# Concordance between self-reports and archival records of physician visits: a case-control study comparing individuals with and without alcohol use disorders in the community

**Joseph E. Glass**[1] and **Kathleen K. Bucholz**[2]

[1]George Warren Brown School of Social Work, Washington University in St. Louis, 1 Brookings Drive, Campus Box 1196, St. Louis, Missouri, United States, 63130

[2]Department of Psychiatry, Washington University in St. Louis School of Medicine, 660 Euclid, St. Louis, MO, United States 63130

## Abstract

**Objective**—The accuracy of self-reported healthcare use among individuals with alcohol use disorders (AUD) has been questioned. The present study attempts to compare the accuracy of self-reported physician visits for individuals who differ with respect to their history of AUDs.

**Methods**—Our data source was a 14-year follow-up of individuals interviewed at the St. Louis site of the 1981-1983 Epidemiologic Catchment Area Study (ECA). We used a case-control design (N=237) to compare the accuracy of self-reports among ECA participants with stably-diagnosed AUDs (cases; n=75) to two comparison groups: those with problem/very heavy drinking (n=81) and those unaffected by alcohol (n=81). Intraclass correlation coefficients (ICC) described the concordance between self-reports and archival records of physician visits in the prior six months. We used multinomial logistic regression to identify characteristics associated with under-reporting and over-reporting, and zero-truncated Poisson regression to identify characteristics associated with discordance severity.

**Results**—Self-reports of cases had substantial concordance with physician records (ICC=0.74, CI=0.61-0.83). As compared to cases, those with problem/very heavy drinking had a significantly higher ICC, and those who were unaffected by alcohol had a significantly lower ICC. However, differences in concordance disappeared when using regression models that adjusted for factors known to affect the accuracy of self-reported healthcare use. Utilization frequency was a strong predictor of inaccurate reporting.

**Conclusions**—These findings suggest AUD status may not independently affect the accuracy of self-reports. Counts of physician visits for those with AUD may be considered accurate when utilization frequency is low.

## Keywords

Alcohol use disorders; service use; concordance; self-report

Correspondence and requests for reprints should be addressed to: Joseph E. Glass, 1 Brookings Drive, Campus Box 1196, St. Louis, Missouri 63130, Ph: 734-678-3709, Fax: 314-934-8511, jeglass@wustl.edu.

## 1. Introduction

While archival healthcare records have been considered the gold standard for measuring services use (Bhandari and Wagner, 2006), their use is often precluded in research conducted outside of healthcare settings. Difficulties may arise when collecting administrative data from a multitude of healthcare agencies, which may be due to feasibility (costs), or concerns about protected health information. On the other hand, the use of other methods to assess service utilization including self-report may be inadequate when precise measurements are needed, such as in the calculation of healthcare costs (Garnick et al., 2002). Given that many healthcare quality measures also rely on sophisticated information (the types of healthcare providers seen, procedures received, or the number of visits attended) (Garnick et al., 2006; Harris et al., 2009; McGlynn et al., 2003), self-report data may be less than optimal as a sole data source when estimating the quality of care.

The accuracy of self-reported healthcare utilization has been given attention in a number of studies. Bhandari and Wagner's (2006) systematic review of this literature proposed a conceptual model describing factors that affect the accuracy of self-reports. Researcher-modifiable factors include, among others, the recall timeframe of services utilization (shorter is better), the use of memory aids, and the domains of services assessed. Whereas survey properties are modifiable by the researcher, participant characteristics are considered fixed. For example, cognitive impairment can affect the accuracy of self-reports, and psychosocial factors such as age, gender, and culture may affect accuracy due to their influence on the interpretation of survey questions.

There has been interest in determining the accuracy of self-reported healthcare utilization in populations with AUD or other substance use disorders (SUDs). Several studies examined the concordance between self-report and archival measures with a focus on validating measures of service utilization developed specifically for individuals receiving specialty SUD treatments. Breslin et al. (2001) found excellent concordance in the assessment of services received at a single SUD agency over a 6-8 month period when self-reports were ascertained using a Timeline Followback method supplemented by numerous recall strategies. On the other hand, Zanis et al. (1997) found low rates of agreement and weak correlations between self-report and records kept by another agency. These studies used small samples and were restricted to the services received at a single SUD agency; more recently there has been an interest in examining the accuracy of self-reports in a broader array of service agencies. For example, Killeen et al. (2004) obtained archival records from emergency departments, inpatient behavioral health and medical units, social service agencies, and legal agencies throughout the community in which a sample of treatment-seeking individuals with AUD received services. They found that self-reports of 12-month service use had moderate to high agreement with agency records across these service domains; however, those with higher scores on the Alcohol Dependence Scale (Skinner and Allen, 1982) had worse agreement. The authors speculated that those with more severe AUD may have had a higher frequency of service utilization or higher levels of cognitive impairment, which could lead to inaccuracies in reporting.

While some evidence suggests that the reliability of self-reported healthcare use may be compromised for some individuals with AUD, it remains unknown whether the accuracy of self-reports among those with AUD actually differs from those who are unaffected by alcohol problems. Furthermore, the examination of accuracy in community samples is warranted, given that those in treatment are just a minority of those with alcohol problems and tend to have a greater problem severity across various psychosocial and clinical domains (Berkson, 1946; Cohen et al., 2007; Mojtabai, 2005).

The present study attempts to compare the accuracy of self-reported physician visits in individuals who differ with respect to their prior history of AUDs. We compare individuals with a stably-diagnosed lifetime history of AUD versus two comparison groups: those without such a history who have at-risk drinking, and those who are unaffected by alcohol. We examine the concordance between self-reported physician visits and archival records of physician visits (medical records and billing/insurance data) that occurred across a range of service settings (e.g. medical services, mental health services, SUD services) in the community. Our focus on physician visits allows us to examine concordance in a majority of individuals, including those who do not need or seek specialty SUD services. We further attempt to identify the extent to which alcohol problem severity, cognitive problems, and psychosocial factors are related to the accuracy of self-reports.

## 2. Methods

### 2.1. Data source

The data source used in this study was the 1997 Health Services Use and Costs Study (HSUC). HSUC was a 14-year follow-up study of individuals who were interviewed in at least two waves of the 1981-1983 Epidemiologic Catchment Area Study (ECA). HSUC included participants from the St. Louis site of ECA, which was one of the five catchment areas of ECA. HSUC used a case-control design to analyze the current health services use of individuals with a stably-diagnosed lifetime alcohol use disorder (AUD) at ECA in comparison to two lower-risk groups. HSUC included three waves of structured telephone interviews that were 6 months apart, as well as a medical abstraction component for participants who consented to a release of all medical records including patient charts and billing/insurance records. The first wave of self-report and medical abstraction data was used in the current study. This study was approved by the Washington University in St. Louis IRB.

### 2.2. Sampling frame and participants

The case group recruited in HSUC, individuals with "stably-diagnosed AUDs" (SA) consisted of ECA participants who met criteria for DSM-III lifetime alcohol abuse and/or dependence in both interview waves (n=243). Approximately 58% of cases had past-year AUD symptoms when assessed at ECA. Prior studies have found that requiring a lifetime diagnosis at two waves, rather than a single wave, improves the diagnostic validity of the disorder under study (Edens et al., 2008; Nelson and Rice, 1997).

Two comparison groups were frequency-matched to cases using age, gender, and race. The first comparison group, "problem/very heavy drinkers" (VHD) (n=242) were identified by meeting any of three criteria: 1) endorsing a very high level of alcohol consumption but no alcohol problems in their lifetime (n=70); 2) endorsing at least one alcohol problem in their lifetime (n=80); or 3) meeting criteria for lifetime AUD in a single, both not both waves (n=92). Approximately 29% of problem/very heavy drinkers had past-year AUD symptoms when assessed at ECA. Those meeting criteria for problem/very heavy drinking based on alcohol consumption endorsed one of ECAs three alcohol consumption measures, which assessed for the lifetime presence of 1) ever drinking seven or more drinks per day every day, for at least two weeks or longer; 2) ever drinking seven or more drinks per day at least once a week, for a couple of months or more, and 3) ever drinking about 20 drinks in one day, at least once. Of the 80 who met criteria for problem/very heavy drinking on the basis that they had at least one alcohol-related problem, 35 had a history of very heavy drinking, and 45 did not. Of the 92 individuals who met criteria for problem/very heavy drinking based on their diagnosis of AUD at a single, but not both waves (i.e. sporadic lifetime diagnosis), 55 did so at wave 1 only, and 37 did so at wave 2 only. We note that while it is

possible that some of the 37 cases with a lifetime diagnosis only at wave 2 may be incident cases rather than unreliable reporters, secondary analyses estimated that over half of these individuals actually endorsed an onset of AUD that occurred before the first interview. The second comparison group, "alcohol-unaffected" individuals (AU) (N=226), consisted of ECA participants who did not meet criteria for DSM-III AUD, very heavy drinking, or alcohol problems at either wave.

A total of 711 individuals (243 SA, 242 VHD, and 226 AU) were targeted for recruitment. Additional details about recruitment may be found in a prior study (Edens et al., 2008). Of the 444 interviewees, 78.6% were male and the average age was 50.1 (SD=10.1). Participants were 77.7% white, 20.3% black, and 2% were Hispanic, American Indian, or another race (which reflected the demographic characteristics of the St. Louis ECA). Written informed consent was obtained from participants after a complete description of the study was provided.

### 2.3. Measures

**2.3.1. Self-reported physician visits—**Telephone interviews assessed health services use that occurred in the six months prior to the interview. Respondent booklets were mailed to participants to assist with the interview, which included a calendar to help with recall. Although the presence of visits to a wide range of providers was queried, we limited the scope of our analyses to physician visits because the rates of visits to non-physicians were low (e.g. less than 3% reported visiting psychologists, social workers, or nurse practitioners). To assess physician visits, participants were asked, "During the past six months, did you receive care or treatment for any health problems from any medical doctor or osteopath like a general practice physician, internist, cardiologist, surgeon, or psychiatrist?" Participants were instructed to report care occurring in any setting, such as offices, clinics, emergency rooms, hospitals, long-term care facilities, mental health care facilities, and drug and alcohol treatment programs. Participants were then asked for the name and address of each provider that was seen. For each provider, the number of visits was queried. To calculate the total number of visits attended, we summed physician visits across providers. Thus, a single variable reflected the count of visits to physicians in the past six months that occurred across service settings. A signed a release form was obtained for the acquisition of records.

**2.3.2. Past-year alcohol problem severity—**We created a hierarchical variable that captured five levels of past-year alcohol problem severity at 14-year follow-up: (1) DSM-IV AUD, (2) problem drinking, (3) at-risk drinking, (4) low-risk drinking, and (5) abstention. For #3, we were only able to approximate the NIAAA definition of risky drinking (National Institute on Alcohol Abuse and Alcoholism, 2005) as drinking to intoxication at least once in the past year, or drinking an average of at least 16 drinks for men and 8 for women per week in the past year.

**2.3.3. Lifetime alcohol problem severity/AUD symptom count—**The HSUC survey included Diagnostic Interview Schedule (DIS) version IV (DIS-IV) modules to assess DSM-IV lifetime alcohol abuse and dependence at 14-year follow-up. Our lifetime alcohol problem severity measure summed the number of DSM-IV alcohol abuse and/or dependence criteria that were ever met, yielding a scale of 0-11.

**2.3.4. Abnormal cognitive status—**Interviewers utilized a checklist to rate respondents as having problems with orientation (time, place, or person), memory (clouding of consciousness, inability to concentrate, amnesia, poor recent memory, poor remote memory, or confabulation), or intellect (below normal intellect, paucity of knowledge, or vocabulary

poor) upon interview completion. Interviewers received training to base these ratings on the respondent's behavior and demeanor during the telephone interview. We created a dichotomous variable that indicated whether or not the respondent's cognitive status was rated as abnormal, based on the identification of "borderline" or "definite" problems (versus no problems) with orientation, memory, or intellect.

**2.3.5. Other self-report measures—**We created a dichotomous variable to indicate whether or not participants were under a doctor's care in their lifetime for any of the following illnesses: heart disease or heart attack, cancer, hepatitis or cirrhosis, stroke, arthritis, asthma, bleeding ulcers, diabetes, tuberculosis, epilepsy, or any other serious and long-lasting physical illness. Last, a variable to reflect any health insurance coverage, including employer-based, public (VA, Medicaid, Medicare), or self-paid was created.

**2.3.5. Physician visits in medical records—**Medical and billing records were obtained from all healthcare providers and institutions that were identified by respondents. For each encounter found in the records, two independent abstractors coded the type of provider seen, provider specialty, and visit setting. Each abstractor received extensive training, and was required to achieve acceptable agreement with several master abstractions prior to commencing abstraction work. The two abstractions were compared and a final abstraction per research participant was arrived at by consensus among senior project investigators. For the present report, only physician visits occurring in any service setting were included. We did not count laboratory visits or telephone contacts.

## 2.4. Data analysis

Analyses were conducted with HSUC respondents who self-reported one or more physician visits and had one or more archival records returned that contained physician visits. Figure 1 depicts the process used to derive the final analytic sample. Of the 444 interviewees, 328 (73.8%) reported any health service use in the prior six months. Of these individuals, 46 (14.0%) reported no visits to physicians, six (1.8%) had missing self-report data on physician visits, eight (2.4%) refused a release of medical records, 23 (7.0%) had no medical records returned, and eight (2.4%) had medical records returned with no physician visits. The remaining 237 participants constituted our analytic sample.

Data analyses were conducted using SAS for UNIX version 8.2. We calculated descriptive statistics for the analytic sample (N=237), then tested for significant differences between case and each comparison group using chi-square tests for categorical measures and ANOVA for continuous measures. Differences between the two comparison groups were not relevant to this study.

**2.4.1. Concordance between self-reports and archival records—**Intraclass correlation coefficients (ICC) were used to estimate the concordance between self-reports and archival records. A two-way mixed-effects ANOVA was used to calculate absolute-agreement ICCs using counts of visits as the response variable (McGraw and Wong, 1996). Each form of measurement (self-report and archival records) was conceptualized as a fixed rater, and raters were nested within participants (participants were the random factors). ICCs were calculated for the overall subsample and then separately for each case and comparison group. We tested for significant differences in ICCs using a distribution-free test for the difference between independent correlations.

**2.4.2. Rates and correlates of under-reporting, over-reporting, and agreement —**Following Bhandari and Wagner (2006), we classified participants who reported fewer visits than were found in archival data as "under-reporters"; who reported more visits than in

the records as "over-reporters"; and those with the same number of visits as in "agreement." Multinomial logistic regression was used to determine the characteristics associated with each report category. We controlled for a number of characteristics known to affect the accuracy of self-reports or administrative data (Bhandari and Wagner, 2006; Garnick et al., 2002; Killeen et al., 2004) including past-year and lifetime alcohol problem severity, age, gender, race (although case and comparison groups were frequency-matched on age, gender, and race, we included these variables to adjust for differences that were potentially introduced by our exclusion criteria), employment status, education status, cognitive impairment, the number of healthcare visits received, self-reported chronic medical illnesses, and insurance status. The past-year alcohol problem severity measure may be the best indicator of the influence of recent alcohol use on the recall of health services. We also included the lifetime alcohol problem severity measure (assessed at follow-up) to capture alcohol problems that may have developed during the 14 years between the ECA and HSUC interviews; it is possible that problematic alcohol use during this period could affect current reporting. A zero-truncated Poisson regression was deemed the most suitable analysis to determine characteristics associated with the severity of discordance among inaccurate reporters because we were modeling the visit count distribution, and only participants with discordance were included.

## 3. Results

### 3.1. Descriptive statistics

Descriptive statistics for the analytic sample (n=237) and stratified by case/comparison group are displayed in Table 1. Few differences were observed; ECA cases had a significantly higher lifetime alcohol problem severity at the follow-up interview than problem/very heavy drinkers and alcohol-unaffected individuals.

### 3.2. Concordance between self-reports and archival records

The ICCs for the overall sample and subsamples were indicative of substantial agreement and are displayed in Table 2. Distribution-free tests revealed that cases had a significantly lower ICC than problem/very heavy drinkers (z=-3.192, $p$<0.01), but a significantly higher ICC than alcohol-unaffected individuals (z=2.032, $p$<0.05).

### 3.3. Over-reporting, under-reporting, and agreement

Overall, slightly more individuals (37.1%) under-reported physician visits than over-reported (32.9%) or agreed (30.0%) (see Table 2). No differences were observed across the case and comparison groups.

### 3.4. Adjusted correlates of over-reporting and under-reporting

Results from the multinomial logistic regression are displayed in Table 3. Only the number of physician visits (per archival data) was significantly associated with over-reporting (RRR=1.18, 95% CI=1.01-1.37) or under-reporting (RRR=1.31, 95%CI=1.14-1.52). Thus, for every additional physician visit, the relative risk of over-reporting and under-reporting increased by 18% and 31%, respectively.

### 3.5. Adjusted correlates of the severity of discordance

Results from the zero-truncated Poisson regression are shown in Table 4. Past-year full-time employment was significantly associated with a lower discordance severity (IRR=.74, 95% CI=.56-.97). Being employed full-time would decrease the expected number of discordant visits by 36%. Additionally, having a greater number of physician visits was significantly

associated with a higher discordance severity (IRR=1.07, 95% CI=1.06-1.09). That is, for every additional physician visit, the expected number of discordant visits increased by 7%.

## 4. Discussion

At 14-year follow-up, self-reports of physician visits for ECA participants with a stably-diagnosed lifetime AUD had substantial concordance with the number of physician visits abstracted from medical records. Most commonly, cases reported more physician visits than were found in the archival data. Further, ICC estimates indicated that self-reports for those with stably-diagnosed lifetime AUD were significantly less accurate than that of those with problem/very heavy drinking, but more accurate than that of those who were not affected at ECA. However, after controlling for factors known to affect the accuracy of self-reported healthcare utilization, differences disappeared. Altogether, these findings suggest that a history of alcohol problems may not independently drive inaccuracies in self-reports of physician visits.

Our expectation that individuals with a history of stably-diagnosed AUD would be less accurate reporters than their counterparts without such history was based on data from treatment samples (Killeen et al., 2004; Zanis et al., 1997). The present study was based on a community sample, where most do not seek treatment and are likely to have a less severe disorder (Cohen et al., 2007). Further, Bhandari and Wagner (2006) discuss the importance of study design characteristics that affect the accuracy of self-reports. The domains of services assessed differed between extant literature and the present study; others did not include routine services as we did, but rather focused on emergency and acute services (Killeen et al., 2004). However, of note is that in a study of Medicare claims data, no relationship between alcohol consumption and over-reporting or under-reporting physician of visits was observed (Wolinsky et al., 2007). Health services researchers should thoughtfully consider the sampling and design characteristics for their study when choosing a method to assess the frequency of healthcare use.

Still our data suggested that the utilization frequency of physician visits was strongly associated with over-reporting, under-reporting, and the severity of discordance. Although others have suggested that shortening the recall period (to 3-6 months) would improve accurate reporting in the context of high utilization (Bhandari and Wagner, 2006), our recall period was indeed 6 months. Perhaps, recall periods of shorter than six-months should be used to achieve optimal concordance. We also found that being employed in the past year was associated with a lower discordance severity for those who reported inaccurately. Similarly, Killeen et al. (2004) found that accurate reporters were more likely to be employed. Because we controlled for other attributes, we feel that it is unlikely that employment is simply a marker of higher cognitive status. Perhaps employment helps to improve recall because it is (typically) a time-structured and scheduled activity. For example, having to take time off of work to attend an appointment with a doctor could serve as cue that increases recall accuracy.

Several alcohol-specific mechanisms that affect the retrieval and self-reporting of information have been hypothesized (Del Boca and Noll, 2000). Such mechanisms include intoxication, physical conditions related to alcohol (e.g. fatigue, withdrawal), alcohol-related psychological states (e.g. anxious and depressed moods), and cognitive impairment. These mechanisms have been implicated in problems with general cognitive functioning in empirical studies (Samokhvalov et al., 2010). Thus, self-report accuracy could be affected by transient states (e.g. intoxication, withdrawal) that affect memory consolidation, retrieval, and the processing of survey questions, as well as by longer-lasting and permanent cognitive impairment (e.g. Korsakoff syndrome, alcohol-related dementia) (Kopelman et al., 2009;

Oslin et al., 1998). In the current study, it is possible that the use of memory cues (i.e. calendars, asking patients to first recall the locations of physician visits before recalling the number of visits attended) lessened the impact of alcohol-related cognitive impairment on the retrieval of information. Alternatively, the nature of these problems may simply be less common and/or less severe in community samples (Fein et al., 2002; Gazdzinski et al., 2008). The results of the current study suggest that the alcohol-specific mechanisms that influence the accuracy of self-reports may be better studied in treatment samples. In community samples, more attention should be given to impact of utilization frequency on the accuracy of self-reports.

### 4.1. Limitations

Several limitations are important to note. Interviewers assessed the number of physician visits by asking "about how many visits" were attended, which could have led participants to use less effort when recalling visits. On the other hand, some recommend the use of phrases such as "your best estimate is fine" in survey research to help decrease the pressure on participants when information is difficult to recall (Dillman, 2000). Additionally, the survey assessed counts of physician visits and the presence of visits within specific service settings using separate questions, thus we were unable to conduct concordance analyses stratified by service setting. It is also noted that the ascertainment of medical record data was conditional on self-reports of having received services from a provider/agency. This may have led to an underestimation of underreporting in this study. This limitation could be partially avoided in studies of concordance that use self-report and administrative data from enrollees of large integrated healthcare systems or insurers. While there was a high response rate from agencies in the return of medical records, data were not available that would indicate that records were not returned, introducing the possibility that some physician visits were missing from archival data. It is also worth noting that there were some differences in reasons for non-response across case and comparison groups (Edens et al., 2008) which could influence the results of the study; although, there were no differences in overall interview rates across groups. In addition, cases were matched to controls based on AUD status at ECA whereas concordance was assessed at 14-year follow-up, thus comparisons across these groups could reflect changes in alcohol severity between ECA and follow-up. Our ICC calculations were unable to adjust for any potential differences. However, we expect that controlling for past-year and lifetime severity at 14-year follow-up in our adjusted models helped to address any unintended differences in cases and comparison groups. We also note that health service use was assessed in the prior six months, yet we controlled for lifetime and past-year alcohol problem severity; it is possible that participants stopped drinking more than six months before the interview.

### 4.2. Conclusions

Using a case-control design and a 14-year longitudinal follow-up of the landmark ECA study, we examined differences between self-reports and archival records of physician visits in a community sample that included individuals with and without an AUD. Our findings suggest that prior and current alcohol problem severity and past-year alcohol consumption patterns do not have an independent effect on the accuracy of self-reports of physician visits in community samples. Thus, we believe that self-reported counts of physician visits for those with AUDs in the community can be considered accurate by policymakers and health services researchers, particularly when utilization frequency is low. It is crucial to consider all aspects of study design when attempting to generalize results regarding the accuracy of self-reported health care utilization.

# References

Berkson J. Limitations of the application of fourfold table analysis to hospital data. Biometrics Bull. 1946; 2:47–53.

Bhandari A, Wagner T. Self-reported utilization of health care services: improving measurement and accuracy. Med Care Res Rev. 2006; 63:217–235. [PubMed: 16595412]

Breslin FC, Borsoi D, Cunningham JA, Koski-Jannes A. Help-seeking timeline followback for problem drinkers: preliminary comparison with agency records of treatment contacts. J Stud Alcohol. 2001; 62:262–267. [PubMed: 11327193]

Cohen E, Feinn R, Arias A, Kranzler HR. Alcohol treatment utilization: findings from the National Epidemiologic Survey on Alcohol and Related Conditions. Drug Alcohol Depend. 2007; 86:214–221. [PubMed: 16919401]

Del Boca FK, Noll JA. Truth or consequences: the validity of self-report data in health services research on addictions. Addiction. 2000; 95(Suppl 3):S347–360. [PubMed: 11132362]

Dillman, DA. Mail and internet surveys: the tailored design method. Wiley; New York: 2000.

Edens EL, Glowinski AL, Grazier KL, Bucholz KK. The 14-year course of alcoholism in a community sample: do men and women differ? Drug Alcohol Depend. 2008; 93:1–11. [PubMed: 17935914]

Fein G, Di Sclafani V, Cardenas VA, Goldmann H, Tolou-Shams M, Meyerhoff DJ. Cortical gray matter loss in treatment-naive alcohol dependent individuals. Alcohol Clin Exp Res. 2002; 26:558–564. [PubMed: 11981133]

Garnick DW, Hodgkin D, Horgan CM. Selecting data sources for substance abuse services research. J Subst Abuse Treat. 2002; 22:11–22. [PubMed: 11849903]

Garnick DW, Horgan CM, Chalk M. Performance measures for alcohol and other drug services. Alcohol Res Health. 2006; 29:19–26. [PubMed: 16767849]

Gazdzinski S, Durazzo TC, Weiner MW, Meyerhoff DJ. Are treated alcoholics representative of the entire population with alcohol use disorders? A magnetic resonance study of brain injury. Alcohol. 2008; 42:67–76. [PubMed: 18358984]

Harris AH, Humphreys K, Bowe T, Kivlahan DR, Finney JW. Measuring the quality of substance use disorder treatment: evaluating the validity of the Department of Veterans Affairs continuity of care performance measure. J Subst Abuse Treat. 2009; 36:294–305. [PubMed: 18835678]

Killeen TK, Brady KT, Gold PB, Tyson C, Simpson KN. Comparison of self-report versus agency records of service utilization in a community sample of individuals with alcohol use disorders. Drug Alcohol Depend. 2004; 73:141–147. [PubMed: 14725953]

Kopelman MD, Thomson AD, Guerrini I, Marshall EJ. The Korsakoff syndrome: clinical aspects, psychology and treatment. Alcohol Alcohol. 2009; 44:148–154. [PubMed: 19151162]

McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, DeCristofaro A, Kerr EA. The quality of health care delivered to adults in the United States. N Engl J Med. 2003; 348:2635–2645. [PubMed: 12826639]

McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods. 1996; 1:30–46.

Mojtabai R. Use of specialty substance abuse and mental health services in adults with substance use disorders in the community. Drug Alcohol Depend. 2005; 78:345–354. [PubMed: 15893166]

National Institute on Alcohol Abuse and Alcoholism. Helping Patients Who Drink Too Much: A Clinician's Guide. Bethesda, MD: 2005.

Nelson E, Rice J. Stability of diagnosis of obsessive-compulsive disorder in the Epidemiologic Catchment Area study. Am J Psychiatry. 1997; 154:826–831. [PubMed: 9167511]

Oslin D, Atkinson RM, Smith DM, Hendrie H. Alcohol related dementia: proposed clinical criteria. Int J Geriatr Psychiatry. 1998; 13:203–212. [PubMed: 9646147]

Samokhvalov AV, Popova S, Room R, Ramonas M, Rehm J. Disability associated with alcohol abuse and dependence. Alcohol Clin Exp Res. 2010; 34:1871–1878. [PubMed: 20662803]

Skinner HA, Allen BA. Alcohol dependence syndrome: measurement and validation. J Abnorm Psychol. 1982; 91:199–209. [PubMed: 7096790]

Wolinsky FD, Miller TR, An H, Geweke JF, Wallace RB, Wright KB, Chrischilles EA, Liu L, Pavlik CB, Cook EA, Ohsfeldt RL, Richardson KK, Rosenthal GE. Hospital episodes and physician visits: the concordance between self-reports and medicare claims. Med Care. 2007; 45:300–307. [PubMed: 17496713]

Zanis DA, McLellan AT, Belding MA, Moyer G. A comparison of three methods of measuring the type and quantity of services provided during substance abuse treatment. Drug Alcohol Depend. 1997; 49:25–32. [PubMed: 9476696]
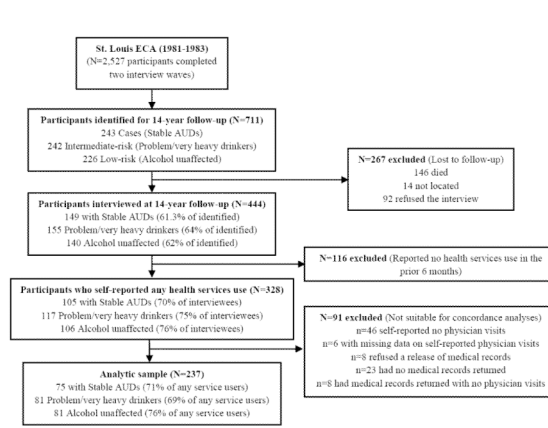
**Figure 1.**
Analytic sample for concordance analyses

**Table 1**

Descriptive statistics of HSUC participants in the analytic sample (N=237).

| | Overall (n=237) | Case group, ECA Stably-diagnosed DSM-III AUDs (n=75) | Comparison group 1, ECA problem/very heavy drinkers (n=81) | Comparison group 2, ECA alcohol-unaffected (n=81) | Cases vs. Group 1[a] | Cases vs. Group 2[a] |
|---|---|---|---|---|---|---|
| | N (%) or Mean (SD) | % or Mean (SD) | % or Mean (SD) | % or Mean (SD) | | |
| Mean age | 52.3 (10.42) | 50.6 (8.83) | 53.2 (10.82) | 53.0 (11.27) | 2.7 (1), p=0.10 | 2.2 (1), p=0.15 |
| Male | 185 (78.1) | 58 (77.3) | 64 (79.0) | 63 (77.8) | 0.1 (1), p=0.80 | 0.0 (1), p=0.95 |
| White | 182 (79.1) | 60 (81.1) | 61 (79.2) | 61 (77.2) | 0.1 (1), p=0.77 | 0.3 (1), p=0.56 |
| HS education[b] | 200 (85.1) | 60 (80.0) | 70 (86.4) | 70 (88.6) | 1.2 (1), p=0.28 | 2.2 (1), p=0.14 |
| Employed in the past year[b] | 168 (71.5) | 57 (76.0) | 57 (70.4) | 54 (68.4) | 0.6 (1), p=0.42 | 1.1 (1), p=0.29 |
| Has health insurance | 225 (94.9) | 70 (93.3) | 78 (96.3) | 77 (95.1) | 0.7 (1), p=0.40 | 0.2 (1), p=0.64 |
| Mean lifetime alcohol problem severity (AUD symptom count) | 2.5 (2.78) | 4.8 (2.87) | 2.2 (2.25) | 0.8 (1.45) | **41.2 (1), p<0.0001** | **123.1 (1), p<0.0001** |
| Past-year alcohol problem severity | | | | | **13.4 (4), p<0.01** | **36.8 (4), p<0.0001** |
| DSM-IV AUD | 11 (4.6) | 5 (6.7) | 6 (7.4) | 0 (0.0) | | |
| Problem drinking | 39 (16.5) | 25 (33.3) | 9 (11.1) | 5 (6.2) | | |
| At-risk drinking | 36 (15.9) | 15 (20.0) | 14 (17.3) | 7 (8.6) | | |
| Low-risk drinking | 92 (38.8) | 17 (22.7) | 31 (38.3) | 44 (54.3) | | |
| Abstention | 59 (24.9) | 13 (17.3) | 21 (25.9) | 25 (30.9) | | |
| Abnormal cognitive status | 21 (8.9) | 6 (8.0) | 7 (8.6) | 8 (9.9) | 0.0 (1), p=0.88 | 0.2 (1), p=0.68 |
| Chronic medical disorder[b] | 158 (67.2) | 54 (72.0) | 50 (61.7) | 54 (68.4) | 1.8 (1), p=0.17 | 1.47 (1), p=0.23 |

[a] Chi-square values (for categorical measures) or F values (for continuous measures) with their degrees of freedom in parentheses are displayed.

SD=standard deviation

[b] Missing data: n=2 for eduction; n=2 for employment; n=2 for chronic medical problems

**Table 2**

ICCs, report category, and mean physician visits (N=237).

| | Overall (N=237) | Case group, ECA Stably-diagnosed DSM-III AUDs (n=75) | Comparison group 1, ECA problem/very heavy drinkers (n=81) | Comparison group 2, ECA alcohol-unaffected (n=81) | Cases vs. Group 1[a] | Cases vs. Group 2[a] |
|---|---|---|---|---|---|---|
| Intraclass correlation coefficient: ICC (95% CI) | 0.78 (0.73–0.83) | 0.74 (0.61–0.83) | 0.90 (0.84–0.93) | 0.55 (0.37–0.68) | [a]$z$=−3.192, $p$ <0.01 | [a]$z$=2.032, $p$ <0.05 |
| Report status: N (%) | | | | | [b]3.66 (2), p=0.16 | [b]0.88 (2), p=0.64 |
| Under-report | 88 (37.1%) | 25 (33.3%) | 31 (38.3%) | 32 (39.5%) | | |
| Agreement | 71 (30.0%) | 20 (26.7%) | 20 (35.8%) | 22 (27.2%) | | |
| Over-report | 78 (32.9%) | 30 (40.0%) | 21 (25.9%) | 27 (33.3%) | | |
| Physician visits: Mean (SD) | | | | | | |
| Per self-report | 4.69 (7.64) | 5.16 (10.11) | 4.26 (6.88) | 4.68 (5.46) | [c]0.43 (1), p=0.51 | [c]0.14 (1), p=0.71 |
| Per archival data | 4.48 (5.44) | 3.80 (4.15) | 4.95 (7.12) | 4.64 (4.38) | [c]1.47 (0.23), p=0.23 | [c]1.51 (1), p=0.22 |

[a] Z values are displayed from a distribution-free test for the difference between independent correlations

[b] Chi-square values with their degrees of freedom in parentheses are displayed.

[c] F values with their degrees of freedom in parentheses are displayed.

### Table 3

Multinomial logistic regression showing the characteristics associated with over-reporting and under-reporting as compared to reporting accurately

|  | Over-reporting RRR (95% CI) | Under-reporting RRR (95% CI) |
|---|---|---|
| *Comparison group* | | |
| Alcohol-unaffected[a] | .83 (.27-2.51) | .89 (.29-2.72) |
| Problem/very heavy drinker[a] | .53 (.20-1.42) | .68 (.26-1.82) |
| *Demographic characteristics* | | |
| Age | .97 (.93-1.02) | 1.02 (.97-1.07) |
| Black | 1.00 (.38-2.64) | 1.13 (.44-2.89) |
| Female | 1.23 (.49-3.10) | 1.17 (.47-2.93) |
| Has a HS education | 1.22 (.37-4.00) | 1.40 (.42-4.68) |
| Employed in the past year | .48 (.16-1.47) | 1.06 (.35-3.21) |
| Insured | 1.17 (.20-6.96) | .64 (.12-3.36) |
| *Illness severity* | | |
| Lifetime alcohol problem severity (AUD symptom count) | 1.07 (.90-1.28) | 1.04 (.87-1.24) |
| Past-year alcohol problem severity | | |
| DSM-IV AUD[b] | 1.01 (.37-2.80) | 1.21 (.45-3.29) |
| Problem drinker[b] | .62 (.18-2.16) | .51 (.15-1.78) |
| At-risk drinker[b] | .78 (.23-2.59) | .94 (.27-3.21) |
| Low-risk drinker[b] | 1.13 (.19-6.83) | .96 (.14-6.81) |
| Abnormal cognitive status | 2.30 (.48-10.99) | 1.03 (.20-5.34) |
| Chronic medical disorder | 1.44 (.65-3.18) | .91 (.41-2.05) |
| Number of healthcare visits | **1.18 [*] (1.01-1.37)** | **1.33*** (1.15-1.54)** |

RRR=Relative risk ratio. CI=Confidence interval.

[*] $p < .05$,

[****] $p < .001$

Model statistics: $\chi^2$ (32, N=226) = 48.36, $p < 0.05$. Psuedo $R^2 = 0.10$.

[a] Reference group was those with stably-diagnosed alcohol use disorder

[b] Reference group was those who abstained from alcohol in the past year

## Table 4

Zero-truncated Poisson regression showing the characteristics associated with discordance severity for those who reported inaccurately

|  | Discordance severity IRR (95% CI) |
| --- | --- |
| *Comparison group* | |
| Alcohol-unaffected[a] | 1.41 (.99-2.01) |
| Problem/very heavy drinker[a] | .95 (.68-1.33) |
| *Demographic characteristics* | |
| Age | .99 (.98-1.00) |
| Black | 1.25 (.98-1.59) |
| Female | .99 (.77-1.28) |
| Has a HS education | 1.08 (.77-1.50) |
| Employed in the past year | **.74** [*] **(.56-.97)** |
| Insured | 1.77 (.76-4.14) |
| *Illness severity* | |
| Lifetime alcohol problem severity (AUD symptom count) | 1.01 (.95-1.07) |
| Past-year alcohol problem severity | |
| DSM-IV AUD[b] | .98 (.74-1.29) |
| Problem drinker[b] | .97 (.70-1.35) |
| At-risk drinker[b] | 1.11 (.72-1.71) |
| Low-risk drinker[b] | 1.08 (.57-2.05) |
| Abnormal cognitive status | 1.40 (.97-2.04) |
| Chronic medical disorder | 1.25 (.92-1.69) |
| Number of healthcare visits | **1.07\*\*\* (1.06-1.09)** |

IRR=Incidence-rate ratio. CI=Confidence interval.

[*] $p<.05$,

[****] $p<.001$

Model statistics: $\chi^2$ (16, N=157) = 152.5, $p<0.0001$. Pseudo $R^2=0.21$

[a] Reference group was those with stably-diagnosed alcohol use disorder

[b] Reference group was those who abstained from alcohol in the past year