Alcohol dependence symptoms and consumption measures were examined for stability and heritability. Data were collected from 12,045 individuals (5376 twin pairs, 1293 single twins) aged 19 to 90 years in telephone interviews conducted in three collection phases. Phases 1 and 2 were independent samples, but Phase 3 targeted families of smokers and drinkers from the Phase 1 and 2 samples. The stability of dependence symptoms and consumption was examined for 1158 individuals interviewed in both Phases 1 and 3 (mean interval = 11.0 years). For 1818 individuals interviewed in Phases 2 and 3 (mean interval = 5.5 years) the stability of consumption was examined. Heritability was examined for each collection phase and retest samples from the selected Phase 3 collection. The measures examined were a dependence score, based on DSM-IIIR and DSM-IV criteria for substance dependence, and a quantity × frequency measure. Measures were moderately stable, with test–retest correlations ranging from .58 to .61 for dependence and from .55 to .64 for consumption. However, the pattern of changes over time for dependence suggested that the measure may more strongly reflect recent than lifetime experience. Similar to previous findings, heritabilities ranged from .42 to .51 for dependence and from .31 to .51 for consumption. Consumption was significantly less heritable in the younger Phase 2 cohort (23–39 years) compared to the older Phase 1 cohort (28–90 years).

Long-Term Stability and Heritability of Telephone Interview Measures of Alcohol Consumption and Dependence

Narelle K. Hansell,1 Arpana Agrawal,2 John B. Whitfield,1 Katherine I. Morley,1 Gu Zhu,1 Penelope A. Lind,1 Michele L. Pergadia,2 Pamela A. F. Madden,2 Richard D. Todd,2 Andrew C. Heath,2 and Nicholas G. Martin1

1 Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia
2 Department of Psychiatry, School of Medicine, Washington University, United States of America
3 Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, School of Population Health and Applied Genetics Unit, ORYGEN Research Centre, Department of Psychiatry, University of Melbourne, Australia

Received 20 March, 2008; accepted 26 March, 2008.

Address for correspondence: Narelle K. Hansell, Genetic Epidemiology, Queensland Institute of Medical Research, Post Office, Royal Brisbane Hospital, Queensland, Australia 4029. E-mail: Narelle.Hansell@qimr.edu.au

A second literature has for many decades documented significant genetic variance in alcohol consumption patterns (Heath, 1995). Short-term test–retest reliability of consumption measures in adults, typically with reassessment at 1 to 6 weeks, is modest to good for a typical quantity measure (drinks per drinking day: retest correlations .48–.88), better for frequency of consumption (number of days used alcohol: retest correlations .64–.85), and highest for quantity × frequency (i.e., approximating total drinks per week: .73–.91; Friesema et al., 2004; Grant, Harford, Dawson, Chou, & Pickering, 1995; Gruenewald & Johnson, 2006; Webb, Redman, Gibberd, & Sanson-Fisher, 1991). Test–retest correlations are notably lower over longer follow-up periods (see Table 1), though still acceptable reliability (Bucholz et al., 1994; Bucholz et al., 1995) and validity (Hesselbrock, Easton, Bucholz, Schuckit, & Hesselbrock, 1999). However, such case series disproportionately represent severely affected individuals, in whom difficulties in the discrimination between unaffected and affected are minimized (Bucholz et al., 1996). In contrast, in general community samples, the majority of individuals reporting a history of alcohol dependence will be individuals with milder disorder (Heath, Bucholz et al., 1994; Lynskey et al., 2005). A growing literature, initially using latent class analysis (Bucholz et al., 1996; Heath, Bucholz et al., 1994; Lynskey et al., 2005), and more recently Item Response Theory (Saha, Chou, & Grant, 2006) suggests that alcoholism may best be viewed as at least a semi-continuous trait, avoiding an arbitrary dichotomization into unaffected and affected individuals. Such an approach would have the potential to increase the power of genetic research approaches.

Downloaded from https://www.cambridge.org/core. IP address: 54.191.40.80, on 21 Aug 2017 at 15:40:30, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1375/twin.11.3.287

287
substantial, particularly for cohorts that were older at baseline assessment and using frequency or quantity × frequency measures. There is evidence for strong genetic overlap between heaviness of consumption and alcohol dependence risk (Heath & Martin, 1994; Whitfield et al., 2004), suggesting that it may be useful to take advantage of quantitative consumption measures as covariates in genetic research on alcoholism.

In this article, the long-term test-retest reliability and heritability of a dependence symptom count measure and a quantity × frequency consumption measure, will be examined in studies in the Australian twin panel. Their potential informativeness for gene-mapping studies of alcohol use disorder will be evaluated.

**Materials and Method**

**Participants**

Data collected from 12,045 individuals (5376 twin pairs and 1293 single twins) were examined. For 2976 individuals (1099 twin pairs and 778 single twins), data were collected on two occasions. The sample was drawn in large part from earlier studies, as shown in Figure 1, with the current report focused on data collected in three phases of telephone interviews. Phase 1 data were collected from 1992 to 1993, Phase 2 data were collected from 1996 to 2000 and Phase 3 data were collected from 2001 to 2005. These studies are described in Table 2.

Sample sizes for each collection phase and retest are shown in Table 4, in addition to the age of participants, the proportion of each sample that is female, and the proportion that is monozygotic (MZ). For the Phase 1/Phase 3 retest sample, the mean time interval between data collection phases was 11.0 years ($SD = 0.9$, range 8.7–13.1). For the Phase 2/Phase 3 retest, the mean time interval was 5.5 years ($SD = 1.6$, range 1.1–9.6).

Ethics approval was received from the institutional review boards (Queensland Institute of Medical Research and Washington University School of Medicine) appropriate to each study. In addition, informed verbal consent was received from participants in the Phase 1 and Phase 2 studies, and informed verbal and written consent was received from Phase 3 study participants. Abstainers (individuals who had never tried alcohol — 3%, 1%, and 4% of individuals at Phases 1, 2, and 3 respectively) were excluded from all analyses and are not included in sample numbers.

**Zygosity**

Zygosity was initially determined by self-report questionnaire with standard questions regarding physical

<table>
<thead>
<tr>
<th>Study</th>
<th>N/Age</th>
<th>Collection method</th>
<th>Interval</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaprio et al. (1992)</td>
<td>13,404 on two occasions (at time 1: older group 24–43 yrs, younger group 18–23 yrs)</td>
<td>Questionnaire</td>
<td>6 years</td>
<td>Older Group</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Quantity: $r = .64 – .70$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Frequency, beer: $r = .52 – .53$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Quantity: $r = .39 – .51$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Frequency, beer: $r = .30 – .32$</td>
</tr>
<tr>
<td>Pape &amp; Hammer (1996)</td>
<td>1084 on 3 occasions (19–22 years at time 1)</td>
<td>Questionnaire</td>
<td>2 years</td>
<td>Quantity × Frequency:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$r = .54$ (2 years)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$r = .37$ (4 years)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$r = .35$ (6 years)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Frequency:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$r = .72$ (2 years)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$r = .64$ (4 years)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$r = .58$ (6 years)</td>
</tr>
<tr>
<td>Wennberg et al. (2000)</td>
<td>80 males on 2 occasions (18 years at time 1, 34–37 years at time 2)</td>
<td>Interview</td>
<td>~18 years</td>
<td>Quantity: $r = .24$</td>
</tr>
<tr>
<td>Kerr et al. (2002)</td>
<td>(a) 7127 on 3 occasions (25–74 yrs at time 1)</td>
<td>Questionnaire</td>
<td>(a) 5 years</td>
<td>(a) Quantity × frequency:</td>
</tr>
<tr>
<td></td>
<td>(b) 3113 on 4 occasions (17–62 yrs at time 1)</td>
<td></td>
<td>−10 yrs</td>
<td>$r = .50$ (5 years)</td>
</tr>
<tr>
<td></td>
<td>(b) 5 years</td>
<td></td>
<td>10 yrs</td>
<td>$r = .26 – .36$ (−10 years)</td>
</tr>
<tr>
<td></td>
<td>15 yrs</td>
<td></td>
<td></td>
<td>(b) Quantity:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$r = .50$ (5 and 10 years)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$r = &lt; .50$ (15 years)</td>
</tr>
<tr>
<td>Whitfield et al. (2004)*</td>
<td>8184 in 1980 Study (18–88 years)</td>
<td>Questionnaire in 1980, 1989 and telephone interview (SSAGA)</td>
<td>4 years</td>
<td>Quantity × frequency:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>−4 years</td>
<td>$r = .66 – .69$ (−4 years)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>−9 years</td>
<td>$r = .57 – .60$ (−9 years)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>−13 yrs</td>
<td>$r = .54 – .56$ (−13 years)</td>
</tr>
</tbody>
</table>

Note: *includes a subsample of data examined in the current study plus data from earlier collection phases.

*SSAGA: Semi-Structured Assessment for the Genetics of Alcoholism (Bucholz et al., 1994)
AUSTRALIAN TWIN REGISTRY
a national volunteer registry supported by the National Health and Medical Research Council - begun in 1978 and recruited through the media, schools and a variety of other sources

OLDER COHORT
(born 1892-1964)
N = 5697 pairs

1980-1982
Canberra Study
N = 3808 complete pairs
576 single twins

1988-1989
Alcohol Cohort 1
N = 2995 complete pairs
337 single twins
2119 partners
2122 parents
2700 siblings
1660 children

(Phase 1)
1992-1993
SSAGA-OZ
N = 2685 complete pairs
519 single twins

1993-1994
SSAGA-II
N = 199 complete pairs
224 single twins

1993-1996
SSAGA-Blood
N = 1404 complete pairs
567 single twins

1993-1996
WOMEN S STUDY
N = 451 complete pairs
65 single twins

1994-1996
SSAGA-Spouse/Partner
N = 3848 partners

1979-1981
Alcohol Challenge
N = 206 pairs

1989-1992
Alcohol Cohort 2
N = 2271 complete pairs
517 single twins
326 partners
1456 parents
1017 siblings

(Phase 2)
1996-2000
Twin 89
N = 2765 complete pairs
735 single twins

(Phase 3)
2001-2005
NAG/IRPG
N = 1205 complete pairs
706 single twins
666 spouse index cases
+ parental/sibling recruitment

Recruitment Pools
Data Collection Phases
Mailed Questionnaire
Telephone Interview
In-Person Testing

1979-1991
Retest Samples
Phase 1/Phase 3 Retest:
N = 446 complete pairs
266 single twins
Phase 2/Phase 3 Retest:
N = 653 complete pairs
512 single twins

YOUNGER COHORT
(born 1964-1971)
N = 4268 pairs

Figure 1
Flowchart showing data sources and data collection phases leading to the present analyses (includes SSAGA-OZ spin-off studies). Twin data collected during Phases 1, 2, and 3 were examined for heritability and retest data were examined for reliability. Retest sample sizes reflect the number of twins with data for one or both of the traits examined. A brief description of each data collection (or study) phase is given in Table 2.
Bold boxes are used to identify the primary sources for the Phase 3 sample.
<table>
<thead>
<tr>
<th>Study</th>
<th>Type of data collection</th>
<th>Summary</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol Challenge (1979–1981)</td>
<td>In-person testing</td>
<td>Blood and urine samples were collected; a questionnaire on drinking and smoking habits was completed; alcohol was ingested (0.75 g ethanol/kg body weight diluted to 10% (v/v) at a steady rate over 20 minutes in a mixture of sugarless lemon and soda water); psychomotor tests were given; ratings of subjective intoxication taken; and 6 blood alcohol levels were taken during and between testing sessions</td>
<td>Two hundred and six twin pairs were recruited from Sydney and Canberra metropolitan areas through the Australian Twin Registry (ATR) for a clinical study of alcohol metabolism and susceptibility to intoxication (Martin et al., 1985; Whitfield &amp; Martin, 1983). To assess repeatability, 87 participants attended a second testing occasion. The sample was not selected on personal or family history of alcoholism, but pairs were excluded if either was a lifetime abstainer (Heath &amp; Martin, 1992; Madden, Heath, Starmer, Whitfield, &amp; Martin, 1995). (See also Birley et al., 2007; Birley et al., 2005; Boomsma, Martin, &amp; Molenaar, 1988; Dickson et al., 2000; Grant et al., 1999; Heath et al., 1999; Heath, Madden et al., 2002; Heath &amp; Martin, 1991a, 1991b; Heath, Whitfield et al., 2001; Madden, Heath, &amp; Martin, 1997; Martin &amp; Boomsma, 1989; Neale et al., 1999; Neale &amp; Martin, 1998; Slutske et al., 1995; Whitfield &amp; Martin, 1983, 1993, 1994, 1996; Whitfield, Nightingale, Bucholz et al., 1998; Whitfield, Starmer, &amp; Martin, 1999; Whitfield et al., 2001)</td>
</tr>
<tr>
<td>Canberra Study (1980–1982)</td>
<td>Mailed questionnaire</td>
<td>Health and lifestyle questionnaire containing questions about personality, health, and lifestyle, including self-reported drinking habits (a history of alcohol problems was not included)</td>
<td>All ATR twins born 1932 to 1963 (i.e., aged 18 or over) were targeted. The questionnaire was mailed to all 5967 twin pairs and completed replies were obtained from 3808 complete pairs (64% pairwise response rate) and 576 single twins (Heath &amp; Martin, 1994). The sample was over-represented for females and MZ pairs and under-represented for uneducated pairs, but representative of the Australian population for measures of personality, anxiety, depression, and weekly alcohol consumption (Heath &amp; Martin, 1988). (See also Bucholz et al., 1998; Dickson et al., 2006; Heath, Jardine, &amp; Martin, 1989; Heath, Meyer, Eaves, &amp; Martin, 1991a, 1991b; Heath, Meyer, &amp; Martin, 1991; Jardine &amp; Martin, 1984; Martin, 1987, 1988, 1991; Slutske et al., 1995)</td>
</tr>
<tr>
<td>Alcohol Cohort 1 (1988–1989)</td>
<td>Mailed questionnaire</td>
<td>Health and lifestyle questionnaire that included a history of alcohol problems (based on criteria proposed by Feighner (1972))</td>
<td>Complete pairs who had responded in the Canberra Study (i.e., 3808 pairs) were targeted. Twins who did not respond to the mailed questionnaire were offered an abbreviated telephone interview that included alcohol consumption questions, but alcohol problems were not assessed. In total, 2987 complete pairs (79% pairwise response rate) and 335 single twins (60% individual response rate) completed questionnaires or were interviewed (Heath &amp; Martin, 1994; Heath, Slutske et al., 1994). (See also Bucholz et al., 1998; Dickson et al., 2006; Heath et al., 1999; Heath, Todorov et al., 2002; Heath, Whitfield et al., 2001; Madden, Heath, &amp; Martin, 1997; Slutske et al., 2002; Slutske et al., 1996; Slutske et al., 1995; Whitfield et al., 2004; Whitfield &amp; Martin, 1996; Whitfield, Nightingale, Bucholz et al., 1998; Whitfield et al., 2000) For a spin-off study of relatives, participants were asked to provide names of parents, siblings, spouses, and children who would be prepared to answer a similar questionnaire and the names of 14,421 relatives were given. A modified version of the questionnaire was prepared for parents and another version for siblings, spouses, and children of twins. In total, 8601 relatives returned questionnaires (Lake, Eaves, Maes, Heath, &amp; Martin, 2000)</td>
</tr>
</tbody>
</table>
### Table 2 (continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Type of data collection</th>
<th>Summary Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol Cohort 2 (1989–1992)</td>
<td>Mailed questionnaire</td>
<td>Health and lifestyle questionnaire that included a history of alcohol problems (based on criteria proposed by Feighner (1972))</td>
</tr>
<tr>
<td>(Phase 1) SSAGA-OZ: Semi-Structured Assessment for the Genetics of Alcoholism (1992–1993)</td>
<td>Telephone interview</td>
<td>Adapted from the SSAGA (Bucholz et al., 1994), a diagnostic instrument developed for genetic research on alcoholism that provides diagnostic assessments of DSM-IV alcohol dependence and other major axis disorders</td>
</tr>
<tr>
<td>(Phase 2) Twin 89 (1996–2000)</td>
<td>Telephone interview</td>
<td>Adapted from the SSAGA (Bucholz et al., 1994), a diagnostic instrument developed for genetic research on alcoholism that provides diagnostic assessments of DSM-IV alcohol dependence and other major axis disorders</td>
</tr>
</tbody>
</table>

The target sample was twins born 1964 to 1971 who were registered with the ATR in 1989. A total of 2428 pairs. These twins were registered as children over the period 1989 to 1992 in response to mass appeals through the Australian school system. Two further individuals were invited to participate in the study based on referrals. For a total of 2951 individuals comprising 1475 male twins, 1476 female twins, and 1475 twin siblings. Of these, 2993 individuals were contactable by telephone, 2449 were reached, 1973 agreed to participate, 1972 provided complete data, and 1971 provided at least one measure of alcohol dependence (at least one twin). In total, 5947 twins were included in the study, 4268 pairs (2993 twins). The target sample was twins born 1964 to 1971 who were registered with the ATR in 1989, a total of 4268 pairs. These twins were recruited through mass appeals through the Australian school system. Two further individuals were invited to participate in the study based on referrals. For a total of 4528 individuals comprising 2264 male twins, 2264 female twins, and 2264 twin siblings. Of these, 4528 individuals were contactable by telephone, 3924 were reached, 3369 agreed to participate, 3368 provided complete data, and 3367 provided at least one measure of alcohol dependence (at least one twin). In total, 9106 twins were included in the study, 6534 pairs (3367 twins).
Table 2 (CONTINUED)
Overview of Studies and Data Collection Phases Shown in Figure 1

<table>
<thead>
<tr>
<th>Study</th>
<th>Type of data collection</th>
<th>Summary</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Phase 3) NAG/IRPG: Nicotine Addiction Genetics/Interactive Research Projects Grant (2001–2005)</td>
<td>Telephone interview</td>
<td>Adapted from the SSAGA (Bucholz et al., 1994), a diagnostic instrument developed for genetic research on alcoholism that provides diagnostic assessments of DSM-IIIR/DSM-IV alcohol dependence and other major axis I disorders</td>
<td>The NAG and IRPG are linked projects with identical interviews, protocols, and procedures. However, the projects differed in terms of their target samples, with the NAG having a smoking focus and the IRPG having an alcohol focus. The NAG was a collaborative project involving data collection in Australia and Finland, but only the Australian data are examined here. Families were targeted based on heavy smoking index cases identified in previous interview and questionnaire surveys. The IRPG projects comprised (a) a Candidate Gene Study, which was designed as an unmatched case-control study, (b) a Big Sibships Study, designed to study only families with 5 or more offspring sharing both biological parents; and (c) an Extreme Discordant and Concordant Sibships (EDAC) Study, designed to focus on families extremely discordant or concordant for heavy drinking and alcohol dependence risk. The families targeted were drawn from earlier surveys. The Candidate Gene Study specifically targeted twins with a history of alcohol dependence, their co-twins, informative siblings, and unrelated controls. Families with 5 or more siblings were targeted for the Big Sibships Study, regardless of earlier-reported alcohol behaviors. In targeted families, biological parents, spouses of twins, and the siblings of spouses were also approached. In addition, a small number of large sibship families were recruited from studies not related to alcohol use, or through the media. For the EDAC Study, the selection of informative sibships was based upon either a history of alcohol dependence or scores on a quantitative index of alcohol consumption (below the 30th percentile and above the 85th percentile). Targeted sibships were either discordant or high-concordant. Also included in the target sample were additional siblings and biological parents. The Australian NAG/IRPG sample comprised 9688 individuals (1205 complete pairs, 706 single twins, 3586 siblings of twins, 889 parents, 668 spouses of twins, and 1431 siblings of spouses). Ancestry was primarily British and European, as shown in Table 3. Funding was provided by NIDA/NCI for the NAG project and NIAAA for the IRPG projects.</td>
</tr>
</tbody>
</table>
similarity and degree to which others could tell co-twins apart. If co-twins gave inconsistent answers, they were followed-up by telephone, and if inconsistency or uncertainty was still apparent, they were asked to send in photographs at various ages, from which a zygosity assignment was made by project staff. Zygosity assignment based on self-report and responses to standard informative questions has been shown to be approximately 97% accurate (Reed et al., 2005). To confirm zygosity assignment, 347 pairs were genotyped at nine independent DNA microsatellite polymorphisms (STR loci: D5S818, D13S317, D3S1358, VWA, D8S1179, D7S820, D21S11, FGA, D18S51) plus the sex marker amelogenin using the Profiler multiplex marker set (AmpFLSTR Profiler Plus, Applied Biosystems, Foster City, CA). The probability of dizygotic twins being concordant for two alleles at the multiple polymorphic loci examined when using this kit, is reported to be less than $10^{-4}$ (Nyholt, 2006). Analysis of genotype data confirmed the accuracy of our zygosity assignment with a 100% confirmation rate for the subset of pairs examined.

**Measures**

**Quantity × Frequency**

A quantity £ frequency measure was derived from questions regarding the number of drinks consumed in a typical drinking day and drinking frequency.

**Typical day.** The telephone interview question at Phase 1 was ‘Think of the times when you’ve had alcohol during the past 12 months. How many drinks do you typically drink on these days when you had an alcoholic drink?’. At Phases 2 and 3, the question was ‘In the past 12 months, how many alcoholic drinks would you have on a typical day when you had any alcoholic drink?’ At Phase 1, the actual number of drinks was recorded. However at Phases 2 and 3, participants were given response choices. Subsequently, all responses were coded into the following categories: 1 = zero drinks, 2 = 1–2 drinks, 3 = 3–4 drinks, 4 = 5–6 drinks, 5 = 7–8 drinks, 6 = 9–11 drinks, 7 = 12–15 drinks, 8 = 16–18 drinks, 9 = 19–24 drinks, 10 = 25–30 drinks, 11 = 31 or more drinks.

**Frequency.** The telephone interview question regarding drinking frequency was ‘During the past 12 months, how often have you had alcoholic drinks?’. Response choices varied slightly between studies. They

### Table 4
Demographic Characteristics of the Samples

<table>
<thead>
<tr>
<th>Sample</th>
<th>$N$ Individuals (Complete Pairs)</th>
<th>Range</th>
<th>M</th>
<th>SD</th>
<th>% Female</th>
<th>% MZ Pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 1 (SSAGA)</td>
<td>5823 (2599)</td>
<td>28–90</td>
<td>43.8</td>
<td>12.2</td>
<td>65</td>
<td>48.8</td>
</tr>
<tr>
<td>Phase 2 (Twin 89)</td>
<td>6151 (2698)</td>
<td>23–39</td>
<td>29.9</td>
<td>2.5</td>
<td>55.3</td>
<td>42.5</td>
</tr>
<tr>
<td>Phase 3 (NAG/IRPG)</td>
<td>3060 (1177)</td>
<td>19–77</td>
<td>41.1</td>
<td>8.8</td>
<td>55.4</td>
<td>22.7*</td>
</tr>
<tr>
<td><strong>Retest Samples:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 1/Phase 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 1 (SSAGA)</td>
<td>1158 (446)</td>
<td>28–65</td>
<td>38.9</td>
<td>7.8</td>
<td>58.2</td>
<td>25.1</td>
</tr>
<tr>
<td>Phase 3 (NAG/IRPG)</td>
<td>—</td>
<td>38–77</td>
<td>49.9</td>
<td>7.8</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Phase 2/Phase 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 2 (Twin 89)</td>
<td>1818 (653)</td>
<td>24–36</td>
<td>30</td>
<td>2.4</td>
<td>54.2</td>
<td>21.4</td>
</tr>
<tr>
<td>Phase 3 (NAG/IRPG)</td>
<td>—</td>
<td>29–41</td>
<td>35.5</td>
<td>2.3</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Note: The total number of participants is 12,045, that is, the sum of the full samples (5823 + 6151 + 3060) less the retest samples (1158 + 1818) less 13 individuals participating in both Phases 1 and 2.

*This percentage is reduced, as the primary aim for this sample was linkage analyses and MZ pairs were only targeted if they had additional siblings.
were recoded into the following six categories that are common to both the telephone interview and earlier mailed questionnaire studies (6 = at least once daily, 5 = 3–6 days per week, 4 = 1–2 days per week, 3 = 1–3 days per month, 2 = less often, 1 = never).

The typical day categories were recoded to reflect the number of drinks per day. Similarly, the frequency categories were recoded to reflect a ‘times per week’ measure. Thus the quantity × frequency measure reflected the number of drinks per week.

**Dependence Score**

Assessment was based on lifetime prevalence. A dependence score was obtained from items based on Diagnostic and Statistical Manual of Mental Disorders, third edition, revised (DSM-III-R) criteria (American Psychiatric Association, 1987) and Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) criteria (American Psychiatric Association, 2000) for substance dependence, which were combined as follows:

1. substance often taken in larger amounts or over a longer period than the person intended
2. persistent desire or one or more unsuccessful efforts to cut down or control substance abuse
3. a great deal of time spent in activities necessary to get the substance or recover from its effects
4. frequent intoxication or withdrawal symptoms when expected to fulfill major role obligations at work, school, or home, or when substance use is physically hazardous
5. important social, occupational, or recreational activities given up or reduced because of substance use
6. continued substance use despite knowledge of having a persistent or recurrent social, psychological, or physical problem that is caused or exacerbated by the use of the substance
7. tolerance, as defined by either of the following:
   (a) a need for markedly increased amounts of the substance to achieve intoxication or desired effect
   (b) markedly diminished effect with continued use of the same amount of the substance
8. characteristic withdrawal symptoms
9. substance often taken to relieve or avoid withdrawal symptoms

All items, with the exception of item 7, were coded as 0 for no and 1 for yes. Item 7 was coded as a three-level ordinal measure with 0 for no, 1 for a marked (but less than 50%) increase in the amount required to achieve the desired effect, and 2 for at least a 50% increase in the amount required to achieve the desired effect. To obtain a single dependence score, items were summed. Note that individuals participating in Phases 2 and 3 were not re-asked dependence-related questions at Phase 3.

**Analyses**

Descriptive statistics were performed using SPSS 13.0 for Windows. Quantity × frequency and dependence scores were treated as continuous variables. In addition, variables were treated as categorical data in confirmatory reliability analyses as data (particularly dependence score) were not normally distributed. Both quantity × frequency and dependence score were positively skewed. Quantity × frequency was log transformed \(\log_{10}(x + 1)\) where \(x\) is the number of drinks per week as has been done previously (Jardine & Martin, 1984). However, transformation did not improve the dependence score distribution and it was consequently left untransformed. Distributions for the full and retest samples are shown in Figure 2.

Raw data on quantity × frequency and dependence score were used to regress out the effects of important covariates, including tests for sex, age at interview, quadratic effects of age (age\(^2\)), sex*age and sex*age\(^2\). PROC REGRESS in SAS was used to regress out the effects of covariates that were significant at \(p < .05\) in Phases 1, 2, and 3 separately and residuals stemming from this regression were used for both the stability and heritability calculations.

All modeling was performed using raw data in the statistical software package Mx (Neale, Boker, Xie, & Maes, 2003) using a full information maximum likelihood (FIML) estimator. The FIML method used in Mx is especially useful in handling missing data. Estimates are unbiased when data are missing completely at random and/or missing at random is predicted by other variables used in the analysis (Little & Rubin, 2000). To assess model fit, the fit of constrained models was compared to the full model by examining the difference in the −2 log likelihood, which is distributed as a chi-square for given degrees of freedom.

**Stability**

The following analyses were conducted for quantity × frequency and dependence score for the Phase 1/Phase 3 retest sample and for quantity × frequency for the Phase 2/Phase 3 retest sample. Residuals, as computed from the full samples, were examined. Submodels were compared to the full model, with the submodels for continuous data being:

- whether mean scores could be equated across phases for each retest sample
- whether variances could be equated across phases for each retest sample
- test–retest correlations with 95% confidence intervals

and for categorical data:

- whether thresholds could be equated across phases for each retest sample (note that each variable was divided into three categories of approximately equal size based on the first interview occasion responses)
Figure 2
Distributions for quantity × frequency (transformed) and dependence score for the full samples (a) and (b), and the retest samples (c), (d), and (e).
test–retest correlations with 95% confidence intervals.

**Heritability**

The extent to which additive genetic (A), shared environmental (C) and nonshared environmental (E) factors influenced population differences in quantity × frequency and the dependence score was examined. Univariate sex-equated models were fit to regression residuals for Phase 1, Phase 2, and Phase 3 data separately. In addition, models comparing quantitative differences in estimates of heritable and environmental influences on the full Phase 1 or Phase 2 samples and the subset of twins targeted for Phase 3, the latter of which was oversampled for smoking/drinking, were examined. The following submodels were tested against the full model (note that dependence score was not available at Phase 3 if collected at Phase 2, and thus a Phase 2/Phase 3 comparison could not be made):

- whether mean scores could be equated across the full Phase 1/Phase 2 sample and their respective Phase 3 components
- whether total variance could be equated across the full Phase 1/Phase 2 sample and their respective Phase 3 components.
- if the total variance could be equated, whether or not additive genetic, shared environmental and nonshared environmental influences could be constrained to be equal across Phase 1/Phase 2 and their respective Phase 3 components was examined. However, if there was statistical evidence in favor of differences in total variance, we used a scalar twin model where the total variance in Phase 3 was modeled as a scalar multiple of the variance in Phase 1/Phase 2. Note that as heritability reflects a proportion of total variance (A/A + C + E), its interpretation is determined by the magnitude of the denominator. Therefore, for instance, if variance was estimated as X in Phase 2 and differently, as Y, in Phase 3, then A/X could not be statistically equated to A/Y as a test for equality of heritable factors. Hence, we used a scalar product (or Y = X^n, where n ranges from –∞ to +∞; if n = 1 then variances are equal) for instances in which variances could not be constrained.
- tests for the statistical significance of A and C were also conducted.

Further, bivariate Cholesky triangular decomposition models of variance (Neale, Boker, Xie, & Maes, 2003) were fit to the full Phase 1/Phase 3 and full Phase 2/Phase 3 datasets to obtain genetic and environmental correlations between phases. Although only a subset of retest data was available, including the full datasets results in more stable estimates of sample characteristics.

**Results**

**Sample Characteristics**

Raw data means are shown in Table 5. Overall, the means for quantity × frequency indicate that weekly consumption is similar between the older (Phase 1) and younger (Phase 2) cohorts. In addition, there is a trend for increased consumption from the earlier collection Phases (1 and 2) to the later collection Phase (3). Means for dependence score show that, at each collection stage, males satisfied a larger number of criteria than females. In addition, the younger cohort (Phase 2) satisfied a larger number of criteria than the older cohort (Phase 1) or the combined cohort (Phase 3). Significance of covariates is shown in Table 6. Their effects were regressed from the variables of interest and residuals were examined for all remaining analyses.

| Table 5 |
| Raw Data Means for Alcohol Variables |

<table>
<thead>
<tr>
<th>Sample</th>
<th>Quantity × frequency</th>
<th>Dependence score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Males</td>
<td>Females</td>
</tr>
<tr>
<td><strong>Full Sample</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 1 (SSAGA)</td>
<td>7.7</td>
<td>3.6</td>
</tr>
<tr>
<td>Phase 2 (Twin 89)</td>
<td>7.9</td>
<td>3.7</td>
</tr>
<tr>
<td>Phase 3 (NAG/IRPG)</td>
<td>10.6</td>
<td>5.4</td>
</tr>
<tr>
<td><strong>Retest Sample</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 1/Phase 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 1 (SSAGA)</td>
<td>9.9</td>
<td>4.2</td>
</tr>
<tr>
<td>Phase 3 (NAG/IRPG)</td>
<td>11.0</td>
<td>5.5</td>
</tr>
<tr>
<td>Phase 2/Phase 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 2 (Twin 89)</td>
<td>9.2</td>
<td>4.4</td>
</tr>
<tr>
<td>Phase 3 (NAG/IRPG)</td>
<td>10.4</td>
<td>5.3</td>
</tr>
</tbody>
</table>

Note: *Dependence-related questions were not asked at Phase 3 if already asked at Phase 2.*
drinking at a low level and fewer at a high level. For dependence score, at Phase 3 for the Phase 1/Phase 3 retest sample, 48% of respondents were in the lower tail and 31% in the upper tail. This indicates a decrease over time in the number of individuals reporting either an intermediate or a high number of dependence symptoms and a corresponding increase in the number of individuals reporting a low number of symptoms. Indeed, 304 individuals (26.3% of the retest sample) reported fewer symptoms at Phase 3 compared to Phase 1. Of the remaining participants, 141 had at least one symptom reported at Phase 1 that was not reported again at Phase 3. Therefore, 38.5% of the retest sample was inconsistent in the reporting of at least one ‘lifetime’ symptom of alcohol dependence.

Prior to the heritability analyses: Means for quantity × frequency could be constrained to be equal across the full Phase 1 and corresponding Phase 3 retest samples, as well as between full Phase 2 and Phase 3 retest samples (Δχ² ranged 0.0–0.8, p > .37). Variances could be equated across Phase 2 and Phase 3 (Δχ², 0.2, p = .65) but not across Phase 1 and Phase 3 (Δχ², 11.8, p < .001). For dependence score, there were statistically significant differences in both the means (Δχ², 25.5, p < .001) and the variances (Δχ², 287.3, p < .001) when comparing Phase 1 with Phase 3. This contrasts with the results reported as preliminary to the stability analyses, where means could be set equal. It is surmised that there is a mean difference between dependence score at Phases 1 and 3, but that this difference is not statistically significant in the smaller retest samples (i.e., when the Phase 1 sample is restricted to individuals retested).

### Analyses of Means, Variances, and Thresholds

When examined as continuous data, quantity × frequency means (of residuals) could be equated across phases for both the Phase 1/Phase 3 and Phase 2/Phase 3 retest samples (Δχ² ranged 0.0–2.0, p > .1). Variances could be equated for the Phase 2/Phase 3 retest sample (Δχ², 0.8, p > .3), but not for the Phase 1/Phase 3 retest sample (Δχ², 13.7, p < .001). In addition, means for dependence score could be set equal across phase for the Phase 1/Phase 3 retest sample (Δχ², 0.3, p > .5). However, variances could not be equated for dependence score for the Phase 1/Phase 3 retest sample (Δχ², 29.7, p < .001).

When examined as categorical data, thresholds could not be set equal across interview occasions for any variable (Δχ² ranged 7.8–63.6, p < .03). For retest samples on first interview occasions, thresholds were set to divide data into three equally-sized categories (low, intermediate, high). However, for quantity × frequency, at Phase 3 for the Phase 1/Phase 3 retest sample the proportion of respondents in the upper and lower tails increased to 37% in each case. Thus, at Phase 3 compared to Phase 1, there were fewer individuals drinking at an intermediate level and a larger number of individuals drinking at both low and high levels. At Phase 3 for the Phase 2/Phase 3 retest sample, 36% of respondents were in the lower tail and 31% in the upper tail. Therefore, on the second interview occasion there was the same proportion of intermediate level drinkers, but more individuals were drinking at a low level and fewer at a high level.

### Table 6

<table>
<thead>
<tr>
<th>Significant covariates</th>
<th>β</th>
<th>SE</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantity × frequency</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 1: Sex</td>
<td>0.2551</td>
<td>0.0115</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Age</td>
<td>0.008</td>
<td>0.0032</td>
<td>.131</td>
</tr>
<tr>
<td>Age²</td>
<td>−0.0001</td>
<td>0</td>
<td>.0017</td>
</tr>
<tr>
<td>Phase 2: Sex</td>
<td>−1.0243</td>
<td>0.3896</td>
<td>.0086</td>
</tr>
<tr>
<td>Sex×Age</td>
<td>−0.8521</td>
<td>0.2613</td>
<td>.011</td>
</tr>
<tr>
<td>Sex×Age²</td>
<td>0.0139</td>
<td>0.0044</td>
<td>.014</td>
</tr>
<tr>
<td>Phase 3: Sex</td>
<td>0.2522</td>
<td>0.0164</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Age</td>
<td>0.0184</td>
<td>0.0079</td>
<td>.208</td>
</tr>
<tr>
<td>Age²</td>
<td>−0.0002</td>
<td>0.0001</td>
<td>.0155</td>
</tr>
<tr>
<td>Dependence score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 1: Sex</td>
<td>0.908</td>
<td>0.1916</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Age</td>
<td>−0.0006</td>
<td>0.0001</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Sex×Age</td>
<td>−0.2516</td>
<td>0.0718</td>
<td>.0005</td>
</tr>
<tr>
<td>Sex×Age²</td>
<td>0.0042</td>
<td>0.0008</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Phase 2: Sex</td>
<td>1.1982</td>
<td>0.0574</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Phase 3: Sex</td>
<td>−7.4612</td>
<td>1.4288</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Age</td>
<td>1.0499</td>
<td>0.1061</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Age²</td>
<td>−0.0095</td>
<td>0.0011</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Sex×Age</td>
<td>−3.1839</td>
<td>0.6322</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>Sex×Age²</td>
<td>0.0276</td>
<td>0.0067</td>
<td>&lt; .0001</td>
</tr>
</tbody>
</table>

### Table 7

<table>
<thead>
<tr>
<th>Sample</th>
<th>Continuous data</th>
<th>Categorical data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase 1/Phase 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantity × frequency</td>
<td>.61 (.57, .65)</td>
<td>.64 (.59, .69)</td>
</tr>
<tr>
<td>Dependence score</td>
<td>.61 (.57, .64)</td>
<td>.58 (.52, .63)</td>
</tr>
<tr>
<td>Phase 2/phase 3*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantity × frequency</td>
<td>.55 (.51, .58)</td>
<td>.59 (.55, .59)</td>
</tr>
</tbody>
</table>

Note: *Dependence-related questions were not asked at Phase 3 if previously asked at Phase 2.
47% of the total variance (Table 8) with the magnitude of heritable influences being somewhat higher in the older versus younger cohorts of twins. In Phase 1 and its component Phase 3, population variance could be attributed to additive genetic (51%) and nonshared environmental (49%) influences. Similarly, in Phase 2 and its component Phase 3, an additive genetic factor accounted for 43% of the total variance, with the remainder due to nonshared environmental influences.

Dependence score. Additive genetic (42–46%) and nonshared environmental factors influenced population variation in dependence score across the three full samples. However, in contrast to quantity × frequency, a statistically significant difference between the variances for Phase 1 and its component Phase 3 was found. Consequently, scalar models were pursued in this comparison. After accounting for these differences in sample variance, heritable factors explained about an equal proportion of this variance (51%) in each sample.

In each of the aforementioned models, additive genetic and nonshared environmental influences were sufficient to explain the variance (Table 8). The influence of shared environmental factors could be dropped in each case without loss of fit (Δχ², ranged 0.0–2.4, p > .1). In the fully saturated models, shared environment accounted for 0% of the variance at all phases (confidence intervals were 0–.02, .0–.06, and 0–.10 for Phases 1, 2, and 3 respectively).

Table 8
Means and Total Variances for the Residuals plus Additive Genetic (A) and Nonshared Environmental (E) Influences with 95% Confidence Intervals for Quantity × Frequency and Dependence Score

<table>
<thead>
<tr>
<th>Sample</th>
<th>Mean (residuals)</th>
<th>Total variance</th>
<th>A (95% CI)</th>
<th>E (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 1 (SSAGA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantity × frequency</td>
<td>0</td>
<td>0.18</td>
<td>0.47 (0.43, 0.51)</td>
<td>0.53 (0.49, 0.57)</td>
</tr>
<tr>
<td>Dependence Score</td>
<td>0.003</td>
<td>2.64</td>
<td>0.46 (0.42, 0.51)</td>
<td>0.54 (0.49, 0.58)</td>
</tr>
<tr>
<td>Phase 2 (Twin 89)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantity × frequency</td>
<td>0.002</td>
<td>0.18</td>
<td>0.31 (0.23, 0.36)</td>
<td>0.69 (0.64, 0.77)</td>
</tr>
<tr>
<td>Dependence score</td>
<td>0.017</td>
<td>5.32</td>
<td>0.45 (0.37, 0.49)</td>
<td>0.55 (0.51, 0.63)</td>
</tr>
<tr>
<td>Phase 3 (NAG/IRPG)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantity × frequency</td>
<td>0.001</td>
<td>0.20</td>
<td>0.40 (0.24, 0.48)</td>
<td>0.60 (0.52, 0.76)</td>
</tr>
<tr>
<td>Dependence score</td>
<td>0.006</td>
<td>2.03</td>
<td>0.42 (0.26, 0.50)</td>
<td>0.58 (0.50, 0.74)</td>
</tr>
<tr>
<td>Combined full/retest</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Phase 1/Phase 3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 3 retest</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantity × frequency</td>
<td>0.002</td>
<td>0.21</td>
<td>0.51 (0.47, 0.54)</td>
<td>0.49 (0.46, 0.53)</td>
</tr>
<tr>
<td>Dependence Score</td>
<td>0.377</td>
<td>5.21</td>
<td>0.51 (0.48, 0.55)</td>
<td>0.49 (0.45, 0.52)</td>
</tr>
<tr>
<td>(Phase 2/Phase 3a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 3 Retest</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantity × frequency</td>
<td>0.001</td>
<td>0.19</td>
<td>0.43 (0.39, 0.53)</td>
<td>0.57 (0.54, 0.61)</td>
</tr>
</tbody>
</table>
| Note: Dependence-related questions were not asked at Phase 3 if previously asked at Phase 2, therefore there is no retest sample for dependence score.

Dependence score. Additive genetic (42–46%) and nonshared environmental factors influenced population variation in dependence score across the three full samples. However, in contrast to quantity × frequency, a statistically significant difference between the variances for Phase 1 and its component Phase 3 was found. Consequently, scalar models were pursued in this comparison. After accounting for these differences in sample variance, heritable factors explained about an equal proportion of this variance (51%) in each sample.

In each of the aforementioned models, additive genetic and nonshared environmental influences were sufficient to explain the variance (Table 8). The influence of shared environmental factors could be dropped in each case without loss of fit (Δχ², ranged 0.0–2.4, p > .1). In the fully saturated models, shared environment accounted for 0% of the variance at all phases (confidence intervals were 0–.02, .0–.06, and 0–.10 for Phases 1, 2, and 3 respectively).

For quantity × frequency, genetic influence was found to be highly stable across the 5.5 and 11.0 year intervals examined (r_g = .96 for both intervals). In addition, the stability of genetic influence over 11.0 years was moderately strong for dependence score (r_g = .62). However, nonshared environmental influences were less stable (r_e = .33 for quantity × frequency over 5.5 years and .34 over 11.0 years, r_e = .40 for dependence score over 11.0 years).

Discussion
The aims of the present article were first, to examine the long-term stability of alcohol dependence symptoms and a quantity × frequency measure of consumption, and second, to examine the heritability of these measures for different collection phases. The results, which are outlined more fully in following paragraphs, indicate that (a) both measures have moderate long-term stability, (b) our dependence measure may not reflect a true ‘lifetime’ measure, (c) both measures are moderately heritable, (d) quantity × frequency appears less heritable in a younger versus older cohort, (e) dependence appears to be less heritable when examined as symptom count than as a yes/no dichotomous measure, (f) selection for drinkers and smokers did not significantly affect the heritability of the measures, and (g) genetic influence
was highly stable over the intervals examined, particularly for consumption.

In the current paper, the stability of alcohol dependence symptoms and a quantity \times frequency measure were examined in an older cohort (ages ranging 28–65 years at baseline) over a mean interval of 11.0 years, and for quantity \times frequency only, in a younger cohort (ages ranging 24–36 years at baseline) over a mean interval of 5.3 years. For both measures and intervals, moderate stability is found with correlations ranging from .55 to .64. Typically, stability decreases slowly over time (e.g., Pape & Hammer, 1996; Whitfield et al., 2004). However, in the present case stability is higher, although not significantly, over the longer compared to the shorter interval (.61-.64 vs. .55-.59 for quantity \times frequency). This may be due to the younger age of participants examined for the shorter interval (41% vs. 2% < 30 years). It has previously been found that the stability of alcohol consumption is lower in younger samples (Kerr, Fillmore, & Bostrom, 2002) and that level of consumption and drinking problems in adolescents and young adults is not highly predictive of later drinking habits and problems (Fillmore et al., 1991).

As expected, the long-term stability estimates are considerably lower than reports of test–retest reliability (.58–.62 vs. .84–.90, (Bucholz et al., 1994), for alcohol dependence symptoms, and .55–.64 vs. .73–.91, (Friesema et al., 2004; Webb, Redman, Gibberd, & Sansom-Fisher, 1991) for quantity \times frequency) indicating changes in drinking habits over time. The stability of dependence symptoms over 11 years is similar to that reported by Slutske, True et al., (1998) for an interval of only 15 months (.58–.61 vs. .61). Further, it is consistent with the general finding of degradation of stability over time, as stability was found to be .77 for dependence symptomatology in a subset of females from the same sample retested after an interval of approximately 2 years (Heath, Bucholz et al., 1997).

Interestingly, close examination of the dependence symptomatology retest data shows that over a third of participants reported at least one symptom at the first interview, which was not subsequently reported at the second interview. However, the interview questions are based on a lifetime diagnosis, that is, they asked if a symptom had ever been experienced rather than if a symptom is currently being experienced or was recently experienced. One explanation for this is that many participants may have responded positively only for symptoms that had been recently experienced, rather than ever experienced. Consequently, the measure of dependence symptomatology examined does not appear to be a true measure of lifetime dependence. Such a finding is not unique to the present study. In a sample of Vietnam era veterans, Slutske, True et al. (1998) found that 13% of their sample had a lifetime alcohol dependence diagnosis at time one, but not 15 months later at time two.

Stability estimates for the longer interval for quantity \times frequency are similar to those based on the Phase 1 telephone interview and earlier questionnaire surveys from the same sample and reported by Whitfield et al. (2004); .61–.64 over ~11 years vs .57–.60 over ~9 years and .54–.56 over ~13 years). These estimates are higher than those reported by Kerr et al. (2002) for quantity \times frequency measures collected by questionnaire over a 10 year interval (.26–.36). However, current estimates for the shorter interval are consistent with estimates reported for a 5-year interval by Kerr et al. (2002); .55-.59 vs .50). Pape and Hammer (1996) reported lower estimates (.35–.54 for 2–6 year intervals), but they examined a younger sample (aged 19–22 at baseline).

As mentioned previously, in addition to having moderate long-term stability, alcohol dependence symptomatology and quantity \times frequency measures both appear to have high test–retest reliability in the short term(Bucholz et al., 1994; Gerstel, Harford, & Pautler, 1980) (Friesema et al., 2004; Webb, Redman, Gibberd, & Sansom-Fisher, 1991). This is an important factor for genetic studies because measurement reliability has a direct impact on the heritability values estimated. The greater the error variance in the measures, the less the measures reflect the true level of individual difference, and the smaller the upper bound of the heritability estimate. Although reliability was not directly investigated in this sample, reports of high reliability for the measures examined suggest a high upper bound for the heritability estimate.

Moderate heritabilities are found for both measures (.42–.51 for dependence symptomatology, .31–.51 for quantity \times frequency). All remaining variance is due to environmental influences unique to the individual. The heritabilities found for dependence symptoms are at the lower end of those generally reported. For example, estimates in the range of .50 to .60 were reported for dependence in a review by Dick and Bierut (2006). Similarly, heritabilities ranging .45 to .65 were reported in a review by Heath, Slutske et al. (1994) for alcoholism in community samples. However, it is notable that a heritability of .64 was found by Heath, Bucholz et al. (1997) in an earlier study of the Phase 1 sample, compared to .46 found in the current study. The major difference between the studies is in how dependence was quantified, which subsequently influenced how it was analyzed. In the Heath, Bucholz et al. (1997) study, dependence was quantified as a dichotomous variable (based on the DSM-IIIR criteria, whereby individuals with three or more symptoms were diagnosed as being dependent) and analyzed using a threshold model. In the present study, symptoms based on the DSM-IIIR and DSM-IV criteria are summed and analyzed as a continuous
variable. These two approaches differ in power, with analysis of continuous data having considerably greater power than analysis of threshold data (Neale, Eaves, & Kendler, 1994). For the Phase 2 sample, similar measurement and methodological differences between the current and an earlier study did not differentiate heritability estimates for dependence symptomatology, with estimates of .45 found in the current study and .47 in the earlier study (Knopik et al., 2004).

Heritability estimates for the quantity × frequency measure fall within the general range of heritabilities reported for alcohol consumption patterns (i.e., .30-.60, Heath & Martin, 1994). As expected, it is similar to that reported previously for data collected from the same sample at Phase 1 and in earlier questionnaire surveys (.40-.54). Of note is the finding of significantly lower heritability in the younger compared to the older cohort (.31 vs. .47). Other studies have hinted at increasing heritability with age for alcohol consumption in adolescents and young adults, but a recent meta-analysis by Bergen et al. (2007) did not find differences to be significant. The difference found in the current analyses may reflect a decreasing environmental influence from young adulthood to middle age. Heritabilities for alcohol consumption appear to be consistent through middle to older age. For example, heritabilities ranging .44 to .47 over a 14- to 18-year period have been reported for older adult males (Reed et al., 1994).

The source of genetic influence on the measures of interest appears to be highly stable over the two intervals examined. For quantity × frequency, almost all of the genes influential at baseline were still influential after intervals of 5.5 or 11.0 years. In addition, most of the genes influencing the dependence symptoms score at baseline were still influential after an 11-year interval. In contrast, the influence of nonshared environmental factors was considerably less stable over time. However, this may be partly due to measurement error, which is included in the estimate of nonshared environmental influence, and which would have a negative impact on stability.

A number of limitations are evident for this study. First, the alcohol phenotypes are based on self-report, which may be associated with problems related to recall, willingness to impart information, or correctly interpreting the question. Nevertheless, self-report measures of alcohol consumption are reported to demonstrate reasonable levels of reliability and validity and have the advantage of being relatively inexpensive, noninvasive, and acceptable to respondents (Del Boca & Darkes, 2003). Second, the data were collected for reasons other than to examine stability, and consequently, in terms of a stability study, the data for interval length are confounded by cohort age. Thus, a clear comparison of stability for interval length cannot be made. Third, the data, particularly for alcohol dependence symptoms, are not normally distributed, but are analyzed as continuous data. This may bias results. However, to address this limitation threshold models were also used to examine data stability and no significant differences were found. Lastly, genetic modeling shows that the variance for both measures is strongly influenced by nonshared environmental factors, but the nature of these influences is not assessed in this study.

The results of the present study will inform decisions made regarding the structure of a combined sample for future linkage and association studies. These decisions include choice of data to analyze when data were collected on two occasions. Due to significantly lower heritability found for alcohol consumption in the younger cohort compared to the older cohort, choosing data collected at Phase 3 rather than at Phases 1 or 2 may maximize power. Further, the heritability estimates derived in the current analyses will be informative regarding the specification of heritability in future quantitative linkage analyses.

More generally, the results suggest that many individuals may not accurately self-report lifetime symptoms of alcohol dependence. Rather, their responses may be disproportionately influenced by current or recent symptoms. Second, oversampling for smokers and drinkers did not significantly influence heritability for the alcohol traits examined. However, heritability did vary significantly for cohort, suggesting that the influence of an individual’s unique life experiences on patterns of alcohol consumption may decrease as they move from young adulthood to middle age. Last, the influence of genetic factors was highly stable over the intervals examined.

Acknowledgments

We thank all participating families, the project coordinators and interviewers under the supervision of Dixie Statham, the data managers under the supervision of John Pearson and David Smyth, and laboratory personnel under the supervision of Grant Montgomery and Megan Campbell. National Institute of Health grants have supported this work, including grants to Drs Andrew Heath (AA07728, AA10248, AA11998, AA13321), Pamela Madden (DA12854), Nick Martin (AA13326), Michele Pergadia (DA019951), Richard Todd (AA13320), and John Whitfield (AA014041). Dr Arpana Agrawal receives support from the Alcoholic Beverages Medical Research Foundation and Dr Katherine Morley is supported by an NHMRC Public health Fellowship (520452).

References

Agrawal, A., Heath, A. C., Grant, J. D., Pergadia, M. L., Statham, D. J., Bucholz, K. K., Martin, N. G., &


