Institute for Policy Research
Northwestern University
Working Paper Series

WP-14-10

Measuring Drug and Alcohol Use Among College Student-Athletes

James Druckman
Payson S. Wild Professor of Political Science
Associate Director and Faculty Fellow, Institute for Policy Research
Northwestern University

Mauro Gilli
PhD Candidate in Political Science
Northwestern University

Samara Klar
Assistant Professor of Political Science
University of Arizona

Joshua Robison
Postdoctoral Researcher, Department of Political Science and Government
Aarhus University, Denmark

Version: October 2014

DRAFT
Please do not quote or distribute without permission.

Abstract

Objective – Few issues in athletics today receive more attention than drug and alcohol usage, especially when it comes to college athletics. The researchers seek to correctly address self-report biases related to banned drug usage and heavy drinking.

Methods – The researchers employ an experimental measurement technique.

Results –Their results suggest that an overwhelmingly greater percentage of student-athletes from a major conference knowingly engage in these two behaviors than self-reports indicate. Specifically, the researchers find 37% of respondents report having knowingly taken banned performance enhancing drugs (compared to 4.9% who directly admit to doing so when asked), and 46% consumed more than five drinks in a week (compared to about 3% who openly admit to doing so).

Conclusions – The researchers provide clear evidence for the tremendous extent of self-under-reporting when it comes to drug and alcohol usage among college athletes.

Drug and alcohol use by college students is a frequently debated and often controversial topic. This subject has received particular attention when it comes to student-athletes. Evidence of the importance of assessing drug and alcohol usage among student-athletes is exemplified by a 2012 NCAA report whose "primary objective [was] to update NCAA policy makers with both current and historical information concerning levels of drug and alcohol use by student-athletes within college athletics" (2012: 4). In this paper, we employ an experimental technique that allows us to offer a more accurate assessment of usage than extant studies provide. We begin in the next section with a literature review that leads us to an explication of our approach. We then present results from our survey. Our evidence demonstrates that the commonly used self-report method for estimating drug and alcohol use found in existing studies, including in the aforementioned NCAA report, immensely understates usage.

The Challenge of Measuring Drug and Alcohol Usage

To our knowledge, there is surprisingly little written on drug use among college student-athletes and, when it comes to student-athletes' own input on this controversial issue, the literature is scarce. We have identified those few instances in which student-athletes' attitudes are measured.[1] While existing studies on this subject are illustrative of college athletes in many ways, the nature of the samples used and the method for measuring usage limit what can be said about the extent of drug and alcohol use. For example, Buckman et al. (2008) find that among male student-athletes, 9.7% say they use "banned performance-enhancers" and 55.8% say they used "performance-enhancing drugs" (which might include legal nutritional supplements). Among female student-athletes, no one said they use "banned performance enhancers" and 29.8% said they used "performance-enhancing drugs." While these are intriguing and important findings, the sample is of limited generalizability since it comes only from those who took part in a mandatory alcohol education program. Green et al. (2001) survey student-

[1] Some of this work looks at other drugs that we do not study, such as chewing tobacco and marijuana, so we do not discuss those studies here.

athletes in Divisions 1, 2, and 3 and find 80.5% use alcohol, but the specifics of the survey are unclear and the survey also was part of a NCAA-sponsored project, for which research teams conducted the survey at each participating school. While this result is clearly important evidence, the way the data was collected creates the possibility that demand effects influenced the validity of usage estimates. For instance, the presence of NCAA authorities during the administration of the survey may have had a substantial influence over respondents' candor, especially given usage was measured via self-reports (also see Wechsler et al. 1997 who similarly rely on self-reports in a study of alcohol use).[2]

Perhaps the most impressive and exhaustive survey of athlete drug use was done by the NCAA (2012) itself in 2009. They drew a stratified sample of institutions from all 1,076 active member institutions of the NCAA and surveyed three pre-specified teams per school with an ultimate sample of 20,474 respondents. Their survey took several steps to ensure anonymity such as providing a pre-addressed and stamped envelope for return to a third party vendor and they did not ask for identifying information from the respondent. The survey asked about a host of drug and alcohol behaviors, finding, for example, that only .4% of respondents report using anabolic steroids within the last 12 months while and over 50% of respondents indicate using alcohol in the past year. The NCAA survey provides vital information. However, like the other studies described above, the NCAA's survey relied on self-reports of behavior which may lead to underreports even with the survey's efforts to ensure anonymity. Indeed, the report acknowledged that (5) "Even with these measures to ensure anonymity, self-report data of this kind can be problematic due to the sensitive nature of the issues. Therefore, absolute levels of use might be underestimated in a study such as this."

In sum, while research to-date provides valuable information, it is plagued by the non-trivial threat of arriving at substantial understatements of usage. Reliance on self-reports leads to under-

[2] Other studies such as those by Lisha and Sussman (2010), Terry and Overman (1991), Ford (2007), and Tricker and Connolly (1997) explore drug and/or alcohol use among athletes (and some non-athletes) but they do not look directly at distinct types of behaviors as we do below.

reporting due to social-desirability and threat of disclosure influences (Tourangeau and Smith 1996, Tourangeau et al. 2000). The former refers to respondents' hesitation to provide an answer that may be deemed as socially unacceptable (e.g., that violates expectations or norms). The latter, meanwhile, occurs when there are "concerns about the possible consequences of giving a truthful answer should the information become known to a third party… [Such a] question … raises fears about the likelihood or consequences of disclosure of the answers to agencies or individuals not directly involved in the survey. For example, a question about use of marijuana is sensitive to teenagers when their parents might overhear their answers" (Tourangeau and Yan 2007: 860). Questions about drug or alcohol usage in general have long been noted as carrying with them social desirability and threat of disclosure problems. For example, Tourangeau and Yan state, "To cite just one line of research… studies that compared self-reports about illicit drug use with results from urinalyses … found that some 30%–70% of those who test positive for cocaine or opiates deny having used drugs recently. The urinalyses have very low false positive rates… so those deniers who test positive are virtually all misreporting" Tourangeau and Yan 2007: 859).

When it comes to student-athletes and drugs/alcohol usage, there is undoubtedly a threat of disclosure issue such that if these student-athletes were discovered to be using banned substances or drinking heavily, they could be prevented from participating in their sport according to NCAA rules. Specifically, the NCAA bans a number of substances including anabolic agents, stimulants, and street drugs; individuals identified as using such substances are banned from participation.[3] While the NCAA only has a limited ban on alcohol usage, it explicitly warns against over-usage in stating: "The following is a list of substances that are commonly abused, and how they can impact a student-athlete's performance and eligibility. **Alcohol**: Alcohol is a nervous system depressant. At high dosages, effects include mood swings, impaired judgment and inability to control motor functions. Alcohol can impair an

[3] See: http://www.ncaa.org/wps/wcm/connect/public/NCAA/Health+and+Safety/Drug+Testing/Resources/

athlete's performance through dehydration, depleting vital nutrients and interfering with restful sleep and recovery."[4] This statement makes reporting use socially undesirable (e.g., it would be violating a possible norm of avoiding any product that may harm performance). Moreover, it may be potentially threatening for athletes to over-drink since their individual school or conference may enforce distinct policies that could put caps on alcohol usage. It is for these reasons that the literature on under-reporting often accentuates biases in self-reported drug and alcohol usage, as the aforementioned NCAA report explicitly recognizes (Tourangeau and Yan 2007: 860). Our goal is to remedy this underreporting problem and identify more accurate rates of usage by employing a procedure that has been shown to overcome underreporting challenges.[5]

There are various ways to elicit more accurate responses (e.g. minimize under-reporting), including the previously discussed anonymity approach employed by the NCAA (for a fuller discussion, see Traugott and Lavrakas 2008). However, perhaps the most powerful approach, and the one we pursue, is called the list experiment or item count technique. This approach has been employed to gauge racial prejudice, homophobia, and substance abuse in other populations than our focus, where is has not been used (e.g., Kuklinski et al. 1997, Druckman and Lupia 2012, Coffman et al. 2013). The technique provides a solid estimate of aggregate responses, although it does not allow for individual level analyses (and again we are unaware of it being employed as we do below when it comes to college athletics).

In this approach, the researcher randomly divides respondents into two groups: one treatment and one control. The respondents in the treatment count the number of arguments with which they agree (or disagree/cause them to be upset) among the (for example) *four* arguments listed in the questionnaire. Of those four arguments provided, one addresses an item of social undesirability (e.g., racism or, in our

[4] See: http://www.ncaa.org/wps/wcm/connect/public/NCAA/Health+and+Safety/Drug+Testing/Resources/Commonly+Abused+Substances

[5] The problem of underreporting also occurs in related areas of mental illness (e.g., Calhoun et al. 2000) and morbidity and health care (e.g., Ansah, et al. 2013, Khantzian 1997, Mountjoy et al. 2010).

case, drug usage). By contrast, respondents in the control group are provided with the same question, except that their argument pool is only comprised of, for example, *three* arguments (e.g., all but the socially undesirable item). Random assignment to the control and treatment groups means that the two groups should be equivalent, on average, in how they answer the items presented on both forms. In turn, this allows for an unbiased estimate for the proportion of respondents who have the socially undesirable trait by subtracting the average number of agreement in the control group from the treatment group.

One notable application is Kuklinski et al. (1997) who employ a list experiment to elicit the extent to which citizens are willing to admit racial anxiety or animus. In the experiment, subjects are presented with a list of items and are asked "how many of them upset you?" Some subjects randomly were assigned to assess a total of three items (e.g., increased gasoline tax, athletes receiving millions of dollars, corporations polluting the environment). Others receive a four-item list where the added item is "a black family moving in next door." Kuklinski and his colleagues report that, among white survey respondents in the American south, the four-item group reported an average 2.37 items made them upset, compared to 1.95 items in the three-item group. Since the groups are otherwise identical, the implication is that 42% of these respondents (i.e., (2.37-1.95)X100) are upset by the thought of a black neighbor. By contrast, when subjects were asked this question directly, only 19% of respondents claimed to be upset. More recently, the National Bureau of Economic Research released a list experiment regarding sexual orientation among Americans (Coffman et al. 2013). They report that the use of a list experiment indicates "substantial underestimation" of non-heterosexuality in conventional surveys. Survey experiments such as these can help us observe opinions that citizens do not readily express due to social desirability and/or threat of disclosure problems. Note too that the experimental (random assignment) nature of this measure means that multivariate analyses are not needed as the groups are, on average, equivalent, and thus the focus is on average percentage differences.

Considerable research shows that list experiments reveal a clear *causal* dynamic of under-reporting. Indeed, differences between the groups have been found to *not* stem from *measurement error*. This argument is supported by three types of evidence. First, studies that have available validation data show that reliance on self-reports, even when coupled with assurances of anonymity as found in the NCAA report cited earlier, generate substantial underreporting of behaviors in comparison to estimates generated by list-experiments; this difference is substantial and is on the order of 40% (see Tourangeau et al. 2000). Second, this argument is consistent with Tourangeau and Yan's (2007: 872) finding that "the use of item count techniques [i.e., list experiments] generally elicits more reports of socially undesirable behaviors than direct questions" (also see Blair and Imai 2012: 47-48 for a long list of examples that employ this approach in other domains). Finally, Kiewiet de Jonge and Nickerson (n.d.) directly investigate the possibility that the added item found in the treatment version of the list experiment by itself leads to a higher number of responses. Their results "imply that the ICT [item count technique] does not *overestimate* socially undesirable attitudes and behaviors and may even provide conservative estimates" (4). In short, they find that there is no evidence that the differing lengths of the lists generate any measurement bias, and instead, differences come only from the experimental treatment of the added "undesirable" item (also see Himmelfarb and Lickteig 1982, Tourangeau et al. 2000: 278, Lensvelt-Mulders et al. 2005, Tourangeau and Yan 2007: 872 for more confirmatory evidence along these lines). Finally, we will later provide direct evidence that measurement error is unlikely since the two groups responded to direct self-report questions in proportions that do not significantly differ and thus the treatment group was not *per se* more likely to count the extra item.

Our causal claim, which is supported by a wealth of prior work as just discussed, is that social desirability and disclosure issues cause under-reporting in direct self-reports relative to a list experiment. Again, this is so because the experimental (random assignment) nature of the approach means the groups

are on average equivalent so any difference in responses is due to distinctions in treatment (see Druckman et al. 2011 for details on the experimental approach and the need for only proportion or mean comparisons between groups and not multi-variate analyses). In short, differences reveal a causal source of under-reporting.

Data and Methodology

Our survey focuses on the NCAA Big Ten conference, which is located primarily in the Midwest, with Nebraska as the western-most point and Penn State to the east (circa 2013, which is relevant since the conference is expanding in 2014). Despite its name, the Big Ten included, at the time of our survey, twelve major universities, all of whom compete in Division I NCAA Athletics. While we recognize the limitations of restricting our sample to only one conference, the Big Ten conference is a strong starting point as it includes a large amount of variance among Universities and includes schools that recruit nationally (for another fruitful study of a single conference, see Fountain and Finley 2009).

In the spring of 2012, we accessed the athletic websites of all twelve Big Ten schools and obtained the full rosters for all sports at every school. We then accessed each school's website to locate and record the email address of every student-athlete listed on those rosters. This information was publicly available at all schools except for the University of Nebraska. We contacted officials at the University of Nebraska to obtain directory information for their student-athletes but were declined and thus they are excluded from our sample.

Overall, we located 6,375 names on rosters. We found no e-mails for 479 student-athletes and subsequently we sent out 5,896 e-mails. Of them, 1,803 bounced back as no longer in service (which could be due to the students no longer being enrolled, database errors, website errors, or some other reason). Thus, we successfully sent a total of 4,093 e-mails that, to our knowledge, reached their intended targets. We also sent out one reminder to all respondents. Sample size varied across schools, in

part due to variations in the number of sports each school sponsors (e.g., Ohio State fields 37 total teams, Michigan has 27 teams, while Northwestern has just 19 teams). We received 1,303 responses leading to response rate of 1303/4093 = 31.8%. This rate exceeds the typical response rate in e-mail surveys of this length, especially those that do not employ incentives (see Couper 2008: 310, Shih and Fan 2008, Sue and Ritter, 2007: 36 for discussion of typical response rates on similar surveys).[6]

While our sample may not be perfectly representative of Big Ten student-athletes, it provides a telling view of drug and alcohol use among student-athletes given the diversity of the schools sampled and given that we have no reason to suspect they differ in terms of reporting relative to other conferences/sports.[7] Additionally, the experimental nature of our key measurement approach means that obtaining a perfectly representative sample is of much less importance than is the random assignment of our experimental treatment between groups (for an extended discussion why this is the case, see Druckman and Kam 2011 who show that given sufficient variance, which we have, experimental findings are robust sampling considerations aside). In short, our sample permits us, as far as we know, to carry out the first study of its kind.

Results

Before turning to our list experiments, we first compare some of our own self-report measures with those from the annual College Senior Survey (sponsored by the UCLA Higher Education Research

[6] While we found notable variance in the success of our e-mails reaching their targets (i.e., not bouncing back), of the 4,093 e-mails that were ostensibly received, we found near identical response rates across Universities with a minimum response rate of 29.59% at Michigan State, a maximum rate of 35.57% at Wisconsin, and an average response rate of 31.79% (standard deviation = .018). In terms of the sample make-up, the following lists the number of respondents and the percentage of the sample that came from each University: Indiana (128; 9.82%), Ohio State (122; 9.36%), Illinois (104; 7.98%), Minnesota (120; 9.21%), Michigan State (100; 7.67%), Purdue (100; 7.67%), Iowa (110; 8.44%), Wisconsin (154; 11.82%), Northwestern (122; 9.36%), Pennsylvania State (116; 8.90%), and Michigan (127; 9.75%). Our responses came from over twenty-four different sports; the specific breakdown is available from the authors. The highest responding sport was track and field which made up 15.12% of the respondents, followed by swimming with 12.97% of the respondents. All other sports consisted of less than 10% of the respondents.

[7] The demographics of our sample are as follows: 60.94% female; 6.29 African American; 27.5% in the first year of school, 26% in their second year, 23.5% in their third year; 22% in the fourth year, and less than 1.0% in years beyond four years; 4% having family incomes below $30,000, 16% between $30,000 and 69,999, 26% between $70,000 and $99,999, 35.5% between $100,000 and $100,999, and 18.5% $200,000 or over; and 51% on athletic scholarship.

Institute; see Franke et al. 2010); the Senior Survey provides an important baseline of comparison between athletes (our survey) and a sample of largely non-athletes.[8] Indeed, the vast majority of the UCLA respondents are, based on probability, non-varsity athletes, given that the sample includes 24,457 individuals from 111 colleges and Universities. For a baseline, we used *identical* questions to those employed by the UCLA survey.

In Table 1 (where we use the label "Ath" for our student-athlete survey and "Gen" for the general survey), the first column lists the question in both surveys. The other columns list the results from the general survey (Gen) and our survey (Ath), with N/A indicating that there were no other response categories on that question.[9]

[Insert Table 1 About Here]

What we clearly find is that, relative to the general student survey, student-athletes are substantially less likely – *in self-reports* – to drink beer, liquor (in general or over the two weeks preceding the survey), and to frequently "party", which can be defined as a social gathering that "typically involves eating, drinking, and entertainment" (Oxford Dictionaries). Nearly 75% of the general student population say they "frequently" or "occasionally" drink beer whereas only 46% of our student-athlete sample say the same ($z = 20.55$; $p < .01$). Similarly, 84% of the general population report drinking wine or liquor "frequently" or "occasionally" over the previous year but only about 36% of student-athletes do so ($z = 39.58$; $p < .01$). In the two weeks prior to the survey, about 53% of the general sample state they partied for 3 or more hours, whereas only 38% of student-athletes do ($z = 9.05$; $p < .01$).[10]

[8] We use results from 2009, as these were the most recent available data. We did not use the freshmen survey as it would provide an incorrect point of comparison since nearly 75% of our respondents were not freshmen.

[9] In our survey, respondents always had the choice to not respond to a particular question and thus the sample size varies across different questions for this reason.

[10] Our main concern is with measuring overall levels of alcohol and drug use, but we also found some intriguing differences between male and female respondents on these items. Specifically, we find that male respondents reported significantly higher rates of drinking and partying and were significantly more likely to admit taking banned drugs (results available from

We speculate that the vast differences in self-report stem from two factors. First, student-athletes may genuinely spend less time engaging in these activities, due perhaps to the fact that they spend much of their time either in classes, practicing their sport, or receiving tutoring. Second, student-athletes may be increasingly aware of the negative influence that alcoholic consumption has on their athletic performance, particularly since on-campus education programs highlight the impact of drinking and other related behaviors on performance. They may share concerns about drugs or perhaps worry about being caught for using drugs banned by their University sports department, their conference, or the NCAA (i.e., threat of disclosure).[11] Of course, the lack of individual level data from the UCLA survey prevents us from making causal claims about the difference between athletes and non-athletes. However, our evidence is suggestive.

Estimating Drug and Alcohol Use with List Experiments

We now turn to our list experiment to gauge the aggregate usage of drugs and alcohol; as explained, this approach is a fruitful one for addressing documented under-reporting in these areas due to social desirability and threat of disclosure issues (e.g., Tourangeau and Smith 1996). As far as we know, we are the first to apply this method to the domain of student-athlete drug use and drinking.

Drug Use[12]

We use the list experiment approach to estimate the percentage of student-athletes who knowingly take banned performance enhancing drugs and consume over 5 alcoholic drinks per week –

---

authors). This is in line with Wechsler et al. (1997) and Buckman et al. (2008). For additional analyses on gender, see the online appendix at: http://faculty.wcas.northwestern.edu/~jnd260/publications.html

[11] We asked two other questions but only to our athletes. First, we asked how worried their about complying with NCAA rules concerning drugs, etc. We found that 48% were slightly, moderately, or extremely worried. Finally we asked about caffeine consumption, in part, because NCAA rules on what is allowable is unclear on this question (as they give amounts in micrograms which are contingent on BMI and other factors) and, as reflected by this reality, we found that nearly 70% student athletes seem to exceed what is allowable, even in self-reports.

[12] As is typical when any measure involves a sensitive behavior, there is an increase in item non-response (Tourangeau and Yan 2007), and thus our N shrinks for the list experiments. Importantly, however, we found no significant differences in terms of respondents who did versus respondents who did not complete the list experiments in terms of income, gender, scholarship type, or party identification. The only variable that distinguished the two groups is year in school which may reflect those toward the end of their college careers as being less concerned about disclosure and thus more likely to respond.

both behaviors to which student-athletes may be extremely hesitant to admit on a self-reported survey. We asked all respondents: "Below is a list of activities that you may have engaged in over your time in college. We are interested in HOW MANY of these activities you have engaged – NOT which ones. Thus, please simply choose the number at the end of the list." All respondents were provided the list of the following four items: "Sustained an injury during a practice or game that prevented you from playing at least one other game;" "Joined a social club whose majority of members did/does not include varsity athletes;" "Skipped a class because you felt so tired from a practice or a game;" and "Was unable to take a class that you hoped to take because of your practice or game schedule." We randomly assigned respondents to receive either just those 4-items or that list with the following key 5th item: "Knowingly took a drug banned by the NCAA that may improve your athletic performance." Thus, respondents who received the 4-item list (i.e., the control group) could provide an answer between 0 and 4, while those who received the 5-item list (i.e., the treatment group) could provide an answer between 0 and 5.[13]

Notice our *focus* is on using drugs that are *knowingly banned*. This differs from some prior work that looks at drug usage more generally or nutritional supplements which may not be banned. In light of the previously discussed work on this approach, if we find differences between the groups, then we will have *clear* evidence that self-reports *cause* under-reporting. Further, as mentioned, the experimental measure means we need only compare averages; indeed, we checked and confirmed the success of random assignment in both experiments meaning any differences stem from the treatment and not another variable.[14]

[13] The NCAA provides a list of banned drugs at: http://www.ncaa.org/wps/wcm/connect/public/NCAA/Health+and+Safety/Drug+Testing/Resources/NCAA+banned+drugs+list

[14] Nonetheless, we did perform multivariate analyses and, as expected for both our list experiments, the experimental condition was by far the most predictive element. We also checked for collinearity and found no variables (i.e., our demographic variables) coming even close to being collinear (e.g., they are not highly correlated).

The results are stark, with the control group registering a mean response of 3.31 (std. dev. = .78; n = 553) while the treatment offering a mean response of 3.68 (.98; 510). This is a significantly marked difference (t_{1061} = 6.90; p< .01). Substantively, it suggests that 3.68-3.31 = 37% of respondents have knowingly taken banned drugs. Remarkably, we also asked at the end of the survey, "Since you started playing sports in the NCAA, have you ever knowingly taken a drug banned by the NCAA that may improve your athletic performance?" To this question, only 4.9% said that they had (with virtually no difference between experimental groups; note that this question was not asked directly in the UCLA survey so does not appear in Table 1).[15] In short, this indicates substantial social desirability and/or threat of disclosure bias. Regardless of how one interprets the moral implications of this finding, if there are any, the central point is that self-reports of drug use among student-athletes are misleading and vastly understate usage.[16] We have no way of knowing specifically what these banned drugs are, as they could range from something as serious as blood doping or something more innocuous such as high concentrations of caffeine (as asked about above, see note 11).[17]

Alcohol Use

We used the exact same technique to explore heavy drinking – another behavior that student-athletes may tend to understate, for potential reasons including the detrimental effects on athletic performance and the possible stigmatization by coaches, trainers, or fellow athletes. We follow the

[15] As mentioned above, empirical evidence of the difference not stemming from measurement error would come in the form of the two experimental groups not significantly differing in terms of self-reports. This is exactly what we find; 5.35% responded yes for the four item group compared to 4.50% for the five item group (z = .64; p < .30).

[16] We also looked specifically at sports where drugs have become an issue or there are ostensible advantages to building strength; these include football, wrestling, rowing, track and field, ice hockey, and water polo. We found larger differences in this subsample with a control mean of 3.20 (.77; 216) and a treatment group mean of 3.77 (1.00; 196) (t_{412} = *6.52*; p < .01). In short, 56% of those participating in football, wrestling, rowing, track and field, ice hockey, and water polo knowingly took banned drugs, 29% higher than among all athletes.

[17] Despite our prior point about, as is typical, not needing multivariate results, we did regress the count on all our demographic variables. We found the experimental condition had the strongest and most consistent effect with all other variables being insignificant other than those later in school tending to report a higher number of activities. Again, the nature of the experiment means these results should not be seen as causal *per se*; the experiment only allows us to make clear statements about underestimates.

UCLA survey and define heavy drinking as consuming more than 5 drinks per week. For this experiment, we again began with: "Below is a list of activities that you may have engaged in over your time in college. We are interested in HOW MANY of these activities you have engaged – NOT which ones. Thus, please simply choose the number at the end of the list. The respondents randomly assigned to the four item group received the following items: "Your choice of which University to attend was determined largely by the sports opportunities (e.g., it weighed in at least in 50% of your decision);" "Stayed up past 1AM, on average, during the season of your sport(s);" "Plan to continue playing your sport after college, although not necessarily on the professional level;" and "Play other sports during the school year at least once a month." The randomly assigned treatment group additionally received this item: "In the typical week during the past academic year, consumed more than five alcoholic drinks (A drink can be a 12-ounce beer or wine cooler a 4 ounce glass of wine, or a shot of liquor either straight or in a mixed drink.)"

Again, we find a statistically and substantively large difference, with the control group reporting a mean of 2.76 (std. dev. = .85; n = 544) and the treatment group a mean of 3.22 (1.10; 506) (t_{1048} = 7.61; p< .01). Perhaps surprisingly, this is even a larger difference than our drug experiment. Substantively, it suggests that (3.22-2.76)X100 = 46% consumed more than five drinks a week. While our experimental item did not perfectly match the UCLA survey (see Table 1), those results suggest that less than 3% of athletes admitted to drinking over 5-times every *two weeks* (and the experimental groups do not reach significant difference between one another; p<.25). This is, again, *causal* evidence of under-reporting given the work that shows this approach reveals clear causal under-reporting of self-reports versus our approach (even with the 4 versus 5 item comparison – as explained above).[18]

[18] We again ran a regression with experimental condition being the strongest predator; otherwise women reported fewer activities, and higher income individuals reported more activities for the drinking list experiment. Details on this and the drug regression are available from the authors.

Conclusion

The magnitude of under-reporting on drug and alcohol usage matches the percentages often found in self-reports versus list experiments on distinct topics. As mentioned, Kuklinski et al. (1997) find evidence that 42% of white southerners express prejudice using a list experiment compared to just 19% when asked directly for a 23% swing. Similarly, Gonzalez-Ocantos et al. (2012: 210) find a 24% swing in those admitting to receiving gifts or personal favors when it comes to Nicaraguan municipal elections. Given these similarly large shifts, we find our results to be credible.

We recognize the constraints of our survey in terms of a potentially limited sample that represents only one collegiate athletic conference at one time period. Nevertheless, the evidence is overwhelming in demonstrating substantial banned drug and alcohol usage (including heavy drinking) among college student-athletes. In short, our results show self-reports *cause* underreporting. That the percentage is greater than self-reports is not a surprise, but the pronounced gap is glaring and suggests usage may be much more widespread than many believe. This result should be troubling to readers, regardless of whether they view drug use as worrisome. For those readers who view drug use as problematic, these results suggest that the problem is much direr than existing data based on self-reports reveal. On the other hand, if one does not see drug use as problematic, then our findings remain important as they provide a useful method for more accurately gauging drug use and for exploring other behaviors affected by social desirability and/or the threat of disclosure bias that could benefit from greater monitoring or education. Further research should try to correctly identify what explains the difference between self-reports and our results – for example, is there a lack of awareness regarding NCAA rules and regulations, or are students rather knowingly violating these terms? What is clear from this study is that measuring these trends requires subtle techniques to provide the most accurate information possible, if so desired.

We realize a drawback of our approach is that it only allows for aggregate rather than individual level findings. Thus, the central point for future work is that the use of self-reports is problematic and that researchers need to take strides to address social desirability and threat of disclosure biases as otherwise estimates will be biased downwards. While a list experiment is one powerful method for correcting for this downward bias, other techniques –such as normalization (i.e. telling respondents undesirable behaviors are common or not threatening to report), sealed booklets, and implicit measures – may also be attractive options for the more precise measurement of these behaviors (see Tourangeau et al. 2000 for greater discussion). The relative merits of these techniques for estimating drug and alcohol use, particularly among this population, are currently unknown so future work can fruitfully compare them against one another until the ideal tactic is discovered and can be put to use in tracking trends and arriving at accurate estimates.

Table 1: Student-Athlete and Student Self-Reported Behaviors

Question	***Response Categories***						
During the past academic year, how often have you: Drank beer. (N for athletes = 1041)	*Frequently* *Ath.: 30.5* *Gen: 33.4%*	*Occasionally* *Ath.:* 15.9 *Gen: 41.4%*	*Not at all* *Ath.:* 53.7 *Gen: 25.1%*	N/A	N/A	N/A	N/A
During the past academic year, how often have you: Drank wine or liquor. (N = 1035)	*Frequently* *Ath.: 20.4* *Gen: 31.5%*	*Occasionally* *Ath.: 15.56* *Gen: 52.5%*	*Not at all* *Ath.: 64.06* *Gen: 15.9%*	N/A	N/A	N/A	N/A
Think back over the past two weeks. How many times in the past two weeks, if any, have you had five or more alcoholic drinks in a row?[A] (N = 1042)	*None* *Ath.: 53%* *Gen: 44.7%*	*Once* *Ath.: 20.6%* *Gen: 14.7%*	*Twice* *Ath.: 14.2%* *Gen: 13.6%*	*3-5 Times* *Ath.: 9.3%* *Gen: 16.8%*	*6-9 Times* *Ath.: 2.11%* *Gen: 6.8%*	*10 or more times* *Ath.: .7%* *Gen: 3.4%*	N/A
During the past academic year, how much time did you spend during a typical week partying? (N = 1042)	*None* *Ath.:18.9%* *Gen: 19.7%*	*Less than 1hr* *Ath.: 22%* *Gen: 11.3%*	*1-2 Hours* *Ath.: 21%* *Gen: 16.4%*	*3-5 Hours* *Ath.: 25.4%* *Gen: 24.7%*	*6-10 Hours* *Ath.: 10.1%* *Gen: 17.2%*	*11-15 Hours* *Ath.: 1.9%* *Gen: 6.1%*	*16-20 Hours* *Ath.: .4%* *Gen: 2.3%* *Over 20 Hours*

							Ath.: .5% *Gen:* 2.3%

[A] The question continued with: "(A drink can be a 12-ounce beer or wine cooler a 4 ounce glass of wine, or a shot of liquor either straight or in a mixed drink.)"
[B] In the question, all words were not underlined but the word "not" was in those response categories.
N/A = not a response category.

References

Ansah, Evelyn Korkor, and Timothy Powell-Jackson, 2013. "Can we trust measures of healthcare utilization from household surveys?"*BMC Public Health* 13: 853.

Blair, Graeme, and Kosuke Imai. 2012. "Statistical Analysis of List Experiments." *Political Analysis* 20: 47-77.

Buckman, Jennifer F., Robert J. Pandina, Helene R. White, and David A. Yusko. 2008. "Alcohol, Tobacco, Illicit Drugs, and Performance Enhancers: A Comparison of Use by Student Athletes and Non-Athletes" *Journal of American College Health* 57(3): 281-290.

Calhoun, Patrick S., William S. Sampson, Hayden B. Bosworth, Michelle E. Feldman, Angela C. Kirby, Michael A. Hertzberg, Timothy P. Wampler, Faye Tate-Williams, Scott D. Moore, and Jean C. Beckham. 2000. "Drug Use and Validity of Substance Self-Reports in Veterans Seeking Help for Posttraumatic Stress Disorder." *Journal of Consulting and Clinical Psychology* 68(5): 923–927.

Coffman, Katherine B., Lucas C. Coffman, and Keith M. Marzilli Ericson. 2013. "The Size of the LGBT Population and the Magnitude of Anti-Ga Sentiment are Substantially Underestimated." NBER Working Paper No. 19508.

Couper, Mick P. 2008. *Designing Effective Web Surveys.* New York: Cambridge University Press.

Druckman, James N., Donald P. Green, James H. Kuklinski, and Arthur Lupia. 2011. *Cambridge Handbook of Experimental Political Science*. New York: Cambridge University Press.

Druckman, James N., and Cindy D. Kam. 2011. "Students as Experimental Participants." In *Cambridge Handbook of Experimental Political Science*, eds. James N. Druckman, Donald P. Green, James H. Kuklinski, and Arthur Lupia. New York: Cambridge University Press.

Druckman, James N. and Arthur Lupia. 2012. "Experimenting with Politics." *Science* 335: 1177-1179.

Ford, Jason A. 2007. "Substance Use Among College Athletes: A Comparison on Sport/Team Affiliation." *Journal of American College Health* 55(6): 367-373.

Fountain, Jeffrey J., and Peter S. Finley. 2009. "Academic Majors of Upperclassmen Football Players in the Atlantic Coast Conference: An Analysis of Academic Clustering Comparing White and Minority Players." *Journal of Issues in Intercollegiate Athletics* 2: 1-13.

Franke, Ray, Sylvia Ruiz, Jessica Sharkness, Linda DeAngelo, and John Pryor. 2010. "Findings from the 2009 Administration of the College Senior Survey (CSS): National Aggregates" Los Angeles: Higher Education Research Institute, University of California Los Angeles. Available at: (http://www.heri.ucla.edu/PDFs/pubs/Reports/2009_CSS_Report.pdf).

Gonzalez-Ocantos, Ezequiel, Chad Kiewiet de Jonge, Carlos Meléndez, Javier Osorio, and David W. Nickerson. 2012. "Vote Buying and Social Desirability Bias: Experimental Evidence from Nicaragua." *American Journal of Political Science* 56(1): 202-217.

Green, Gary A., Frank D. Uryasz, Todd A. Petr, and Corey D. Bray. 2001. "NCAA Study of Substance Use and Abuse Habits of College Student-Athletes" *Clinical Journal of Sports Medicine* 11(1): 51–56.

Himmelfarb, Samuel, and Carl Lickteig. 1982. "Social Desirability and the Randomized Response Technique." *Journal of Personality and Social Psychology* 43: 710-717.

Khantzian, Edward J. 1997. "The Self-Medication Hypothesis of Substance Use Disorders: A Reconsideration and Recent Applications." *Harvard Review of Psychiatry* 4(5): 231-244.

Kiewiet de Jonge, Chad P., and David W. Nickerson. N.d. "Artificial Inflation or Deflation?: Assessing the Item Count Technique in Comparative Surveys." *Political Behavior*. Forthcoming.

Kuklinski, James H., Paul M. Sniderman, Kathleen Knight, Thomas Piazza, Philip E. Tetlock, Gordon B. Lawrence, and Barbara Mellers. 1997. "Racial Prejudice and Attitudes Toward Affirmative Action." *American Journal of Political Science* 41(2): 402-419.

Lensvelt-Mulders, Gerty J. L. M., Joop J. Hox, Peter G. M. van der Heijden, and Cora J. M. Maas. 2005. "Meta-Analysis of Randomized Response Research: Thirty-Five Years of Validation." *Sociological Methods & Research* 33(3): 319–348.

Lisha, Nadra E., and Steve Sussman. 2010. "Relationship of High School and College Sports Participation with Alcohol, Tobacco, and Illicit Drug Use: A Review." *Addictive Behaviors* 35(5): 399-407.

Mountjoy, Margo, Astrid Junge, Juan Manuel Alonso, Lars Engebretsen, Ioan Dragan, David Gerrard, Mohamed Kouidri, Eide Luebs, Farhad Moradi Shahpar, and Jiri Dvorak. 2010. "Sports injuries and illnesses in the 2009 FINA World Championships (Aquatics)." *British Journal of Sports Medicine* 44: 522-527.

National Collegiate Athletic Association (NCAA). 2012. *National Study of Substance Use Trends Among NCASS College Student-Athletes*. Indianapolis: National Collegiate Athletic Association,

Shih, T.H., and X. Fan. 2008. Comparing response rates from Web and mail surveys: A meta-analysis. Field Methods 20(3) 249-271.

Sue, Valerie M., and Lois A Ritter. 2007. *Conducting Online Surveys*. Sage Publications.

Terry, Toni., and Steven J. Overman. 1991. "Alcohol Use and Attitudes: A Comparison of College Athletes and Nonathletes" *Journal of Drug Education* 21(2): 107-117.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. New York: Cambridge University Press.

Tourangeau, Roger, and Tom W. Smith. 1996. "Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context." *Public Opinion Quarterly* 60(2): 275-304.

Tourangeau, Roger, and Ting Yan. 2007. "Sensitive Questions in Surveys." *Psychological Bulletin* 133: 859-883.

Traugott, Michael W., and Paul J. Lavrakas. 2008. *The Voters' Guide to Election Polls*. Lanham MD: Rowman & Littlefield.

Tricker, Raymond, and Declan Connolly. 1997. "Drugs and the College Athlete: An Analysis of the Attitudes of Student Athletes at Risk" *Journal of Drug Education* 27(2): 105-119.

Wechsler, Henry, Andrea E. Davenport, GeorgeW. Dowdall, Susan J. Grossman, and Sophia I. Zanakos. 1997. "Binge Drinking, Tobacco, and Illicit Drug Use and Involvement in College Athletics." *Journal of American College Health* 45(5): 195-200.

Reviewer Appendix on Gender

One consistent finding in some prior work concerns gender differences in these behaviors (e.g., Wechsler et al., 1997, Buckman et al. 2008), such that that males are more likely to drink and use illegal drugs than are females. This is tangential to our study but we provide an appendix here on some details (which can be deleted as it is noted in a note above, or put on-line.)

We explored gender differences in our survey by using multi-variate regressions with a series of control variables, as listed in an above note, and a variable we measured that asks respondents to rate the importance of religion in their life. The full results are available from the authors but here we focus on gender, for the following two reasons: (1) as explained, gender is the one variable that consistently influenced respondents across variables, and (2) this is consistent with the aforementioned work that also finds gender effects. (Aside from gender, we found: religious importance decreased the likelihood of partying, drinking booze, or drinking beer; income increased the likelihood of partying, drinking beer, and drinking in general; and year in school influenced drinking for the assumed reason that seniors may reach the legal drinking age. Other effects were less consistent and again are available from the authors.)

[Insert Table A-1 About Here]

In Table A-1, we present the average score for our key questions, broken out by gender with the statistical significance listed in the last row. We find significant differences in all cases. Specifically, we see that men party significantly more, drink significantly more, and are significantly more likely to admit to taking banned drugs. On the flip side, women display a distinct drinking behavior by consuming significantly more booze (wine or liquor). In short, gender does seem to matter in behavior among student-athletes with men generally engaging in what may be deemed socially less desirable behavior. Of course, many of these results beg the questions of the veracity of responses. Like all prior work, we have relied on self-reports.

Table A-1: Gender Differences [A]

Activity	Average Total	Average for Men	Average for Women	Stat Sig. for Men versus Women
Party (8 point scale)	2.95 (1.41; 1028)	3.21 (1.38; 411)	2.77 (1.40; 617)	$t_{1028} = 5.01$; $p < .01$
Booze (3 point scale)	2.44 (.81; 1021)	2.37 (.83; 410)	2.48 (.80; 611)	$t_{1021} = 2.20$; $p < .05$
Beer (3 point scale)	2.23 (.89; 1027)	2.37 (.78; 410)	2.14 (.94; 617)	$t_{1027} = 4.06$; $p < .01$
Take Banned Drugs (% no)	95.20 (1041)	93% (411)	97% (630)	$z = 3.04$; $p < .01$
How much drink (6 point scale)	1.88 (1.16; 1028)	2.18 (1.28; 410)	1.70 (1.02; 618)	$t_{1028} = 6.63$; $p < .01$

[A]For the "average total" column we only included those who also answered the gender question.