APPENDIX A: TECHNICAL ESTIMATION DETAILS

In this appendix, we present additional technical details around the calculations we conducted to obtain estimates of total medical care costs that could have been avoided if alcohol consumption were reduced. First, we describe the formula used to produce estimates of the alcohol attributable fraction (AAF) (%) of female breast cancer cases, in which we separate drinkers from non-drinkers in the probability distribution over alcohol consumption. Second, we describe the quadrature routine that was used to approximate the integrals in our AAF formula. Third, we describe the approach took to account for underreporting of alcohol consumption in the National Survey on Drug Use and Health survey data used to measure the prevalence and distribution of alcohol consumption (a key input into the formula for AAFs). Finally, we describe the Monte Carlo approach that was used to assess uncertainty around all of our point estimates.

A1. The Alcohol-Attributable Fraction Formula

The basic formula for an AAF is:

$$AAF = \frac{\int_0^\infty RR(x) f_{X\,|\,X>0}(x\,|X>0)dx - 1}{\int_0^\infty RR(x) f_{X\,|\,X>0}(x\,|X>0)dx}. \tag{A1.1}$$

Where f(x) denotes the probability density function over alcohol consumption, and where RR(x) denotes the relative risk mapping from alcohol consumption levels to risk of breast cancer.

To account for the large proportion of survey respondents who report no alcohol consumption (i.e., non-drinkers), we substitute the continuous probability density function (PDF) in Eq. (A1.1) with a mixed discrete-continuous PDF. Specifically, we used

$$f_X(x) = \alpha + (1 - \alpha) \cdot f_{X\,|\,X>0}(x\,|\,X > 0), \tag{A1.2}$$

where α denotes the probability that X=0 (i.e., the probability of being a non-drinker). Since RR(0)=1 and because we can interchange integration and summation operations, substituting Eq. (A1.2) into Eq. (A1.1) results in the following alternative formula:

$$\mathrm{AAF} = \frac{\alpha + (1-\alpha)\cdot \int_0^\infty \mathrm{RR}(x) f_{X\,|\,X>0}(x\,|X>0)dx - 1}{\alpha + (1-\alpha)\cdot \int_0^\infty \mathrm{RR}(x) f_{X\,|\,X>0}(x\,|X>0)dx}. \qquad \text{(A1.3)}$$

Eq. (A1.3) is almost equivalent to the formula in Rehm et al.,[1] except that it makes clear that the continuous portion of the alcohol exposure distribution needs to be weighted by the probability that X>0 [i.e., (1–α)] so that the overall exposure distribution integrates one.

A2. Quadrature

To approximate the integral in Eq. (A1.3) we used Gauss-Legendre quadrature after applying a change of variables to transform the range of integration from $(0, \infty)$ to [-1, 1]. We found this approach to have better numerical stability than using Gauss-Laguerre quadrature after normalizing by exp(-x). The following change of variables changes the bounds of integration as desired:

$$x = \frac{1-t}{1+t} \qquad \text{(A2.1)}$$

Note that as t approaches -1, x approaches infinity. Also, as t approaches 1, x approaches 0. Accordingly, after the change of variables in Eq. (A2.1) the integral in Eq. (A1.3) can be rewritten as follows:

$$\int_0^\infty \mathrm{RR}(x)\ f(x)\ dx = -2\int_1^{-1} \mathrm{RR}\left(\frac{1-t}{1+t}\right)\ f\left(\frac{1-t}{1+t}\right)\ \frac{dt}{(1+t)^2}$$

$$= 2\int_{-1}^{1} \mathrm{RR}\left(\frac{1-t}{1+t}\right)\ f\left(\frac{1-t}{1+t}\right)\ \frac{dt}{(1+t)^2} \qquad \text{(A2.2)}$$

Equation (A2.2) can be approximated as follows:

$$2\int_{-1}^{1} \text{RR}\left(\frac{1-t}{1+t}\right)\ f\left(\frac{1-t}{1+t}\right)\ \frac{dt}{(1+t)^2} \ \cong\ 2\ \sum_{i=1}^{n} \text{RR}\left(\frac{1-x_i}{1+x_i}\right)\ f\left(\frac{1-x_i}{1+x_i}\right)\ \frac{w_i}{(1+x_i)^2} \qquad \text{(A2.3)}$$

where x_i and w_i are the nodes and weights for Gauss-Legendre quadrature.

A3. Accounting for Underreporting

As mentioned in the paper, we estimate subpopulation-specific AAFs for each subpopulation defined by age (18–25 years, 26–34 years, 35–49 years, and $\geq$50 years), race/ethnicity (white, non-Hispanic; black, non-Hispanic; other, non-Hispanic; and Hispanic), and insurance (Medicaid, private, and other/no insurance). We modeled the alcohol consumption distributions within each of these subpopulations using the best-fitting parametric distribution as determined by Kolmogorov-Smirnov testing.

In this section we present additional details on how distributions (generalized gamma, gamma, Weibull, and lognormal) were adjusted separately. We concluded the section by presenting a detailed, but simplified, numerical example of shifting the overall mixture distribution that utilizes more than one statistical distribution to model alcohol consumption across demographic sub-strata (e.g., females aged 21–25 years, 26–34 years, etc.). All of these shifting procedures depend on estimating a multiplier, m, which compares the overall survey average alcohol consumed with an estimate of the per capita volume of alcohol sold in the U.S. Specifically, we use:

m = Per capita sales/Average alcohol consumed. (A3.1)

Generalized Gamma

The PDF for the three-parameter generalized gamma distribution can be written as

$$f_X(x; \alpha, \beta, \theta) = \frac{|\alpha|\, x^{\alpha\beta - 1}}{\Gamma(\beta)\, \theta^{\alpha\beta}} \times \exp\left(-\left\{\frac{x}{\theta}\right\}^{\beta}\right), \qquad \text{(A3.2)}$$

where α and β are shape parameters, and θ is the scale parameter. Mittelhammer.[2] Γ(β) denotes the gamma function, which is defined as

$$\Gamma(\beta) = \int_0^{\infty} x^{\beta-1}\, e^{-x}\, dx. \qquad \text{(A3.3)}$$

We have found an alternative parameterization of the generalized gamma distribution presented in Manning et al.[3] to have smoother convergence properties. The parameterization in Manning et al. is

$$f_X(x; \kappa, \mu, \sigma) = \frac{\gamma^{\gamma}}{\sigma\, x \sqrt{\gamma}\, \Gamma(\gamma)} \times \exp\left(z\sqrt{\gamma} - u\right), \qquad \text{(A3.4)}$$

where $\gamma = \text{abs}(\kappa)^{-2}$, $z = \text{sign}(\kappa)\cdot\{\log(x) - \mu\}/\sigma$, and $u = \gamma\cdot\exp(\text{abs}(\kappa)\cdot z)$. There is no explicit scale parameter in the parameterization in Eq. (A3.4). However, Manning et al.[3] showed that the scale parameter in Eq. (A3.2) can be written as a function of the parameters in Eq. (A3.3). In particular,

$$\theta = \exp(\mu) \cdot \gamma^{-\left(\frac{\sigma}{\kappa}\right)}. \qquad \text{(A3.5)}$$

Thus one can estimate the parameters of Eq. (A3.4) and adjust the scale by adding the logarithm of the multiplier [i.e., Eq. (A3.1)] to the parameter μ, which is equivalent to multiplying θ by the multiplier. In summary, the two steps required to adjust the generalized gamma distribution are:

1. Estimate the generalized gamma parameters in Eq. (A3.4) using the observed survey data.
2. Adjust the parameter μ by *adding* the logarithm of the multiplier defined in Eq. (A3.1).

Gamma

The PDF for the standard gamma distribution can be written as:

$$f_X(x; \alpha, \theta) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} \cdot x^{\alpha-1} \exp\left(-\frac{x}{\theta}\right), \qquad \text{(A3.6)}$$

where α is a shape parameter and θ is the scale parameter. Since the parameterization in Eq. (A3.6) includes a single scale parameter, the gamma distribution can be adjusted by multiplying the scale parameter by the multiplier in Eq. (A3.1). In summary, the two steps required to adjust the gamma distribution are:

1. Estimate the gamma parameters in Eq. (A3.6) using the observed survey data.
2. Adjust the parameter θ by multiplying by the multiplier defined in Eq. (A3.1).

Weibull

The PDF for the Weibull distribution can be written as:

$$f_X(x;\alpha, \theta) = \frac{\alpha}{\theta}\left(\frac{x}{\theta}\right)^{\alpha-1} \times \exp\left(-\left\{\frac{x}{\theta}\right\}^{\alpha}\right), \qquad \text{(A3.7)}$$

where α is a shape parameter and θ is the scale parameter. Since the parameterization in Eq. (A3.7) includes a single scale parameter, the Weibull distribution can be adjusted by multiplying the scale parameter by the multiplier in Eq. (A3.1). In summary, the two steps required to adjust the Weibull distribution are:

3. Estimate the Weibull parameters in Eq. (A3.7) using the observed survey data.
4. Adjust the parameter θ by multiplying by the multiplier defined in Eq. (A3.1).

Lognormal

In the latter three cases, we adjust the distribution by multiplying the scale parameter by the multiplier defined in Eq. (A3.1). The lognormal distribution is slightly different. To see that a lognormally distributed random variable can be thought of as a transformation of a normally distributed random variable. If X is lognormally distributed with parameters μ and σ, then

$$X = \exp(Y), \quad \text{where } Y \sim N(\mu, \sigma^2). \tag{A3.9}$$

Accordingly, $X \cdot m = \exp[Y + \log(m)]$ for any constant m. Thus, mis-measured and accurately measured alcohol consumption can be related in terms of a normally distributed random variable, Y, as follows:

$$Y = \log(m) + Y^{*}. \tag{A3.10}$$

Eq. (A3.10) implies that Y is normally distributed with mean given by $\mu + \log(m)$ and variance given by σ^2. Combining this with Eq. (A3.9), it follows that a lognormally distributed random variable can be adjusted by adding the logarithm of the multiplier defined in Eq. (A3.1) to the location parameter, μ, rather than the scale parameter, σ, as we have done in all cases above. In summary, the two steps required to adjust the lognormal distribution are:

1. Estimate the lognormal parameters using the observed survey data.
2. Adjust the parameter μ by *adding* the logarithm of the multiplier defined in Eq. (A3.1).

A4. Details Around Monte Carlo Simulations

While Delta Method or bootstrapped SEs could be used here, we used a Monte Carlo approach to assess uncertainty around our point estimates (i.e., AAFs, attributable cases, and attributable medical care costs). In a Monte Carlo approach one simulates a distribution for all of the uncertain components within each calculation, and then repeats the calculation R times to simulate a distribution of estimates. The 2.5th and 97.5th percentiles provide the lower and upper bounds of a 95% CI, respectively. We used R=1,000 Monte Carlo repetitions.

The uncertain components include:

1. The parameters of the alcohol exposure distribution (including the probability of being a non-drinker);
2. The parameters of the dose-response function [RR(x)]; and
3. The incremental medical care cost estimates.

The total number of breast cancer cases are not estimated, but represent a complete census of cases that occurred in 2013. The parameters of the alcohol exposure distribution include α and the parameters of the continuous PDF used to model alcohol exposure among drinkers. We used the proportion of respondents who are non-drinkers to estimate α. The beta distribution is an appropriate statistical distribution to simulate proportions. We used the weighted number of non-drinkers as the first shape parameter, and the number of drinkers as the second shape parameter to generate pseudo-random beta variates. The parameters of the continuous PDF used to model alcohol consumption among drinkers are estimated using maximum likelihood models. Under standard assumptions these parameters are asymptotically normal. Thus we used estimates of the distributional parameters and associated variance-covariance matrix to generate pseudo-random multivariate normal variates. Similarly, the dose-response function parameter was estimated using a generalized least-squares trend estimator. Under standard assumptions the dose-response function parameter is also asymptotically normal. Accordingly, we used the parameter estimate and its associated variance estimate to generate pseudo-random normal variates. Finally, cost estimates are nonnegative and potentially skewed. A typical distributional assumption for medical costs is the gamma distribution. Appendix Table 3 summarizes this information.

Appendix Table 1. Average Number of Breast Cancer Cases From the U.S. Cancer Statistics Database by SEER Summary Stage and Age, 2012–2013[a,b]

	SEER summary stage[c]			
Breast cancer incident cases	**Localized**	**Regional**	**Distant**	**Overall**
18-44 years	11,448	8,952	1,281	21,681
45-64 years	62,964	30,373	5,786	99,123
Total	74,412	39,325	7,067	120,802

[a]Data are from selected statewide and metropolitan area cancer registries that meet the data quality criteria for all invasive cancer sites combined.
[b]Source: 2012 data: U.S. Cancer Statistics Working Group. United States cancer statistics: 1999–2012. Incidence and mortality web-based report. Atlanta, GA: U.S. DHHS, CDC, National Cancer Institute; 2015. 2013 data: U.S. Cancer Statistics Working Group. United States cancer statistics: 1999–2013. Incidence and mortality web-based report. Atlanta, GA: U.S. DHHS, CDC, National Cancer Institute; 2016.
[c]http://seer.cancer.gov/tools/ssm/

Appendix Table 2. Incremental Annual Medical Care Costs by Stage of Cancer at 12-months for Medicaid Beneficiaries and for Women With Private Health Insurance[a]

	Younger women, aged 18-44 years		Older women, aged 45-64 years	
Insurance type/breast cancer stage	**Incremental cost estimate (12-months)**	**95% CI**	**Incremental cost estimate (12-months)**	**95% CI**
Medicaid				
Localized	$46,616	($43,394-$49,837)	$28,674	($27,122-$30,226)
Regional	$59,431	($56,603-$62,260)	$45,288	($43,265-$47,311)
Distant	$93,471	($83,203-$103,739)	$62,868	($57,464-$68,271)
All stages	$66,596	($63,551-$69,641)	$45,914	($57,464-$68,271)
Private insurance				
Localized	$79,432	($74,885-$83,980)	$59,719	($57,910-$61,529)
Regional	$115,416	($110,034-$120,798)	$104,749	($101,431-$108,067)
Distant	$142,797	($117,509-$168,084)	$126,691	($111,529-$141,853)
All stages	$97,299	($93,443-$101,155)	$75,667	($73,893-$77,441)

[a]Medical care costs were adjusted to 2014 dollars using a gross domestic product deflator.[4]

Appendix Table 3. Uncertain Parameters and the Statistical Distribution Assumed in the Monte Carlo Analysis of Uncertainty

Uncertain parameter(s)	Statistical distribution
Probability of being a non-drinker	Beta
Prevalence distribution parameters	Multivariate normal
Dose-response parameter	Normal
Incremental medical care cost estimates	Gamma

Supplementary Tables

In this section we present supplementary result tables used as inputs into our calculations. Specifically, we present: (1) the dose-response modeling results; (2) results determining the best-fitting distribution for alcohol consumption among each modeled subpopulation as well as model predicted average daily consumption before and after adjusting for underreporting; (3) substrata-specific AAFs; (4) the total number of breast cancer incident cases obtained from the U.S. Cancer Statistics database; and (5) estimated incremental medical care costs by stage of cancer at 12-months for women who are Medicaid beneficiaries and those with private health insurance.

The dose-response modeling results can be summarized as follows:

$$\begin{array}{cc} \log RR(x) = & 0.0097784x \\ (S.E.) & (0.000427) \end{array}$$

Appendix Table 4. Best-fitting Statistical Distributions for Alcohol Consumption Model by Race/Ethnicity, Age Group, and Insurance Status

Stratum	Model choice	Model predicted average (before adjustment)[a]	Model predicted average (after adjustment)[a]
Female; white, non-Hispanic; 18-25 years; Medicaid	Generalized gamma	10.82	24.85
Female; white, non-Hispanic; 18-25 years; private	Generalized gamma	12.18	27.99
Female; black, non-Hispanic; 18-25 years; Medicaid	Lognormal	8.62	19.81
Female; black, non-Hispanic; 18-25 years; private	Lognormal	9.27	21.29
Female; Hispanic; 18-25 years; Medicaid	Weibull	12.01	27.61
Female; Hispanic; 18-24 years; private	Gamma	8.21	18.87
Female; other, non-Hispanic; 18-25 years; Medicaid	Lognormal	8.34	19.16
Female; other, non-Hispanic; 18-25 years; private	Generalized gamma	10.30	23.66
Female; white, non-Hispanic; 26-34 years; Medicaid	Lognormal	8.69	19.98
Female; white, non-Hispanic; 26-34 years; private	Lognormal	10.09	23.19
Female; black, non-Hispanic; 26-34 years; Medicaid	Lognormal	9.81	22.54
Female; black, non-Hispanic; 26-34 years; private	Lognormal	7.83	17.99
Female; Hispanic; 26-34 years; Medicaid	Generalized gamma	7.79	17.91
Female; Hispanic; 26-34 years; private	Lognormal	6.14	14.10
Female; other, non-Hispanic; 26-34 years; Medicaid	Weibull	8.22	18.88
Female; other, non-Hispanic; 26-34 years; private	Lognormal	9.62	22.10
Female; white, non-Hispanic; 35-49 years; Medicaid	Generalized gamma	9.51	21.86
Female; white, non-Hispanic; 35-49 years; private	Generalized gamma	8.76	20.13
Female; black, non-Hispanic; 35-49 years; Medicaid	Lognormal	13.16	30.23
Female; black, non-Hispanic; 35-49 years; private	Lognormal	5.55	12.75
Female; Hispanic; 35-49 years; Medicaid	Gamma	10.19	23.42
Female; Hispanic; 35-49 years; private	Lognormal	3.85	8.84
Female; other, non-Hispanic; 35-49 years; Medicaid	Weibull	7.58	17.42
Female; other, non-Hispanic; 35-49 years; private	Lognormal	6.64	15.25
Female; white, non-Hispanic; ≥50 years; Medicaid	Lognormal	8.75	20.11
Female; white, non-Hispanic; ≥50 years; private	Lognormal	8.82	20.26
Female; black, non-Hispanic; ≥50 years; Medicaid	Gamma	4.42	10.15
Female; black, non-Hispanic; ≥50 years; private	Lognormal	5.39	12.39
Female; Hispanic; ≥50 years; Medicaid	Weibull	4.25	9.77
Female; Hispanic; ≥50 years; private	Gamma	4.27	9.81
Female; other, non-Hispanic; ≥50 years; Medicaid	Generalized gamma	3.27	7.52
Female; other, non-Hispanic; ≥50 years; private	Lognormal	7.58	17.42

Appendix Table 5. Estimates of Alcohol-attributable Fractions by Race/Ethnicity, Age Group, and Insurance Status

Stratum	Medicaid		Private insurance	
	Estimate, %	95% CI	Estimate, %	95% CI
Female; White, non-Hispanic; 18-25 years	9.69	8.52-10.98	16.91	15.54-18.23
Female; Black, non-Hispanic; 18-25 years	8.43	7.26-9.89	11.51	9.95-13.02
Female; Other, non-Hispanic; 18-25 years	6.97	3.98-10.26	16.89	14.99-18.94
Female; Hispanic; 18-25 years	8.57	7.03-10.30	14.49	12.24-16.92
Female; White, non-Hispanic; 26-34 years	10.61	8.95-12.44	17.60	16.34-18.83
Female; Black, non-Hispanic; 26-34 years	10.30	7.54-13.24	13.22	10.71-15.93
Female; Other, non-Hispanic; 26-34 years	9.98	5.36-15.28	11.31	8.63-14.40
Female; Hispanic; 26-34 years	9.45	5.00-14.46	14.84	11.85-18.62
Female; White, non-Hispanic; 35-49 years	8.83	7.31-10.37	15.12	14.40-15.88
Female; Black, non-Hispanic; 35-49 years	6.25	2.15-9.73	11.53	9.71-13.41
Female; Other, non-Hispanic; 35-49 years	6.97	3.58-11.40	5.38	3.45-7.31
Female; Hispanic; 35-49 years	10.51	4.14-16.95	9.50	7.41-11.19
Female; White, non-Hispanic; ≥50 years	7.52	6.83-8.19	8.10	7.52-8.76
Female; Black, non-Hispanic; ≥50 years	6.77	4.12-9.88	3.71	2.34-4.99
Female; Other, non-Hispanic; ≥50 years	6.48	3.17-11.07	5.57	2.97-9.13
Female; Hispanic; ≥50 years	1.21	0.22-2.59	2.01	0.64-3.31

Appendix Table 6. Alcohol-attributable Breast Cancer Incidence Cases and Medical Care Costs in Older Women Aged 45-64 Years by Stage at Diagnosis and Insurance Status

	Medicaid insurance		Private insurance		Medicaid and private insurance	
Outcome	**Estimate**	**95% CI**	**Estimate**	**95% CI**	**Estimate**	**95% CI**
The number of alcohol-attributable breast cancer incidence cases						
Localized	**2,193**	1,982-2,413	**3,169**[a]	2,945-3,418	**5,362**	5037-5,682
Regional	**1,058**	956-1,164	**1,529**[a]	1,421-1,649	**2,586**	2,430-2,741
Distant	**202**	182-222	**291**[a]	271-314	**493**	463-522
Overall	**3,453**	3,120-3,799	**4,989**[a]	4,636-5,381	**8,441**	7,930-8,944
Attributable annual medical care costs (in millions $)						
Localized	**$12.61**	$11.42-$13.83	**$146.90**[a]	($134.90-$159.60)	**$159.50**	$147.40-$172.30
Regional	**$13.33**	$12.13-$14.55	**$115.40**[a]	($106.10-$125.30)	**$128.70**	$119.10-$138.40
Distant	**$5.12**	$4.53-$5.71	**$21.99**[a]	($19.16-$25.20)	**$27.11**	$24.18-$30.35
Overall	**$31.06**	$28.37-$33.78	**$284.30**[a]	($261.70-$308.30)	**$315.40**	$292.40-$338.70

Notes: Estimates in boldface are statistically significantly different from zero (p<0.05).
[a]The estimate for women enrolled in Medicaid is statistically significantly different (p<0.05) from the corresponding estimate for women with private health insurance.

APPENDIX REFERENCES

1. Rehm J, Kehoe T, Gmel G, Stinson F, Grant B, Gmel G. Statistical modeling of volume of alcohol exposure for epidemiological studies of population health: the U.S. example. *Popul Health Metr*. 2010;8:3. https://doi.org/10.1186/1478-7954-8-3.
2. Mittlehammer RC. *Mathematical statistics for economics and business.* New York, NY: Springer-Verlag; 1996. https://doi.org/10.1007/978-1-4612-3988-8.
3. Manning WG, Basu A, Mullahy J. Generalized modeling approaches to risk adjustment of skewed outcomes data. *J Health Econ.* 2005;24(3):465-488. https://doi.org/10.1016/j.jhealeco.2004.09.011.
4. Table 1.1.9. Implicit Price Deflators for Gross Domestic Product. www.bea.gov/iTable/iTable.cfm?ReqID=9&step=1#reqid=9&step=3&isuri=1&903=13. Accessed March 9, 2016.