Supporting Table 1. ICD-9 and ICD-10 codes used for cohort selection and comorbidities

Disease	ICD-10 Code (started 10/1/2015)	ICD-9 Code
Cirrhosis, unspecified	K74.6x	571.5
Alcoholic cirrhosis	K70.3x	571.2
Alcohol use	F10.x, T51.x, K86.0, I42.6, K29.20, K29.21, G62.1	790.3, 425.5, 535.30, 535.51, 577, 357.5, 291.x, 303.x, 305.0x, V113.0, E860.01-.03, 980.x
Alcoholic liver disease	K70.x	571.1, 571.3
Upper GI bleed	K29.01, K29.21, K29.31, K29.41, K29.51, K29.61, K29.71, K29.81, K29.91, I85.01, I85.11, K22.11, K26.0, K26.2, K26.4, K26.6	456.0, 456.2, 530.82, 530.00, 530.01, 531.20, 531.21, 531.40, 531.41, 531.60, 531.61, 532.00, 532.01, 532.21, 532.40, 532.41, 532.60, 532.61, 533.00, 533.01, 533.20, 533.21, 533.40, 533.41, 533.60, 533.61, 534.00, 534.20, 534.21, 534.40, 534.41, 535.01, 535.11, 535.21, 535.31, 535.41, 535.51, 535.61, 537.83, 578.0, 578.1, 578.9,
Hepatitis C	B18.2	070.41, 070.44, 070.51, 070.54, 070.70, 070.71
Ascites	R18.8, K71.51	789.59, 789.50
Varices without bleed	I85.00, I85.10	456.1, 456.21
Variceal bleed	I85.01, I85.11	456.0, 456.2
Portal hypertension	K76.6	572.3
Hepatic encephalopathy	K72.91	572.2
Hepatorenal syndrome	K76.7	572.4
Hepatocellular carcinoma	C22.0	155.0
Diabetes	E08.xx, E10.xx, E11.xx, E13.xx	250.x
Acute kidney injury	N17.xx	584.5, 584.6, 584.7, 584.8, 584.9
Spontaneous bacterial peritonitis	K65.2	567.23
More specific alcohol use codes and definitions used in alcoholic cirrhosis ascertainment		
ICD-9 Codes		**Definitions**

790.30	Excessive blood alcohol level
425.5	Alcoholic cardiomyopathy
535.30, 535.31	Alcoholic gastritis, with and without bleeding
577	Alcoholic pancreatitis
357.5	Alcoholic polyneuropathy
291.x	Alcohol-induced mental disorders
303.x	Alcohol dependence syndromes
980.x	Toxic effect of alcohol
305.0x	Nondependent alcohol abuse
E860.0, E860.1	Accidental poisoning by alcohol
V113	Personal history of alcoholism

Supporting Table 2. Gender and portal hypertensive complication event rates.

	Baseline N=169,531				1-year				2-year			
Characteristic	**AC Men** N=44,611	**AC Women** N=21,442	**Non-AC Men** N=53,282	**Non-AC Women** N=50,196	**AC Men**	**AC Women**	**Non-AC Men**	**Non-AC Women**	**AC Men**	**AC Women**	**Non-AC Men**	**Non-AC Women**
HCV	21%	14%	35%	20%	30%	21%	42%	24%	32%	23%	43%	25%
Ascites	21%	23%	8%	7%	38%	40%	15%	14%	45%	46%	18%	17%
Hepatic Encephalopathy	5.3%	6.2%	1.2%	1.3%	19%	20%	5%	5%	26%	26%	7%	7%
Variceal bleeding	4,2%	2.3%	1.3%	1.0%	11%	7.6%	4%	3.1%	15%	11%	5%	4.5%
GI Bleeding	16%	14%	8%	7%	29%	27%	14%	13%	36%	33%	18%	17%
HCC	2%	<1%	3%	1.3%	7%	3%	7%	3%	9.4%	4%	9%	4%
SBP	<1%	<1%	<1%	<1%	4%	4%	1%	1%	6%	5%	2%	1%
HRS	<1%	1%	<1%	<1%	4%	4%	1%	1%	6%	6%	2%	1%
AKI	8%	8%	5%	3%	20%	18%	10%	8%	27%	24%	13%	10%
Decompensation[@]	27%	29%	10%	9%	47%	47%	19%	17%	55%	53%	23%	21%

AC: alcoholic cirrhosis; HCV: Hepatitis C virus; HCC: hepatocellular carcinoma; SBP: spontaneous bacterial peritonitis; HRS: hepatorenal syndrome; AKI: acute kidney injury
*P<0.05 for all between-group comparisons (AC versus non-AC). [§]Gastroenterology outpatient visits were defined as a single code for an outpatient visit with a GI specialist.
[@]Decompensation defined as at least 1 occurrence of ascites, hepatic encephalopathy, or variceal bleeding.

Appendix B: Supplementary Methods

Prevalence estimation

For prevalence estimation, the ESI population was stratified into 32 strata based on the following four factors: age (18-45 or 45-64), sex, census region (4 categories), and employee status (policy holder versus covered dependent). For each of these strata, and for each year from 2009-2015, the total ESI population size was calculated from the Medical Expenditure Panel Survey(1). For each stratum, we estimated the cirrhosis or ALD/cirrhosis prevalence in Marketscan as discussed below, and then multiplied this prevalence estimate by the corresponding ESI stratum size to determine a per-stratum population count. These counts were then summed to produce an overall estimated prevalence count per year. This prevalence count was then divided by the ESI population size to produce a prevalence rate.

Estimation of the per-stratum Marketscan ESI prevalence is complicated by the need to reduce bias resulting from differing lengths of observation among the subjects. If we consider a person's cirrhosis status on a particular date to be determined by having had a cirrhosis code on or prior to that date in our data, the prevalence estimates for the earlier years will be biased downward, due to shorter lengths of preceding data. This leads to two potential areas of bias in the data: 1) under-reported prevalence and 2) an artificially steep estimate of trend increase. The first bias, of under-reported prevalence, is due to subjects being diagnosed with cirrhosis in one year, remaining enrolled in the dataset but not interacting with the healthcare system in subsequent years, and then re-appearing in the dataset once a claim is entered (for example, a patient is diagnosed with cirrhosis in 2009, remains enrolled but does not interact with the healthcare system in 2010, and then does so in 2011 when a claim appears in the dataset). Because cirrhosis is a chronic condition that does not wax and wane but, once present remains present, failure to account for these "skipped" years would produce artifically low prevalence estimates. The second bias results from underestimation of cirrhosis prevalence in the early years of the cohort because we are unable to include patients with diagnoses in 2008 or earlier, which is outside our dataset window. This produces an artificially steep estimate of the rate of increase in cirrhosis prevalence, since the later years in our data record will be less biased than the earlier years.

To mitigate these two biases, we adjusted results as follows: To estimate yearly cirrhosis prevalence, for each year from 2009-2015 we selected the people who were enrolled for the entire year, and directly calculated the proportion of these people who had a cirrhosis code during that year in each stratum. These rates underestimate the true prevalences since people with cirrhosis can go a year or longer without receiving a cirrhosis code, as discussed above, even while they remain covered. Therefore, we used the 2015 data to estimate the downward bias in this rate so that we can correct for it. We considered everyone who was covered for all of 2015 (the year for which we have the greatest length of historical data) and then calculated two rates: one based on subjects having a cirrhosis code in 2015, and one based on subjects having a cirrhosis code in any year from 2009-2015. The ratio of these rates (approximately 1.7) reflects the undercount when considering only people who have cirrhosis codes in a particular year. To produce our final prevalence estimates, we adjusted the per-stratum Marketscan counts by this factor prior to projecting them to the ESI population.

Cost analysis

Raw per-person costs were calculated by summing the "netpay" values in the Marketscan O, S, and D tables. Costs were summed over the first year following diagnosis, capping the total cost at 1 million dollars. A "time" value was calculated as the duration of coverage from diagnosis up to 1 year. Subjects who lost coverage or who died during the first year were coded with time equal to their coverage/survival duration, all other subjects had time=1. We then fit linear regression models to model the costs as a function of numerous controls. Forward variable selection was used to select main effects and interactions from the following variables: year of diagnosis, time (as described above), female, age at diagnosis, Elixhauser score, ascites, variceal bleed, hepatic encephalopathy, alcoholic liver disease, gastrointestinal bleeding, liver cancer, hepatitis C, HRS or AKI, portal hypertension, liver transplant, SBP, and census region. All time-varying covariates were coded based on their value at the end of the cost-summation period (1 year for most subjects). The final model had 32 terms in all. We then used the fitted values from this model to produce a modeled cost for each Marketscan person. When forming these fitted values, the time variable was reset to 1 for everyone who did not die (thereby accounting for censoring), but remained at the survival duration for people who died.

These predicted log costs were then averaged within subgroups of interest using 2012 weights (derived as in the prevalence analysis) to account for discrepancies between the Marketscan and ESI populations. Subgroups of interest included, for example, subjects who do and who do not have ascites during their first year post-diagnosis. This allowed us to compare average per-person costs for these subgroups in a way that reflects the ESI population, and is minimally biased due to differing follow-up times. These cost estimates do not attempt to isolate a given risk factor from other factors that correlate with it. For example, the difference in mean costs between people with and without ALD is inflated by costs of ALD-associated comorbidities, such as ascites or variceal bleeding.

We also calculated regression effects for factors of interest, expressing the results as the percent change in cost associated with each risk factor. Since the regression analysis considers cost on the log-scale, associations between individual factors, such as ascites, and costs can be calculated directly from the regression slopes, and presented as percent changes, e.g. the percentage difference in total costs associated with having, versus not having ascites.

Admissions and readmissions analysis

The admissions and readmissions analyses used a similar model-based strategy as the cost analyses. Negative binomial regression models were used for admission counts in place of the linear model used in the cost analysis. Regression effects, controlling for all other modeled factors were estimated as in the cost analysis. In addition, comparisons between risk groups of interest were conducted in two ways. First, as in the cost analysis, the weighted mean modeled admissions counts in two subgroups were computed, e.g. for comparing people who do and who do not have ascites within the first year after cirrhosis diagnosis. The results are presented as excess events per 100 subjects, meaning the difference between the number of admissions for 100 subjects with ascites and for 100 subjects without ascites. These differences are referred to as the “excess admissions per 100 cases”. Note that these comparisons are influenced by the factor of interest (e.g. ascites) and by other risk factors correlated with it.

References

1. Medical Expenditure Panel Survey. https://meps.ahrq.gov/mepsweb/.