

PERSPECTIVES

Methodological Challenges and Limitations of Research on Alcohol Consumption and Effect on Common Clinical Conditions: Evidence from Six Systematic Reviews

Barbara J. Turner, MD, MEd¹ and A. Thomas McLellan, PhD^{2,3}

¹Division of General Internal Medicine, University of Pennsylvania, Philadelphia, PA, USA; ²Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA, USA; ³Treatment Research Institute, Philadelphia, PA, USA.

BACKGROUND: Despite the high prevalence of alcohol consumption in the US, 'mainstream' physicians generally consider it to be peripheral to most patient care. This may be due in part to a dearth of rigorous research on alcohol's effect on common diseases.

METHODS: To evaluate this issue, we examined six systematic reviews, four of which were conducted as part of a research initiative supported by the Robert Wood Johnson Foundation, the Program of Research to Integrate Substance Use Information into Mainstream Healthcare (PRISM). PRISM aimed to assimilate and improve the evidence on the medical impact of alcohol (and other drugs of abuse) on common chronic conditions.

RESULTS: From these reviews, we summarize the methodological limitations of research on alcohol's impact on development and/or clinical course of depression, hypertension, diabetes, bone disease, dementia, and sexually transmitted diseases. The studies included in these reviews were largely fair to good quality, and few were in primary care settings. Syntheses were hampered by the myriad of definitions of alcohol consumption from any/none to seven levels and a plethora of types of alcohol use disorders.

CONCLUSION: We recommend more high-quality observational and experimental studies in primary care settings as well as a more standard approach to quantifying alcohol use and to defining alcohol use disorders.

KEY WORDS: alcohol; standardized measurement; medical impacts; research limitations; alcohol drinking/adverse effects.

J Gen Intern Med 24(10):1156-60

DOI: 10.1007/s11606-009-1072-z

© Society of General Internal Medicine 2009

A 2005 Institute of Medicine (IOM) report warned that physicians' failure to screen for and address the effects of alcohol on health and disease compromised the quality of care

for Americans¹. The IOM report is responsive to warning signs from national survey data. For example, according to 2006 National Center for Health Statistics data, 61% of Americans drink alcohol, and 33% of drinkers report bingeing (i.e., five or more drinks/day)². Despite the high prevalence of alcohol consumption, often at levels deemed to be unhealthy³, most physicians pay little attention to their patients' alcohol consumption or to its effects on common diseases⁴⁻⁶. When physicians do ask about alcohol use, they record it in the social history section of the chart so alcohol use information is usually not accessible when diagnosing and treating diseases.

The limited relevance and quality of research on the effects of alcohol use on common clinical conditions seen in primary care practice may contribute to this lack of attention. Thus, with support from the Robert Wood Johnson Foundation, the Program of Research to Integrate Substance Use Information into Mainstream Healthcare (PRISM) was launched in 2004 to assimilate and improve the evidence on the medical impact of alcohol (and other drugs of abuse) on common chronic conditions. To these ends, we commissioned systematic reviews about the risks and benefits of alcohol use on a variety of common medical conditions that internists manage on a daily basis. In addition, other critical reviews have recently been published on the impact of alcohol use disorders on common clinical conditions. In this commentary, we identify themes from the critiques of the evidence from six reviews and offer suggestions to improve the validity and relevance of the research on the impact of alcohol consumption on common clinical conditions.

For this methodological critique, we included six reviews: four reviews commissioned by PRISM⁷⁻¹⁰ and two identified in a search of the literature. We searched PubMed, MEDLINE, and Cochrane Libraries for all English language systematic reviews published from 1/2004 through 5/2009 using the search terms "alcohol drinking/adverse effects" AND "systematic reviews." The search found 37 unique studies, of which 2¹¹⁻¹² met the following criteria shared by the PRISM reviews: examined the impact of alcohol on the development, clinical course, or management of common diseases; performed explicit, reproducible searches; critiqued methods in identified studies; and written for a general audience (i.e., published in non-addiction specialty journals).

As shown in Table 1, most of these six systematic reviews initially identified large numbers of potentially relevant studies, but only small fractions were deemed eligible for inclusion in the final review. Most identified studies employed a prospective cohort design, but experimental trials were also

Supported by the Robert Wood Johnson Foundation and the Betty Ford Foundation

Received December 13, 2008

Revised June 12, 2009

Accepted June 30, 2009

Published online August 12, 2009

Table 1. Evidence from Six Systematic Reviews of Alcohol Consumption and Effect on Common Medical Conditions

Domain	Depression	Hypertension	Diabetes mellitus	Bone disease	Dementia, cognitive decline in elderly	Sexually transmitted diseases (STDs)
Research questions on the effect or association of alcohol consumption	Prevalence of alcohol problems, Clinical course of depression, Response to treatment	Blood pressure	Incidence of diabetes Diabetes management, Complications	Fracture risk Bone density, Bone loss over time	Incident cognitive Decline/dementia	STD acquisition
Databases searched	MEDLINE, PsychINFO, Cochrane libraries	MEDLINE	MEDLINE	MEDLINE, PsychINFO, current contents, Cochrane libraries	MEDLINE, Embase, Psychinfo	MEDLINE
Publications in initial search (N)	1,579	834	974	914	94	1,777
Final eligible studies (N)	35	9	32	35	26	42
Study types	Prospective cohort Cross-sectional surveys Community household survey	Experimental trials	Prospective cohort Experimental trials	Prospective cohort Case control Experimental trials	Prospective cohort Retrospective case-control Nested in a cohort	Prospective cohort Cross-sectional surveys
Meta-analysis	No	Yes	No	Yes	Yes, but significant heterogeneity	No
Relevance of study to primary care	7 all or partially in primary care Observational cohorts	Only experimental settings	All observational cohorts from general populations	All observational cohorts from general populations	Some observational population-based cohorts	One population-based cohort None in primary care
Measures of study quality assessment	Jadad quality score ²⁶ Quality index score ¹³	None	USPSTF* internal validity criteria	USPSTF* internal validity criteria Rate and frequency of alcohol assessment Evaluation of confounders	Alberta Heritage Foundation for Medical Research ²⁷	No uniform quality measure Critique of alcohol use measures
Summary of quality assessment	Randomized controlled trials—2 excellent and 1 poor 7 cohort and 3 case series studies—poor or moderate	No grading of study quality except method of measuring blood pressure	Most fair except 3 good No standard diagnosis of diabetes, Limited adjustment for confounding	All fair Limited adjustment for confounding	Quality >15 of a possible 22 score Biased subject selection, Limited baseline data Lack of detail in results	Wide variety of measures of alcohol and STDs Cross-sectional design

*USPSTF, US Preventive Services Task Force

conducted for several of the review topics. A meta-analysis was not conducted by two reviews because of the heterogeneity of research designs, the myriad ways of defining alcohol consumption, and the diversity of the outcome measures.

In regard to the relevance of this research to primary care practice, the depression review included seven studies from primary care settings, but four of these were combined with psychiatric care settings. The other systematic reviews did not specifically address whether the studies were conducted in primary care patients. However, in four reviews, some evidence came from observational cohorts drawn from general populations that are likely to be relevant to primary care practice.

Four reviews employed previously published quality assessment tools to evaluate the rigor of identified studies (Table 1).

The depression study reported that several randomized trials were of excellent quality⁷, but, based on the Quality Score Index score¹³, the observational studies on alcohol and depression were of lower quality than studies included in other systematic reviews¹⁴. The hypertension review focused primarily on critiquing the method of measuring blood pressure⁸. The diabetes and bone reviews judged the eligible cohort studies to be primarily of “fair quality” because most were deficient in adjustment for confounders. For example, in the diabetes review, the authors noted failure to adjust for waist-to-hip ratio, family history of diabetes, and race, while, for the bone disease review, the authors found that most studies did not adjust for obvious correlates of bone disease such as estrogen use⁹⁻¹⁰. The dementia review did not report serious flaws in

selected studies, but they employed a measure that did not provide an overall research quality grade¹¹. The STD review primarily critiqued the types of alcohol measures and STD outcomes in identified studies¹².

The overriding methodological concern noted by the systematic reviews was the heterogeneity of the alcohol use measures employed by the research studies (Table 2). The review of alcohol's effect on blood pressure was least affected by this variability because all of the included experimental trials reported on the quantity of alcohol that was administered. But these trials frequently did not describe study subjects' usual alcohol consumption prior to the study. Further, a dose-response relationship of alcohol use on blood pressure was not estimated by this review, presumably because of the variability in amounts and timing of the doses of administered alcohol.

Studies included in the other five reviews used a myriad of quantities, time frames, and terms to describe patterns of alcohol use. Literally dozens of approaches were used to define the amount of alcohol consumed, ranging from a simple yes/no measure to an idiosyncratic seven-category measure. The units of analysis in summary measures also varied from grams of alcohol per day to "drinks" per varying time periods. The diversity of approaches used to examine alcohol consumption is best demonstrated by the STD review where the authors were forced to examine four different summary categories of alcohol use measures¹². The depression review included only studies of persons with "alcohol problems" but had to create a detailed table of the definitions for the many terms used for these problems including: at risk, hazardous, harmful, abuse, dependence, and alcoholism⁷. In other reviews, the spectrum of alcohol consumption in the included studies ranged from rare to heavy, but both the diabetes and bone disease reviews

noted that few studies included women consuming larger amounts of alcohol⁹⁻¹⁰. In addition, five reviews critiqued many studies because the 'non-drinker' category combined persons who never drank alcohol with former users who may have previously suffered from an alcohol use disorder. Finally, few studies addressed the type of alcohol consumed.

Despite the ubiquity of alcohol consumption in the US and its potential to affect the diagnosis, management, outcomes, and costs of common chronic diseases, outside of the PRISM-sponsored reviews, we found few additional reviews in medical journals that systematically synthesized the evidence of the impact of alcohol drinking on diseases that are routinely treated by primary care and other mainstream physicians. The best evidence identified by these reviews was generally of mediocre quality as judged by established research quality measures. Except for hypertension, the reviews found few randomized clinical trials regarding the effect of alcohol on the selected diseases. Admittedly, randomized trials of alcohol consumption over an extended time frame may be impractical for some diseases. However, there is good reason for physicians to be skeptical of the results of simple observational designs concerning this or any other topic.

Beyond the fundamental design issues regarding the studies in these reviews, there are other important methodological flaws including limited adjustment for potential confounders. However, the most serious but potentially correctible flaw across all these reviews is the unacceptable variability in approaches to measure the quantity, frequency, and duration of alcohol use. To standardize experimental research with alcohol, Brick has proposed several mathematical approaches to determine ounces of pure (100%) alcohol provided in a single drink and over the course of an entire trial¹⁵. Observational studies included in these reviews frequently analyzed

Table 2. Alcohol Use Measures in Studies Included in Six Systematic Reviews

Domain	Depression	Hypertension	Diabetes mellitus	Bone disease	Dementia, cognitive decline in elderly	Sexually transmitted diseases (STDs)
Alcohol consumption measures	Qualitative Quantitative Alcohol problems (at-risk, abuse, dependence, alcoholism) Former +/- never user	Quantitative Sustained not episodic (binge) use No dose-response examined	Qualitative Quantitative Most group moderate + heavy use Most use measured only at baseline	Qualitative Quantitative Few data on >2 drinks per day Most use measured only at baseline	Qualitative Quantitative Most report drinks per varying time intervals	Qualitative Quantitative (Any/none, quantity/frequency, use in specific situations, problem drinking)
Time frame for measure	Current and lifetime	Few comment on baseline use before trial, washout period in half	Current	Current and lifetime	Current and lifetime	Current
Distinguish former from never user	Rarely	N/A	Rarely	Rarely	Rarely	Rarely
Type of alcohol used	No comment	Spirits or vodka	1 Study of beer drinkers	Not reported separately	12 Studies	1 Study
Summary alcohol use measures	None	Grams of alcohol/day	Drinks/day (one drink equals 12.6 g)	Drinks/day (alcohol in a drink varies by country)	2 Measures (any/none, baseline dose vs. none)	Drinker vs. non-drinker

the effect of alcohol as quantified by a 'standard drink.' The US Department of Agriculture (USDA) defines a standard drink as 13.7 g (0.6 ounces) of pure alcohol, which correlates to a 12-oz can of beer, 5-oz glass of wine, or 1.5-oz glass of distilled spirits¹⁶. However, in Australia, for example, a standard drink has 10 g of alcohol¹⁷, making syntheses of the health effects of alcohol from international studies more challenging. The USDA and the National Institute on Alcohol Abuse and Alcoholism (NIAAA)¹⁸ set the standard for "non-problematic" drinking for a healthy adult man as no more than 14 drinks per week and, for healthy women, as no more than 7 drinks per week. However, these two federal agencies differ on the amount that would be considered excessive for a single day. The USDA recommends only one drink for women and two for men, whereas the NIAAA defines excessive consumption (binging) as more than three drinks for women and four for men. Special attention to binge drinking is highly relevant to defining alcohol use disorders because researchers using the Behavioral Risk Factor Surveillance System showed that adding information about bingeing to an average daily alcohol consumption measure increased the relative prevalence of "heavy drinking" by up to 42%, depending on how bingeing is measured¹⁹. In addition to the need for standardized measures of alcohol consumption, researchers also must routinely assess the pattern, quantity, and time frame/duration of alcohol consumed at baseline by subjects in a trial, as noted in the hypertension review⁸.

Other non-standard approaches abound in research on the effects of alcohol consumption above acceptable levels. Terms used to describe these patterns of alcohol consumption include: alcohol problems, unhealthy alcohol use, alcohol use disorders, excessive drinking, alcohol abuse, alcohol dependence, and harmful drinking. The depression review provided a separate table just to clarify the definitions of the multiple measures used in the identified studies⁷. Current Diagnostic and Statistical Manual of Mental Disorders (DSM) IV categories add to the confusion with 15 codes used to describe alcohol consumption patterns, such as abuse, dependence, withdrawal, alcohol-related disorder, and intoxication²⁰. Hopefully, DSM V will correct this morass of terms that fail to even offer a category for potentially excessive alcohol consumption without currently evident negative health effects, such as for a woman drinking two glasses of wine a night. Research increasingly supports the use of the AUDIT instrument as a valid and reliable measure to identify a full range of alcohol use disorders ranging from abstinence through non-problem use to dependence²¹. Thus, standardization of the instruments used to define alcohol use disorders must also be a goal.

A serious threat to validity in studies of the health effects of alcohol occurs when the reference group of non-drinkers includes current abstainers who have suffered health consequences and ceased to drink²². Even the "never drinkers" may be heterogeneous because one study reported that over half of the persons who claimed to be lifetime abstainers had previously reported heavy to problematic use alcohol²³. Among drinkers, under-reporting alcohol consumption may present an even greater threat to the validity of research on the health effects of alcohol. Klatsky and colleagues reported that persons with hypertension who reported one to two alcoholic drinks a day were 75% more likely to have high liver transaminase enzymes than persons reporting no or less frequent use²⁴. These laboratory results suggest alcohol-related liver injury in

some of these 'moderate' alcohol drinkers. These challenges reinforce the need to continue to develop innovative approaches to reduce the stigma of reporting about alcohol use and to improve the validity of self-report data. A final dimension that most studies have not attempted to address is the type of alcohol consumed because it adds another layer of complexity in addition to the multiple alcohol use measures.

CONCLUSIONS

Based on the evidence synthesized by these reviews, we offer the following recommendations. First, wherever possible, research on health effects of alcohol needs to use experimental designs with a standard measure of pure alcohol consumed over a specific time frame. Specifically, we suggest the reporting of standard (NIAAA) drinks per day of beer or wine and of spirits over a 1-week period—permitting the calculation of total ounces of alcohol consumed within a report period that should be easy for participants to remember.

Second, observational studies need to be conducted in primary care populations using standard measures of quantity/frequency as well as maximum daily drinking (binging). Specifically, we recommend the use of valid and reliable measures such as the AUDIT or the shorter AUDIT-C.

Third, we recommend standardization of terms for alcohol use disorders based on the NIAAA guidelines, but, in the near future, DSM V may offer a better codification of these terms.

Fourth, the rigor of observational studies needs to be improved by measuring alcohol consumption at multiple points in time and by assessing a broad array of key confounders. The list of potential confounders varies by study outcome, but must at least include measures of tobacco and other drug use because of their correlation with alcohol use and the independent, powerful effects they can have on health²⁵. Because alcohol use may promote health or can have powerful negative effects that cause disease, research in the future needs to be more rigorous, relevant, and practical for clinical practice.

Acknowledgements and Conflict of Interest: Both authors gratefully acknowledge the support of the Robert Wood Johnson Foundation for PRISM and the support of the Betty Ford Foundation for the development of this paper. Dr. Turner receives support from Pfizer, Inc., through a grant to the University of Pennsylvania for unrelated research. Dr. McLellan reports no sources of other support or conflicts of interest.

Corresponding Author: Barbara J. Turner, MD, MEd; Division of General Internal Medicine, University of Pennsylvania, 1123 Blockley Hall/6021 423 Guardian Drive, Philadelphia, PA 19104, USA (e-mail: bturner@mail.med.upenn.edu).

REFERENCES

1. Institute of Medicine Committee on Quality of Health in America. The quality of health care for mental and substance-use conditions. Washington, DC: National Academy Press; 2006.
2. Centers for Disease Control and Prevention, National Center for Health Statistics Health, Health, United States. 2007. Chartbook on Trends in the Health of Americans, Hyattsville, MD: Library of Congress Catalog Number 76-641496

3. **Allen JP, Wilson VB.** Assessing Alcohol Problems: A Guide for Clinicians and Researchers. 2003; 2nd ed. NIH Publication No. 03-3745.
4. **Spandorfer JM, Israel Y, Turner BJ.** Primary care physicians' views on screening and management of alcohol abuse: inconsistencies with national guidelines. *J Fam Pract.* 1999;48:899-902.
5. **Vinson DC, Elder NC, Werner JJ, Vorel LA, Nutting PA.** Alcohol-related discussions in primary care: a report from ASPN. *J Fam Pract.* 2000;49:28-33.
6. **Rush BR, Urbanoski KA, Allen BA.** Physicians' enquiries into their patients' alcohol use: public views and recalled experiences. *Addiction.* 2003;98:895-900.
7. **Sullivan LE, Fiellin DA, O'Connor PG.** The prevalence and impact of alcohol problems in major depression: a systematic review. *Am J Med.* 2005;118:330-341.
8. **McFadden CB, Brensinger CM, Berlin JA, Townsend RR.** Systematic review of the effect of daily alcohol intake on blood pressure. *Am J Hyperten.* 2005;18:276-286.
9. **Howard AA, Arnsten JH, Gourevitch MN.** Effect of alcohol consumption on diabetes mellitus: A systematic review. *Ann Intern Med.* 2004;140:211-9.
10. **Berg KM, Kunins HV, Jackson JL, et al.** Association between alcohol consumption and both osteoporotic fracture and bone density. *Am J Med.* 2008;121:406-18.
11. **Peters R, Peters J, Warner J, Beckett N, Bulpitt C.** Alcohol, dementia and cognitive decline in the elderly: a systematic review. *Age Ageing.* 2008;37(5):505-12.
12. **Cook RL, Clark DB.** Is there an association between alcohol consumption and sexually transmitted diseases? A systematic review. *Sex Transm Dis.* 2005;32(3):156-64.
13. **Downs SH, Black N.** The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Comm Res.* 1998;52:377-384.
14. **Jordan KM, Arden NK, Doherty M, et al.** EULAR Recommendations 2003: an evidence based approach to the management of knee osteoarthritis: Report of a Task Force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCI-SIT). *Ann Rheum Dis.* 2003;62:1145-55.
15. **Brick J.** Standardization of alcohol calculations in research. *Alcohol Clin Exp Res.* 2006;30(8):1276-87.
16. United States Department of Agriculture and United States Department of Health and Human Services. In: *Dietary Guidelines for Americans.* Chapter 9 - Alcoholic Beverages. Washington, DC: US Government Printing Office; 2005, p. 43-46. Available at <http://www.health.gov/DIETARYGUIDELINES/dga2005/document/html/chapter9.htm> Accessed June 4, 2009
17. Australian Government Department of Health and Aging. The Australian Standard Drink. <http://www.alcohol.gov.au/internet/alcohol/publishing.nsf/Content/standard>, Accessed June 4, 2009
18. National Institute on Alcohol Abuse and Alcoholism. 2005. A Pocket Guide for Alcohol Screening and Brief Intervention. <http://pubs.niaaa.nih.gov/publications/Practitioner/PocketGuide/pocket.pdf>. Accessed June 4, 2009.
19. **Stahre M, Naimi T, Brewer R, Holt J.** Measuring average alcohol consumption: the impact of including binge drinks in quantity-frequency calculations. *Addiction.* 2006;101(12):1711-8.
20. <http://Psychweb.com>, http://psyweb.com/Mdisord/DSM_IV/jsp/dsmab.jsp accessed June 4, 2009
21. **Bradley KA, Bush KR, McDonnell MB, Malone T, Fihn SD.** The Ambulatory Care Quality Improvement Project (ACQUIP). Screening for problem drinking: Comparison of CAGE and AUDIT. *J Gen Intern Med.* 1998;13(6):379-88.
22. **Fan AZ, Russell M, Stranges S, Dorn J, Trevisan M.** Association of lifetime alcohol drinking trajectories with cardiometabolic risk. *J Clin Endocrinol Metab.* 2008;93(1):154-61.
23. **Rehm J, Irving H, Ye Y, Kerr WC, Bond J, Greenfield TK.** Are lifetime abstainers the best control group in alcohol epidemiology? On the stability and validity of reported lifetime abstinence. *Am J Epidemiol.* 2008;168(8):866-71.
24. **Klatsky AL, Gunderson EP, Kipp H, Udaltsova N, Friedman GD.** Higher prevalence of systemic hypertension among moderate alcohol drinkers: an exploration of the role of underreporting. *J Stud Alcohol.* 2006;67(3):421-8.
25. **Anthony JC, Echeagaray-Wagner F.** Epidemiologic analysis of alcohol and tobacco use. *Alcohol Res Health.* 2000;24(4):201-8.
26. **Jadad AR, Moore RA, Carroll D, et al.** Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled Clinical Trials.* 1996;17:1-12.
27. **Kmet LM, Lee RC, Cook LS.** Standard quality assessment criteria for evaluating primary research papers from a variety of fields. Alberta Heritage Foundation for Medical Research 2004.