economics, poverty, and social services. However evidence has demonstrated these measures often contain error. We compare the one-week test-retest reliability of income and finance data from 24 adult psychiatric outpatients using assessment-as-usual (AAU) and a new instrument, Timeline Historical Review of Income and Financial Transactions (THRIFT). Reliability estimates obtained with the THRIFT for Income (.77), Expenses (.91), and Debt (.99) domains were significantly better than AAU. Reliability estimates for Balance did not differ. THRIFT reduced measurement error and provided more reliable information than AAU for assessment of personal finances in psychiatric patients receiving Social Security benefits. The instrument also may be useful with other low-income groups.

Keywords

Reliability; Instrument; Income; Financial Management; TLFB

Introduction

Finances received and managed by individuals with disabling psychiatric conditions, who may rely on federal financial benefits, are an important indicator of personal well-being and access to resources, as well as a measure of the impact of programs targeting the needs of this population. Standardized psychosocial assessments routinely include measures of income or financial status in overall evaluations of functioning and treatment planning. Indeed, the measurement of income is ubiquitous across the social sciences, and is fundamental in studies of welfare, poverty, public policy, and related issues. The efficacy of the national welfare program has been evaluated by financial outcomes largely derived by survey data (Hotz and Sholz, 2001); the U.S. Census and other surveys of income have both informed allocation of federal support services, and provided information about program success.

Individually and nationally, the impact of the decisions based upon income survey data from financially disadvantaged persons mandates accurate and reliable measurement of target variables. However, measuring income can be remarkably difficult, and results often are

Anne C. Black is corresponding author. Please address correspondence to VACHS; 950 Campbell Avenue; 116A-4; West Haven, CT 06516 anne.black@yale.edu.

Conflicts of Interest and Source of Funding: All authors declare no conflicts of interest related to the conduct of this study or preparation of the manuscript.

inaccurate. Details about personal finances frequently are derived by retrospective personal account, requiring individuals to recall total income and expenses in various domains over a specified period of time. There is considerable evidence that individuals receiving benefits income have difficulty with basic financial facts that are even simpler than recalling and compiling income and expenses over the preceding month. Lynn et al. (2004) described rates as high as 50% of respondents under-reporting income when data were collected by survey from low-income beneficiaries. Rosen et al. (2007) noted benefits income claimed by a large proportion of surveyed homeless individuals was inaccurate when compared to Social Security Administration payment records. Among low-income respondents, substantial discrepancy was observed in self-reported receipt of benefits services provided two weeks apart (Reichert and Kindelberger, 2000). In that study, having income in the poverty range was positively related to response variance.

Heuristics used by respondents to provide retrospective personal information lead to inaccurate accounts due to systematic cognitive biases and error (Hufford and Shiffman, 2003). Responses may be overly influenced by most recent or most memorable events during the target time period, and therefore may not represent true values (Mathiowetz et al., 2001; Stull et al., 2009). Additionally, respondents may misjudge the timing of a remote event, such as receipt of income or accrual of expense, and erroneously report its occurrence during the target timeframe (telescoping error; Bradburn et al., 1987; Mathiowetz et al., 2001). A problem in asking respondents their income over a period of time and then their expenses over that same period of time is that they may try to derive values for both categories that agree. This phenomenon was observed in a survey of single mothers receiving welfare benefits who tended to underestimate expenses to match reported low income (Edin and Lein, 1997).

Other sources of error and bias in self-report of finances include underreporting of sensitive or illegal behavior related to money acquisition or exchange, and the difficulty of recalling often complex and variable patterns of income and financial transactions (Mathiowetz et al., 2001).

To elicit accurate retrospective data, event history calendars have been more effective than standard survey methods (Martyn and Belli, 2002). Timeline follow-back (TLFB) is an interview technique that incorporates reference to a calendar to cue accurate recall of the daily occurrence of events over a specific time period. The technique has provided desirable test-retest reliability for retrospective self-reports of alcohol use (Sobell et al., 1979) and has been extended to self-reports of other behaviors including smoking (Lewis-Esquerre et al., 2005) psychoactive substance use (Fals-Stewart et al., 2000), residential stability and homelessness (Tsemberis et al., 2007), and sexual risk behavior (Weinhardt et al., 2002).

In this study, we tested the performance of a new assessment of personal finances designed to improve measurement with adults with psychiatric issues, and applicable to those receiving federal financial support. The instrument, entitled "Timeline Historical Review of Income and Financial Transactions" (THRIFT), includes content from an unpublished assessment-as-usual (AAU) personal budgeting questionnaire, and incorporates TLFB and additional prompts to improve recall. We compared the test-retest reliability of responses on the THRIFT to those elicited by the AAU questionnaire in a sample of adults with psychiatric and substance use issues receiving Social Security financial benefits, hypothesizing that responses would be more reliable with the THRIFT.

Methods

Sample

We enrolled 28 adults with chronic psychiatric disabilities receiving outpatient services at a state-operated mental health clinic who reported using cocaine in the last 60 days, and who received Social Security benefits (SSI or SSDI). Patients were recruited for participation by study flyers and clinician referral. Participants were scheduled to report for two assessment sessions, one week apart, and were paid for participation. Three participants attended only the first assessment session, and were not included in the analysis. A fourth participant was omitted from analyses because he received a retroactive benefit check at the time of the first assessment that was not representative of his typical income.

Study Design

Participants attended two sessions approximately one week apart. At each session, participants were interviewed to determine their financial status using AAU and the THRIFT (both described in more detail below). Between the two money management assessments on each occasion, participants completed a computer activity to reduce the chance that responses on the first assessment would be memorized, and re-reported in the second assessment. To control for sequence effects, the order of the two financial assessments for each patient was randomly assigned, and this order was maintained for both testing occasions (the initial, and one week later).

Assessment of Finances—Personal income and financial transactions were assessed on each occasion by (1) AAU, asking participants to “think back over the last 30 days” about income, account balance, debt, and expenses and (2) the THRIFT, a revised version of the first instrument incorporating TLFB, timeline anchors, and additional prompts to elicit the same financial information. Each instrument assessed personal financial activity over the last 30 days, but the instruments differed in how domains of finance were sub-divided into items, and in how domains were reviewed.

Assessment As Usual (30-day recall)—Participants were asked about items in four domains; expenses, debt, income, and balance. Domains were presented separately, in sequence, and the interviewer provided prompts for specific items within domains. For example, within the expenses domain, participants were asked how much (in dollars) had been spent on utilities, rent, food, clothes, transportation, health care, and seven other distinct expenses. Participants were also asked how each expense was paid. In the debt domain, items included debt to family, friends, and credit card company. Items were similarly provided, and served as prompts, for income and balances domains. Administration time for the assessment was approximately 15 minutes.

THRIFT—The THRIFT was developed based on anecdotal information gathered from prior research in which participants’ self-reported financial status using AAU was compared to more comprehensive information collected during nine months of money-management based treatment (Rosen et al., 2010). Major changes from AAU included addition of an In-Kind domain (described below), more item prompts for Expense and Debt domains, and simultaneous (rather than sequential) review of expenses and debt. The THRIFT is available for download at www.behaviorchange.yale.edu.

By design, administration of the THRIFT assessment began with a review of the last 30 days on a calendar, identifying the dates to be covered by the interview, anchoring the beginning and end dates, and documenting major and routine events that occurred during that time. Dates of check receipt and payment of bills were among key events reviewed. The

assessment was divided into five domains, adding “In-Kind” benefits to the four domains covered in AAU. Administration time was approximately 15 minutes.

The In-Kind domain was defined to include any service or donation provided to the participant free-of-charge, or in exchange for something other than money, for which the participant would otherwise have had to pay. It also included bills or debts waived or paid on the participant’s behalf, and purchases made by others for the participant’s benefit. Examples of in-kind benefits included cable television purchased by a roommate, drugs provided by a friend, debt waived by a lender, and housing (rent) provided in exchange for maintenance work.

Expenses and Debt domains were expanded from the original assessment to include additional items, and were reviewed simultaneously during administration of the survey. As such, the review of expenses within each item served as a prompt for debts within the same item.

In-Kind benefits were reviewed immediately prior to Expenses and Debt, cueing recall of bills or services that were covered free-of-charge that may have altered typical expenses and debt for that time period. This new sequence of prompts was designed for greater coordination of information across domains, given the non-independence of items.

Kaufman Brief Intelligence Test, Second Edition (KBIT-2)—To account for the potential confounding effect of intelligence on stability of responses across occasions, participants were administered the KBIT-2 (Kaufman and Kaufman, 2004), a brief assessment of verbal and non-verbal intelligence.

Data Analysis

Items within the domains of income, expenses, balance, and debt were summed to form total dollar estimates for each domain. We used descriptive statistics to explore distributions of domain estimates, as well as discrepancies in estimates, between methods. Wilcoxon signed ranks tests were used to compare estimates by method. Pearson’s correlation coefficient quantified the reliability of domain estimates across occasions within the AAU and THRIFT conditions. We estimated zero-order and partial correlations between the first and second administration of each measure, controlling for KBIT scores. Correlations were transformed to z-scores using Fisher’s Z transformation and were compared across assessment methods by domain using the “Z-based version” of the Pearson-Filon test statistic for dependent, non-overlapping correlations. This method was found to provide valid results with samples as small as $n=20$ (Raghunathan et al., 1996). We also assessed the effect of assessment order (AAU or THRIFT first within a session) on response reliabilities by comparing correlations across order, averaging over assessment methods, using the same analytical technique.

Results

Sample Descriptives

Participants were 58% male, 63% African-American, 25% Caucasian, and 12% Hispanic, with a mean age of 45 years (range 22 – 58). Participants had diagnoses of schizophrenia (37%), schizoaffective disorder (25%), bipolar disorder (22%), major depressive disorder (21%), and PTSD (13%), and four participants (17%) had more than one Axis I diagnosis. Four participants (17%) had a technical degree and on average participants had 11 years of education. The group mean composite IQ score as measured by the Kaufmann Brief Intelligence Test (K-BIT 2; Kaufman and Kaufman, 2004) was 78 (range = 51–126; $SD=16$). With regard to substance use in the 28 days prior to the study, participants averaged

6 days of cocaine use, 4 days of cannabis use, 3 days of alcohol use, and 2 days of drinking to intoxication. The sample averaged less than one day of use of other substances. According to the THRIFT (Time 1), mean monthly income was \$933 (range \$303 – \$1297; $SD=212$), mean in-kind benefits were valued at \$79 (range \$0 – \$600; $SD=144$); mean expenses totaled \$1063 (range \$335–\$4493; $SD=816$); mean debt was \$3499 (range \$0 – \$59,205; $SD=11994$); and mean balance totaled \$25 (range –\$0 – \$230; $SD=56$).

Test-Retest Reliability

Reliability estimates for each assessment by domain are reported in Table 1. As hypothesized, score reliability was significantly higher with the THRIFT for three of the four domains: debt, income, and expenses ($p<.001$, $p<.05$, $p<.05$, respectively for each one-tailed test). Test-retest reliability was best for estimates of debt and expenses. Scores were substantially less stable over time for the balance variable. There was no consistent effect of assessment order on score reliability. KBIT scores were unrelated to response stability.

Domain Estimates by Method

Differences in domain estimates between methods varied by person; some participants estimated larger amounts with AAU than THRIFT and others estimated smaller amounts such that deviations between estimates were both positive and negative. These distributions of differences suggest random error in recall rather than systematic bias. This was supported by the Wilcoxon signed ranks tests which failed to reveal reliable, directional differences in estimates between methods (Table 2).

Discussion

Incorporating the TLFB interview method, and structuring assessment of finances to include additional prompts and simultaneous inquiry about dependent items (such as expenses and debt within the Rent item) resulted in decreased error in measurement as indicated by greater reliability in responses for three of four domains. Although our inquiry focused on the relative stability of responses for one instrument over another, the absolute reliability for three of four domains was also remarkable (.77 to .99), particularly considering the degree of disability among participants.

The addition of the in-kind category may not only improve reliability of estimates of expense and debt, but also provides an opportunity to gain insight into the unique economics of individuals with low income, who may be more likely to make exchanges that do not involve money and to receive income-equivalent benefits through informal arrangements with others. Questions probing about services provided free-of-charge, debts waived, and bills paid by others shed light on the means by which some individuals in poverty meet basic needs.

The balance domain appears to be measured with a great deal of error. However, that domain was unique in that respondents were asked about their balances in accounts and on-hand at the time of assessment, rather than averaged over 30 days. Thus, it is likely that inconsistency in values across time points reflects true differences in balance from one time to the next, rather than error in recall.

Conclusions

The structure and event cues provided by the THRIFT seem to obviate respondents' reliance on heuristics and other sources of error to provide retrospective estimates of income and financial transactions. Decreasing error in the measurement of these important indicators of

functioning and opportunity permits more accurate determinations of need and access to services, and can improve inferences about the effects of targeted interventions.

Whereas the current results represent improvement with a small, specific sample of individuals, the structure of the THRIFT was informed by strategies with demonstrated effects on the accuracy of recall of personal information across diverse samples. Thus, we encourage evaluation of the THRIFT with other groups for whom information about personal financial management is desired.

Finally, although the results indicate that the THRIFT reduced random error in recall in this confidential research study, it may not reduce or eliminate bias in estimates when reports of money earned or spent have financial or other consequences for respondents. Under high-stakes conditions, it is possible that information could be falsified, regardless of assessment method. Establishing rapport with respondents and providing a comfortable and trusting environment are critical elements of good assessment, and will maximize the validity of results.

Acknowledgments

This study was funded by NIDA: R01DA12952, R01DA025613.

The authors thank Samuel Ball for his review of an earlier version of the paper.

References

- Andrews, DA.; Bonta, J. The Level of Service Inventory–revised. Toronto, Canada: Multi-Health Systems; 1995.
- Bradburn NM, Rips LJ, Shevell SK. Answering autobiographical questions: The impact of memory and inference on surveys. *Science*. 1987; 236:158–161.
- Edin, K.; Lein, L. Making ends meet: How single mothers survive welfare and low-wage income. New York: Russell Sage Foundation; 1997.
- Fals-Stewart W, O’Farrell TJ, Freitas TT, McFarlin SK, Rutigliano P. The timeline follow-back reports of psychoactive substance use by drug-abusing patients: Psychometric properties. *J Consult Clin Psychol*. 2000; 68:134–144. [PubMed: 10710848]
- Hotz, VJ.; Scholz, JK. Measuring employment and income outcomes for low-income populations with administrative and survey data. In: Ver Ploeg, M.; Moffitt, RA.; Citro, CF., editors. *Studies of welfare populations: Data collection and research issues*. Washington, DC: National Academy Press; 2002. p. 275–315.
- Hufford MR, Shiffman S. Assessment methods for patient-reported outcomes. *Dis Manag Health Out*. 2003; 11:77–86.
- Kaufman, AS.; Kaufman, NI. Kaufman Brief Intelligence Test, second edition (KBIT-2). Bloomington, MN: Pearson, Inc; 2004.
- Lewis-Esquerre JM, Colby SM, Tevyah TO, Eaton CA, Kahler CW, Monti PM. Validation of the timeline follow-back in the assessment of adolescent smoking. *Drug Alcohol Depen*. 2005; 79:33–43.
- Lynn, P.; Jäckle, A.; Jenkins, SP.; Sala, E. Working Papers of the Institute for Social and Economic Research, paper 2004-28. Colchester: University of Essex; Dec. 2004 The impact of interviewing method on measurement error in panel survey measures of benefit receipt: Evidence from a validation study.
- Martyn KK, Belli RF. Retrospective data collection using event history calendars. *Nurs Res*. 2002; 51:270–274. [PubMed: 12131240]
- Mathiowetz, NA.; Brown, C.; Bound, J. Measurement error issues in surveys of the low income population. *Data collection on low income and welfare populations*. Washington, DC: National Academy Press; 2001.

- Raghunathan TE, Rosenthal R, Rubin DB. Comparing correlated but non-overlapping correlations. *Psychol Methods*. 1996; 1:178–183.
- Reichert JW, Kindelberger JC. Reliability of income and poverty data from the current population survey annual demographic supplement. *US Census Bureau Report*. 2000:151–156.
- Rosen MI, McMahon TJ, Rosenheck RA. Homeless people whose self-reported SSI/DI status is inconsistent with Social Security Administration records. *Social Secur Bull*. 2007; 67:53–61. [PubMed: 17972788]
- Rosen MI, Rounsaville BJ, Ablondi K, Black AC, Rosenheck RA. Advisor-teller money manager (ATM) therapy for substance abuse. *Psychiatr Serv*. 2010; 61:707–713. [PubMed: 20592006]
- Sobell LC, Maisto SA, Sobell MB, Cooper AM. Reliability of alcohol abusers' self-reports of drinking behavior. *Behav Res Ther*. 1979; 17:157–160. [PubMed: 426744]
- Stull DE, Leidy NK, Parasuraman B, Chassany O. Optimal recall periods for patient-reported outcomes: Challenges and potential solutions. *Curr Med Res Opin*. 2009; 25:929–942. [PubMed: 19257798]
- Tsemberis S, McHugo G, Williams V, Hanrahan P, Stefancic A. Measuring homelessness and residential stability: The residential timeline follow-back inventory. *J Community Psychol*. 2007; 35:29–42.
- Weinhardt LS, Carey MP, Maisto SA, Carey KB, Cohen MM, Wickramasinghe SM. Reliability of the timeline follow back sexual behavior interview. *Ann Behav Med*. 2002; 20:25–30. [PubMed: 9755348]

Table 1

Test-Retest reliability estimates by method and domain (n=24).

Method	Expense	Debt	Income	Balance
THRIFT	.91 *	.99 ***	.77 *	.37
AAU	.83	.92	.40	.29

 $p < .001$,

*
 $p < .05$; one-tailed tests of probability.

Table 2

Total dollar estimates by domain and method (n=24).

Domain	THRIFT	AAU
	Mean (SD)	Mean (SD)
Expense	1063.42 (816.22)	1039.75 (750.73)
Debt	3499.17 (11994.13)	4671.13 (11170.93)
Income	933.46 (211.55)	1049.21 (336.39)
Balance	25.25 (55.69)	22.42 (50.77)