Lifetime consumption of alcohol typically reaches its highest levels during an individual’s late teens and early 20s. Numerous studies document a range of negative consequences of high levels of alcohol consumption, including violence, date rape, accidents, academic problems, and family conflict. The college campus is one setting where the pattern of youthful heavy drinking is felt acutely. College students, on average, drink more than their noncollege peers of the same age and routinely report negative consequences from both their own and others’ drinking. While heavy drinking in college is associated with personality factors such as impulsivity, nonconformity, and depression, contextual factors such as distance from parents, close association with peers, dormitory residence, association with fraternities and sororities, large-group social events, and athletics appear to support and exacerbate heavy drinking as well.

Heavy drinking in college should not be confused with dependent drinking in later life, however. Several studies note significant reductions in heavy drinking in the 20s and 30s. At the individual level, Schulenberg and colleagues followed the drinking patterns of high school seniors for at least 6 years and noted that a subset of individuals (17%) reported more than isolated patterns of heavy drinking over time. Thus, college administration and health officials, who are under increasing pressure to provide both preventive and treatment services as a public health service for college students, are faced with a multifaceted social problem that is common, risky, and limited in time for most but chronic for some. Unfortunately, few interventions have a documented positive impact in changing college drinking behavior. In particular, commonly offered educational programs have little impact. We recently reported on one model of indicated prevention, which involves personalized individual feedback and brief motivational interventions for high-risk students during a single brief, nonconfrontational counseling session. Students who received this preventive intervention reported significantly greater reductions in alcohol-related problems at the 2-year follow-up compared with a randomly assigned control group.

In the current analysis, we examine the natural history of drinking patterns and related problems over 4 years, within both high-risk and normative samples. The preventive intervention is evaluated over this extended period of time. Finally, using individual classification of reliable change, we describe developmental trajectories of drinking among college students who drank heavily in high school and describe rates of clinically significant change.

**METHODS**

**Participants and Recruitment**

All students younger than 19 years who had committed themselves to attending the University of Washington in the fall of 1990 (n = 4000) were mailed a questionnaire during the spring before matriculation. For completing the questionnaire, students were paid $5 and entered into a drawing for a prize. The questionnaire included items about quantity and frequency of alcohol consumed during a typical week and peak alcohol consumption in the past 3 months. The questionnaire also included the Rutgers Alcohol Problem Inventory to assess drinking-related consequences during the previous 3 years. A total of 2041 students (51%) provided complete questionnaires and indicated a willingness to participate in a future study.

From the completed questionnaires, 508 individuals were identified as being at “high risk” by the following criteria: drinking at least once a month and consuming 5 to 6 drinks on at least 1 occasion in the last month, or experiencing at least 3 negative consequences from drinking (Rutgers Alcohol Problem Inventory items) on 3 to 5 different occasions in the previous 3 years. An additional normative comparison sample was selected randomly from the entire pool of responders (n = 151), including 33 persons who...
met high-risk criteria, to track the natural history of changes in drinking behavior within the cohort over time.

Potential participants were invited by letter and by phone. Participants were paid $25 for entering the study, completing the baseline questionnaire and interview, and providing 2 collateral contacts who could confirm the participant’s rates of alcohol consumption. Of those contacted, 348 (68%) who were identified as being at high risk agreed to participate and were randomized into either the intervention or no-intervention control groups; 113 (75%) of those randomly selected for the natural history comparison agreed to participate (28 of the 113 also met high-risk criteria). At baseline assessment in the autumn of their freshman year, high-risk students reported, on average, drinking more than 10 drinks per week during fewer than 2 drinking occasions, reaching a typical estimated peak blood alcohol level of 0.12%. Normative comparison students reported drinking about 5 drinks per week during 1 drinking occasion, reaching a typical estimated peak blood alcohol level of 0.08%. The high-risk sample was 55% female and 84% White. The normative comparison sample was 54% female and 78% White.

Confidence in the representativeness of the sample is potentially limited by the 50% response rate to the initial mailing. However, drinking rates within the normative comparison sample are quite consistent with averages noted in national databases (e.g., Monitoring the Future). The level of drinking among high-risk participants is more extreme than in the normative comparison sample. Rates of drinking were compared: listwise deletion and multiple imputation. Participants with missing data at the 4-year follow-up were eliminated from analyses as missing data. Analyses indicated no differences in parameter estimates between imputation of missing data with list-wise deletion and multiple imputation. Two methods of processing missing data were compared: listwise deletion and multiple imputation. Participants with missing data at the 4-year follow-up were eliminated from analyses as missing data. Missing data occurring at the 1-, 2-, or 3-year assessments were replaced by a multiple imputation method using maximum likelihood estimation.
line mean and standard deviation, yielding relative scores from a normative baseline whose mean equals 0 and whose standard deviation equals 1. We also converted the standardized outcome measures in each drinking domain to 3 unit-weighted factor scores by computing the mean of the standardized relevant measured indicators.

RESULTS

Success of Randomization

No significant baseline differences were observed between prevention and control conditions for alcohol consumption, related consequences, or demographic and individual difference factors.

Longitudinal Attrition

At the 4-year follow-up, 363 of the 433 unique participants (84%) from the 2 high-risk groups and the normative comparison group completed assessments. Complete data sets at baseline and all 4 follow-up periods were provided by 328 participants (76%), and 346 participants (80%) provided data at 4 of 5 time points, including year 4. Consistent with university norms, 53% of the sample reported having graduated at the 4-year follow-up.

Attrition rates at the 4-year assessment did not differ significantly between high-risk and normative comparison groups, or between randomized high-risk group conditions (prevention vs control). Attrition was not significantly related to participant’s sex or to baseline quantity, frequency, or consequences of drinking.

Changes in Drinking and Negative Consequences by Prevention Condition

Three distinct model analyses provided hypothesis tests with restricted maximum likelihood parameter estimates for the effects of group (prevention and control) and time (baseline and 1-, 2-, 3-, and 4-year follow-up) on the 3 major outcome factor scores: frequency, quantity, and negative consequences. The mixed models included specifications for a priori contrasts for adjacent time points (baseline vs 1-year follow-up, 1-year vs 2-year follow-up, etc.), repeated-measures effects for time, time-by-treatment interactions, and random-effects estimates for subjects. A priori contrasts of time-by-treatment effects allow evaluation of increasing or decreasing differences between experimental groups at the 3- or 4-year assessments.

Standardized mean differences from normative baselines among high-risk participants for drinking frequency, drinking quantity, and negative drinking consequences over the 4-year follow-up are listed in Table 1. Figure 1 displays drinking rates and consequences for high-risk participants and normative comparison samples. Significant main effects over time were noted in all 3 analyses. Over 4 years, the magnitude of change was greatest for measures of negative drinking consequences ($F_{4,322} = 45.65$, $P < .001$), compared with those of drinking quantity ($F_{4,322} = 28.22$, $P < .001$) and drinking frequency ($F_{4,322} = 7.98$, $P < .001$), which demonstrated the smallest effect.

<table>
<thead>
<tr>
<th>TABLE 1—Mean Standardized Factor Scores (SD) and Differences (SEM) Between High-Risk Prevention Group and High-Risk Control Group Over 4 Years: Preventive Intervention for College Drinkers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Drinking Factor</strong></td>
</tr>
<tr>
<td>Frequency</td>
</tr>
<tr>
<td>Control</td>
</tr>
<tr>
<td>Prevention</td>
</tr>
<tr>
<td>Difference</td>
</tr>
<tr>
<td>Quantity</td>
</tr>
<tr>
<td>Control</td>
</tr>
<tr>
<td>Prevention</td>
</tr>
<tr>
<td>Difference</td>
</tr>
<tr>
<td>Negative consequences</td>
</tr>
<tr>
<td>Control</td>
</tr>
<tr>
<td>Prevention</td>
</tr>
<tr>
<td>Difference</td>
</tr>
</tbody>
</table>

Note. Factor scores reflect elevations in SD units relative to baseline mean for the normative comparison group. All mean values are significantly greater than 0.0. Difference score = control score minus prevention score, and difference deviations are standard errors as computed by SAS PROC MIXED. Rounding of factor scores creates slight discrepancies in difference scores. Statistical tests of difference scores represent bivariate tests where $H_0: T_{baseline} = T_{prevention}$.

* $P < .05$; ** $P < .01$; *** $P < .001$ for all tests.
Significant prevention group–by–time interactions over 4 years were observed with respect to negative drinking consequences ($F_{3,32} = 2.38$, $P < .05$) and for drinking quantity ($F_{4,32} = 4.33$, $P < .001$), but not with respect to drinking frequency. Thus, from a multivariate perspective, drinking problems declined significantly over time, and the preventive intervention produced significant differences in alcohol use and related problems over 4 years. Group-by-time effects (differential change) are shown in Table 1 to result in significant group differences in negative consequences at all follow-up assessments.

A priori contrasts of between-group differences in change scores (complete case analysis, 2-tailed test) represent group-by-time interactions for each interval and thus reveal when treatment effects occur. Differences in the magnitude of change between the high-risk prevention and high-risk control groups from baseline to 1-year follow-up were evident for frequency ($P = .03$), quantity ($P = .0002$), and negative consequences ($P = .0095$). All other adjacent time change contrasts between groups were nonsignificant ($P > .05$). Thus, the prevention program appears to have its primary effect between the baseline and 1-year assessments. Differences noted at the 3- and 4-year assessments can be interpreted as a continuation of effects noted earlier in the follow-up period.

Trajectories of the normative comparison subjects over time are also evident in Figure 1. For normative participants (separate analysis on all normative control group subjects), drinking frequency was significantly increased relative to baseline at the 3-year follow-up ($t_{39} = 2.18$, $P < .05$) and 4-year follow-up ($t_{39} = 2.51$, $P < .05$). Drinking quantity showed minor variation over time, with the 2-year assessment being significantly above baseline ($t_{39} = 2.18$, $P < .05$). Mean negative consequences of drinking remained quite stable through the college years.

We reanalyzed the drinking patterns described above to evaluate the possible moderating effects of the participant’s sex. Sex did not significantly moderate the prevention effects previously noted for drinking quantity or negative drinking consequences. A 3-way interaction for drinking frequency ($F_{3,32} = 2.22$, $P < .05$) proved difficult to interpret. Examinations...
tion of means suggested that there was greater variability over time among control group males and that, relative to the control condition, the preventive intervention may have reduced drinking frequency among women but not among men.

**Analysis of Individual Change**

Average changes in drinking rates and problems may mask considerable variability in individual responses, and thus they do not provide a measure of individual risk status and do not indicate how many cases of problem alcohol use might be prevented as a result of this prevention program. To describe outcomes and natural history in terms reflective of individual risk status for college students, we computed risk cutpoints and measures of individual change by using an algorithm reported elsewhere. We chose to examine the Alcohol Dependency Scale for the current analysis because it best reflects the negative consequences of drinking (where the prevention program had the greatest impact) and does not contain college-specific items (e.g., slept late for a class) that are less relevant to those who have quit school or graduated by the 4-year follow-up. The risk cutpoint of 5 for the Alcohol Dependency Scale was established empirically to reflect the point at which a subject is equally likely to be a member of the high-risk population or of the normative population containing no high-risk participants (a “functional” comparison). As described in an earlier report, we used an empirical-percentiles approach to determine cutpoints and confirmed the stability of estimates through bootstrapping. A reliable change index was defined as a difference from baseline scores that meets or exceeds 2 standard errors of the estimate of difference scores.

Participants were then categorized on the basis of their baseline scores and 4-year follow-up scores. Owing to the multidimensional definition of “high risk,” not all high-risk participants were above the risk cutpoint for the Alcohol Dependency Scale at baseline. Similarly, the normative comparison group included some individuals with scores above risk cutpoints at baseline. Participants were classified as “resolved” if their score began above the risk cutpoint, changed reliably (more than the reliable change index), and ended below the cutpoint. Participants were classified as “reliably improved” if their score changed reliably in a direction of fewer dependence symptoms but did not cross the cutpoint. Participants were classified as “reliably worse” if their score moved reliably in the direction of more dependence symptoms but did not cross from below to above the risk cutpoint. “New cases” represents those individuals who became reliably worse over time and crossed from below to above the risk cutpoint. If change did not exceed the reliable change index, participants were classified as “no change.”

Table 2 presents change categories, based on the Alcohol Dependency Scale, from baseline to 4-year follow-up for all participants. Within the normative comparison group, which represents the population of students at this university and thus includes some high-risk participants, 22.9% of students with scores below the risk cutpoints at baseline became new cases; the modal trajectory was no change (55.7%). Among the 35 normative comparison students whose scores were above a risk cutpoint at baseline, 8.6% reliably improved and 31.4% were resolved at the 4-year follow-up. Trends for high-risk participants as well as preventive intervention effects can also be observed in Table 2. For example, among high-risk participants with Alcohol Dependency Scale scores above the risk cutpoint at baseline, 32.8% of those in the control condition were resolved at the 4-year follow-up, indicating that roughly a third of those who drink heavily in high school resolve drinking risk over the course of 4 years. This rate improved to 42.6% among those in the prevention condition. Surprisingly, among high-risk participants with Alcohol Dependency Scale scores below the risk cutpoint at baseline, 18.5% of the control group were reliably worse or were new cases 4 years later, compared with 10% of the prevention group. More generalized comparisons also follow from Table 2. Among high-risk participants, 67% of the prevention group had good outcomes over 4 years (resolved, reliably improved, or no change from a baseline score below the risk cutpoint), compared with 55% of controls.

**DISCUSSION**

Central findings suggest that much heavy drinking among college students is transitory, despite some students who report a pattern of continued or worsening consequences over time. Changes in drinking are reflected in specific dimensions of drinking behavior. Compared with a high-risk control sample, in this randomized trial, participants receiving a brief individual preventive intervention had significantly greater reductions in negative consequences that persisted over a 4-year period. Individual change analyses suggest that the dependence symptoms of those receiving the brief intervention are more likely to decrease and less likely to increase.

To understand the impact of the preventive intervention in a broader context, we report the natural history of different dimensions of drinking through the participants’ 4 years of college. The frequency of drinking did not
change dramatically over 4 years for either high-risk or normative samples. Normative comparison participants, representing the general student body, reported slight increases in drinking frequency over time, particularly at the 3-year follow-up, when many in the sample had reached 21 years of age. Drinking frequency among high-risk students declined only minimally over the 4-year period.

Mean drinking quantity and negative consequences increased only marginally within our normative comparison group, suggesting that students in general do not commonly or routinely develop drinking problems during the college years. However, drinking quantity and associated problems declined steadily over time for high-risk students who entered college with a history of heavy drinking. Recruitment of high-risk students was based on drinking during high school, and thus both developmental and statistical trends probably move toward less extreme behavior (but these same individuals reported mean increases in drinking upon entry into college). Four years after matriculation, our high-risk control students continued to drink more frequently than the normative comparison students, but their problem scores had markedly decreased and, although still above those of the comparison group, were much less elevated (standardized difference about 0.62) than they were at baseline (standardized difference about 1.40). This pattern of data is consistent with other longitudinal studies showing that adolescents with problem behaviors (including heavy drinking) remain less conventional than others as they age into adulthood, but do not have worse psychosocial adjustment.

Prevention effects were observed for only some dimensions of drinking behavior. The dimension of greatest interest, negative consequences of drinking, shows the greatest effects. This is important not only because negative consequences measure the degree to which individuals may be harmed as a result of drinking, but also because the preventive intervention targeted individual choices and reduction of risk, rather than drinking rates per se. Our findings are consistent with the goals of harm reduction interventions, approaches that focus on minimizing the harmful effects of high-risk drinking. The duration of our prevention effects is also noteworthy.

Modest differential changes in drinking quantity and frequency, described in our earlier report of 2-year outcomes, do not appear to persist for longer periods of time, yet we found significantly reduced negative consequences 3 1/2 years after the preventive intervention. We are unaware of other studies (much less randomized trials) of prevention efforts among college students that demonstrate such long-lasting effects.

Our analyses of individual change suggest that, regardless of baseline risk status, about 1 in 3 college students, as represented by the normative comparison participants, reliably worsened during the college years. Another third do not change, and a third reliably improve. Among high-risk samples, many more students reliably improve than worsen, and those with scores beginning below risk cutpoints most often report no change in drinking status. Group differences in rates of reliable worsening and improvement among high-risk participants suggest that feedback and advice may function by accelerating a normative developmental process of reducing drinking, as well as by slowing a less typical developmental process that may otherwise lead to an escalation of drinking for some individuals during college years.

The results described in this report should, of course, be interpreted with attention to the inherent limitations of the research method. The study included students from only one large public university. Given variability in drinking across educational settings potentially based on size, private funding, and entry criteria, our results may not be generalizable to all other student populations. Assessment was based on self-report, which could result in inaccurate or socially desirable reporting. Self-report of alcohol use, however, has been shown to be quite accurate in many contexts where no penalties exist for specific responses. Self-report indices could also reflect students' desire to please researchers, and those receiving the feedback may experience a greater demand for such reporting. Our experimental design did not include an attention-only control to test this effect. However, it is difficult to conclude that response biases would result from such a limited contact (1 hour during freshman year and mailed feedback during sophomore year), would extend for 4 years, and would produce treatment effects in one dimension but not another. Also, participants were reminded that collateral reports were obtained at each assessment point to encourage honest responding.

Although our data are consistent with a broad literature showing that brief interventions are effective in reducing alcohol and drug use, to date, we do not know more precisely how these interventions work (our data suggest that developmental processes underlying drinking reduction and drinking escalation might be affected). Nor do we know the critical components for content and delivery of the preventive intervention. Is the primary component merely increased attention to the issue? Can interventions based on this component be conducted by peers in groups or with only mailed feedback? Continued research on these types of issues will facilitate the adoption of effective prevention programming and the continued reduction in harm based on youthful heavy drinking.

About the Authors
John S. Baer and Daniel R. Kivlahan are, and at the time of the study Patrick McKnight was, with the Center of Excellence in Substance Abuse Treatment and Education, VA Puget Sound Health Care System, Seattle, Wash. John S. Baer, Arthur W. Blume, and G. Alan Marlatt are with the Department of Psychology, and Daniel R. Kivlahan is with the Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle.

Requests for reprints should be sent to John S. Baer, PhD, S116-ATC VA Puget Sound Health Care System, 1660 S Columbian Way, Seattle, WA 98108 (e-mail: jshaer@u.washington.edu).

This article was accepted October 4, 2000.

Contributors
J. S. Baer, D. R. Kivlahan, and G. A. Marlatt designed the study, obtained funding, and supervised data collection. A. W. Blume assisted in data management, conducted data analyses, and wrote parts of the “Methods” section. P. McKnight conducted mixed model and individual change data analyses and wrote parts of the “Results” section. J. S. Baer wrote the initial draft of all other sections. D. R. Kivlahan served as primary editor. All authors contributed to editing and revision.

Acknowledgments
This research was supported by grant R37 AA05591 from the National Institute on Alcohol Abuse and Alcoholism.

We gratefully acknowledge the support and assistance in the preparation of this report by Thad Leffingwell, Maxine Follock, Dan Neal, and Lisa Roberts. Others providing essential support for the ongoing management of the research project include Mary Larimer, Jason Kimer, Lori Quigley, and Ken Weingart.


Baer et al. | Peer Reviewed | Research Articles | 1315
References

34. Shadish WR. Planned critical multiple testing: some elaborations. Behav Assess. 1986;8:75-103.
Copyright of American Journal of Public Health is the property of American Public Health Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.