APPENDIX A. SEARCH STRATEGY

Key Question 1 (Screening)

Medline (via PubMed) Searched 1/29/2013.

HCC	"hepatocellular carcinoma"[Title/Abstract] *OR* "HCC"[Title/Abstract] *OR* "Carcinoma, Hepatocellular"[Mesh] *OR* "liver cancer"[Title/Abstract] *OR* "Liver Neoplasms"[Mesh]
AND	
Screening	"alpha-Fetoproteins"[Mesh] *OR* alpha-fetoprotein*[Title/Abstract] *OR* alfa-fetoprotein*[Title/Abstract] *OR* alpha-foetoprotein*[Title/Abstract] *OR* alfa-foetoprotein*[Title/Abstract] *OR* alpha-fetalprotein*[Title/Abstract] *OR* alfa-fetalprotein*[Title/Abstract] *OR* alpha fetalprotein*[Title/Abstract] *OR* alfa fetalprotein*[Title/Abstract] *OR* alphafetoprotein*[Title/Abstract] *OR* alfafetoprotein*[Title/Abstract] *OR* alphafoetoprotein*[Title/Abstract] *OR* alfafoetoprotein*[Title/Abstract] *OR* alphafetalprotein*[Title/Abstract] *OR* alfafetalprotein*[Title/Abstract] *OR* CT[Title/Abstract] *OR* "Tomography, X-Ray Computed"[Mesh] *OR* mri[Title/Abstract] *OR* magnetic resonance imaging[Title/Abstract] *OR* "Magnetic Resonance Imaging"[Mesh] *OR* screen-detected[Title/Abstract] *OR* screening[Title/Abstract] *OR* "Mass Screening"[Mesh] *OR* "Early Detection of Cancer"[Mesh] *OR* ultrasonography[Title/Abstract] *OR* "Ultrasonography"[Mesh] *OR* "ultrasonography"[Subheading]

Key Question 2 (Harms of Screening)

(Note: medical adverse effects of screening would be included in above search. An additional search was designed to capture psychological harms of screening specifically.)

Medline (via PubMed) Searched 3/5/2013.

HCC	"hepatocellular carcinoma"[Title/Abstract] *OR* "HCC"[Title/Abstract] *OR* "Carcinoma, Hepatocellular"[Mesh] *OR* "liver cancer"[Title/Abstract] *OR* "Liver Neoplasms"[Mesh]
AND	
Psych harms	"False Positive Reactions"[Mesh] *OR* "False Negative Reactions"[Mesh] *OR* "Anxiety"[Mesh] *OR* "Depression"[Mesh] *OR* "Stress, Psychological"[Mesh] *OR* "Patient Acceptance of Health Care"[Mesh] *OR* "psychology"[Subheading]

An additional search for psychological harms of screening was conducted in PsycInfo (via OVID) on 6/28/2013:

((("hepatocellular carcinoma"[Title/Abstract]) *OR* "HCC"[Title/Abstract])) *OR* ("Carcinoma, Hepatocellular"[Mesh])*OR*(("liver cancer"[Title/Abstract])) *OR* ("Liver Neoplasms"[Mesh])

The search of PsycInfo yielded 160 citations; none were found to be relevant.

Key Question 3 (Treatment)

Medline (via PubMed) Searched 3/5/2013.

HCC	"hepatocellular carcinoma"[Title/Abstract] OR "HCC"[Title/Abstract] OR "Carcinoma, Hepatocellular"[Mesh] OR "liver cancer"[Title/Abstract] OR "Liver Neoplasms"[Mesh]
AND	
Treatment	ablation[Title/Abstract] *OR* "Ablation Techniques"[Mesh] *OR* hepatectomy[Title/Abstract] *OR* resection[Title/Abstract] *OR* excision[Title/Abstract] *OR* "Hepatectomy"[Mesh] *OR* Sorafenib[Title/Abstract] *OR* Nexavar[Title/Abstract] *OR* "sorafenib"[Supplementary Concept] *OR* transplant[Title/Abstract] *OR* transplantation[Title/Abstract] *OR* "Liver Transplantation"[Mesh] *OR* treatments[Title/Abstract] *OR* treatment[Title/Abstract] *OR* "Therapeutics"[Mesh] *OR* "therapy"[Subheading]
AND	
Mortality	mortality[Title/Abstract] *OR* survival[Title/Abstract] *OR* "Mortality"[Mesh] *OR* "mortality"[Subheading] *OR* "Survival Rate"[Mesh]

Additional searches:

The searches developed for MEDLINE were adapted for the Cochrane Central Register of Controlled Trials and searched on 6/28/2013.

ClinicalTrials.gov was searched on 9/3/2013 with the term "Hepatocellular Carcinoma" and no limitations on study type, recruitment status, etc.

APPENDIX B. FIGURES

Figure 1. Analytic Framework

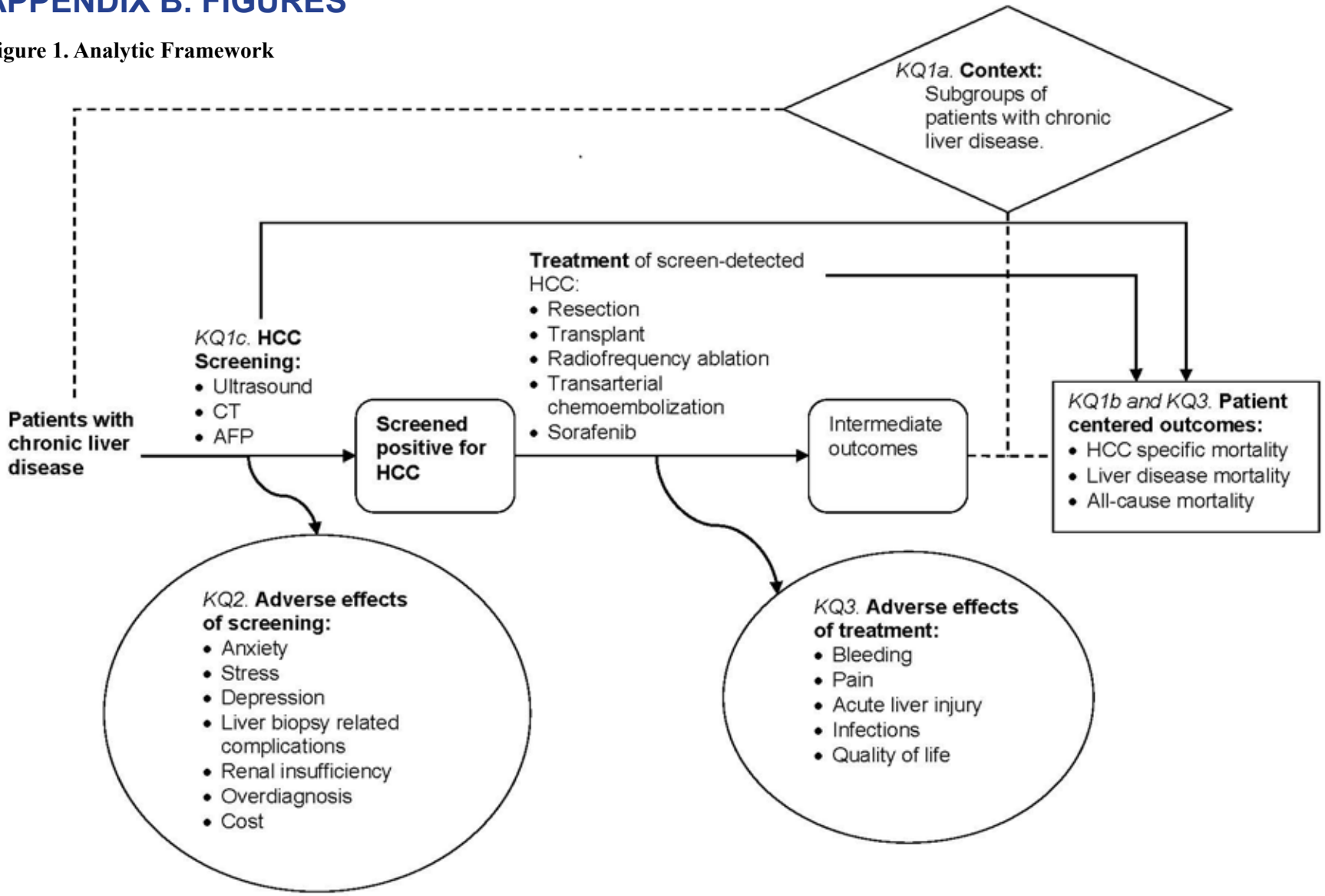


Figure 2. Literature flow diagram

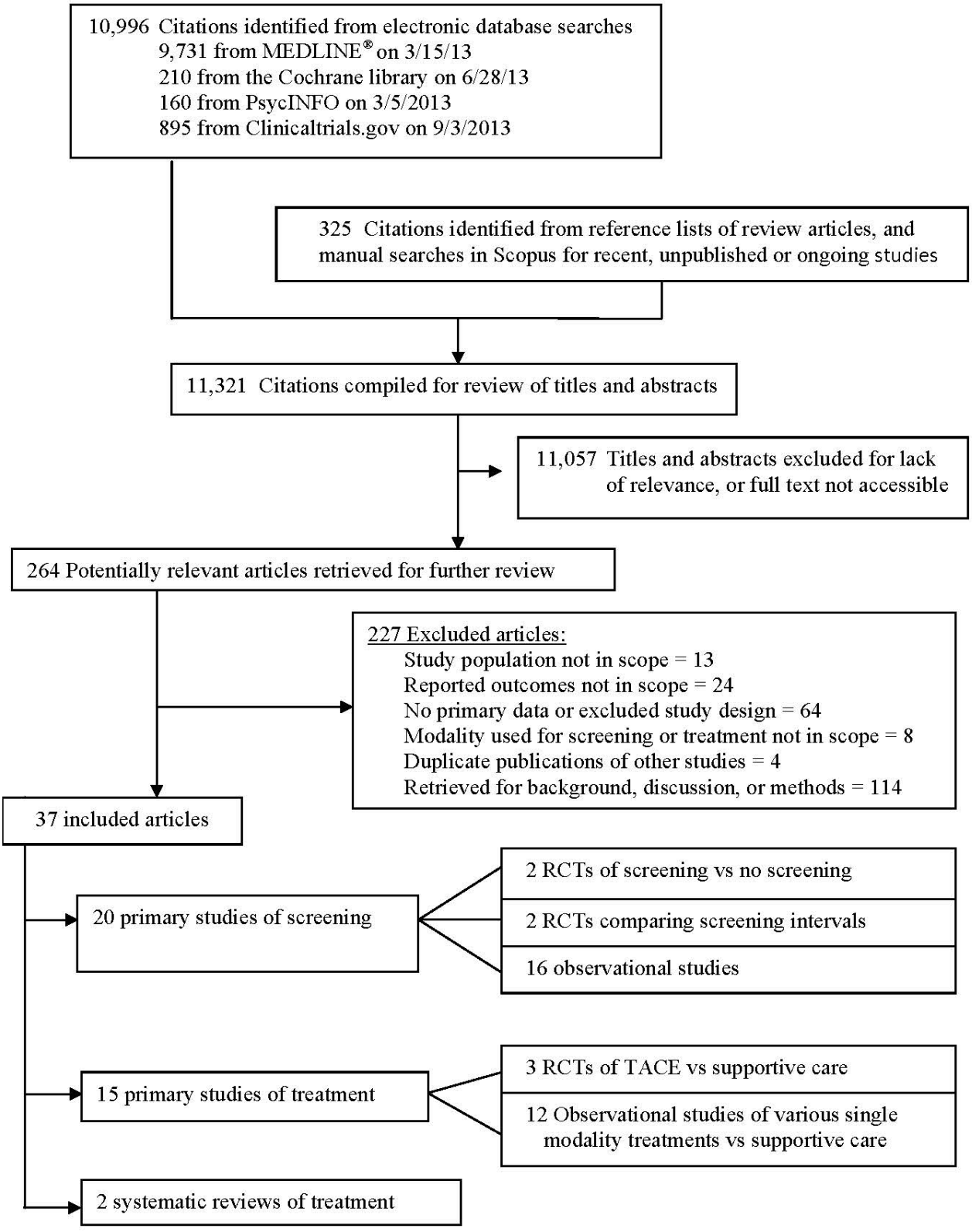
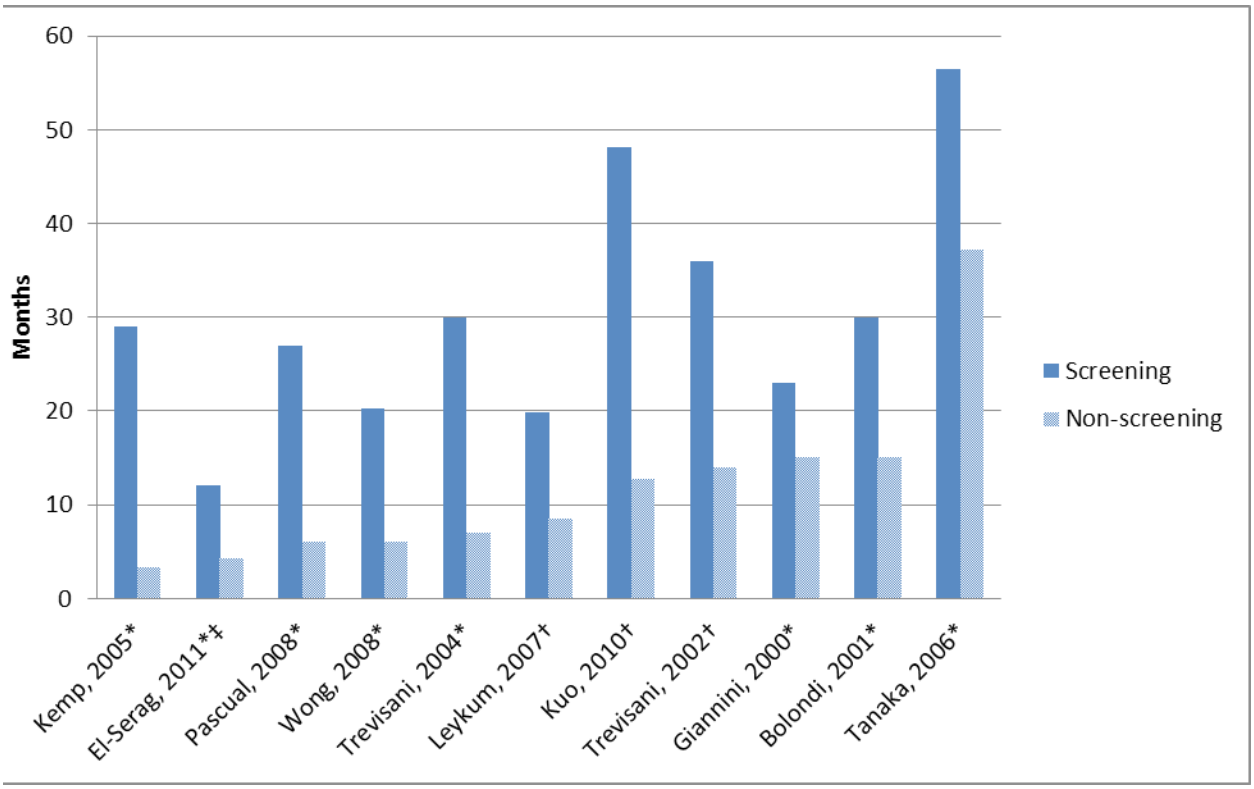


Figure 3. Median survival in cohort studies of HCC patients diagnosed through screening programs compared with non-screening



*P < 0.05

†P-value was not reported.

‡Screening group includes patients screened at both 0-6 and 7-24 month intervals before HCC diagnosis.

APPENDIX C. INCLUSION/EXCLUSION CRITERIA

Code	Definition	Exclusion criteria/notes	Screening studies inclusion criteria	Treatment studies inclusion criteria
I-Screening **I-Treatment**	Include – screening Include – treatment		**KQ1 –Benefits of screening:** 1a. In which subgroups of patients with chronic liver disease have the effects of HCC screening on patient survival been evaluated? 1b. What are the effects of HCC screening on disease-specific and all-cause mortality in these patient subgroups? 1c. Are there particular HCC screening modalities that are more effective on patient survival than others? **KQ2 –Harms of screening:** 2. What are the harms of HCC screening among patients with chronic liver disease?	**KQ3** – What are the benefits and harms of treating early stage HCC?
I-SR	Include – systematic review		Systematic review or meta-analysis that addresses any of the key questions. Code X9-SR for comparative effectiveness reviews of treatment modalities.	
X1	Non-English language	Most foreign language abstracts have been filtered out, but can be retrieved for further review as needed.		
X2	Not relevant to HCC			
X3	Study population is not in scope for either screening or treatment KQs.	Exclude: Patients with prior, advanced, or metastatic HCC; in vitro studies.	Adults with chronic active viral hepatitis, alcohol-related liver disease, non-alcoholic steatohepatitis, non-alcoholic fatty liver disease, hemochromatosis, primary biliary cirrhosis, primary sclerosing cholangitis, autoimmune hepatitis – all with or without cirrhosis.	Patients with early stage HCC, defined as patients with the equivalent of BCLC Stage A (3 or fewer nodules, <3cm, or 1 nodule <5 cm). Milan criteria = early stage HCC. Include studies for which at least a portion of the population is Stage A and B (these studies may be useful for addressing harms).
X4	No primary data, or study design not in scope	Exclude: Non-systematic or narrative reviews, opinions, case studies, case series, quasi-experimental studies, or other excluded study designs.	Include studies that compare screened patients with unscreened patients, using any of the following study designs: • Observational studies, e.g., cohort or case-control designs • Controlled studies, e.g., RCT, controlled clinical trial, controlled before/after designs. Also include: active-controlled/head-to-head trials and observational studies that compare screening modalities or screening intervals. *For cost studies:* include primary data collected in U.S. settings. Exclude modeling and simulation studies, and primary studies in non-U.S. settings.	Included study designs: • Randomized, placebo-controlled trials comparing a single treatment or combination of included treatment modalities vs no active treatment/placebo/active screening without treatment (analogous to watchful waiting); • Observational studies of a single or combination treatment modality that: - include a comparison group of untreated HCC patients, and - have a sample size ≥ 100 patients (treated plus untreated) - adjust for potential confounders. Studies that do not examine the effects of potential confounders (age, sex, baseline liver disease) are excluded. *Specific exclusions for treatment studies:* Code X9 for head-to-head/active-controlled treatment trials; X9-SR for comparative effectiveness reviews; add combo (e.g. X9-combo) to indicate multiple treatment modalities. Use code X10 for observational studies include an untreated comparison group, and contain data on harms of treatment but sample size is <100 treated patients.

Code	Definition	Exclusion criteria/notes	Screening studies inclusion criteria	Treatment studies inclusion criteria
X5	Modality used for screening or treatment is not in scope	Excluded screening modalities: Biomarkers, thrombocytopenia, DNA/RNA analyses. Excluded treatment modalities: Exclude percutaneous alcohol injection (no longer in use, and not in 2010 guidelines). Specify excluded treatments as they occur in the screening process.	Ultrasound, CT, MRI, and/or alpha-fetal protein screening for primary HCC.	Early stage/curative treatments include resection, transplant, radiofrequency ablation, transarterial chemoembolization, and sorafenib.
X6	None of the reported outcomes are in scope	Exclude studies that do not report any of the outcomes of interest. Exclude diagnostic accuracy studies.	Benefits: • Mortality due to HCC, liver disease, or all causes Harms: • Psychological effects (eg, anxiety, stress, depression, labeling) • Liver biopsy-related complications (eg, bleeding, infection) • Renal insufficiency • Overdiagnosis (ie, identifying cancers that would not have caused disease undetected) • Cost – include primary data collected in U.S. settings.	Benefits: • Mortality • Quality of life Harms: • Hospitalizations • Bleeding • Pain • Acute liver injury • Infections • Quality of life • Reports of any adverse event
X7	Other reason: specify	Add comments or keywords as needed.		
X9		Exclude head-to-head/active-controlled treatment trials. Code X9-SR for systematic reviews/meta-analyses on comparative effectiveness. X9-combo, where applicable.		
X10		Exclude relevant observational studies on treatments with sample size <100 treated patients (we may pull these later if low yield of studies with n≥100).		
X11	Duplicate publication	Exclude older publications or conference proceedings that have been subsequently published as full-text articles		
Note: Excluded articles should each receive a single X code, according to the order listed. Articles coded for background ('B') should also receive an X code.				
B	Background	Add 'B' any of the above X codes (e.g., 'X6–B') if the article contains information that may be useful for the introduction, discussion, limitations, future research, or other contextual purposes. Add comments or keywords as needed.		

APPENDIX D. TABLES

Table 1. Randomized trials of hepatocellular carcinoma screening in patients with chronic liver disease

Author, Year, Setting Years of enrollment	Screening modality, frequency (months)	N, screening vs no screening	Demographics: mean age; % male; race	Etiology, %	Liver disease severity, %	Stage at Diagnosis, %	Treatment received, %	Mortality, screening vs no screening, or interval comparison
Chen, 2003[29] Asia: China (Qidong county, Jiangsu Province) 1989-1995	AFP+ALT, 6 vs None	3712 vs 1869	age: 41.0 vs 41.3	HBV: 100	NR	*I: 29.6 vs 6.0 II: 50.6 vs 53.0 III: 19.8 vs 41.0	NR	All-cause mortality rate (per 100,000): 1.842 person-yr vs 1.788 person-yr HCC mortality rate: 1.138 person-yr vs 1.113 person-yr, p=0.86
Zhang, 2004[28] Asia: China (Shanghai) 1993-1995	AFP+US, 6 vs None	9757 vs 9443	age: 42 vs 41 male: 62.6 vs 63.3	HBsAg+: 64.8 vs 63.8 HBsAg+ and history of hepatitis: 26.8 vs 28.0	NR	I: 60.5 vs 0 II:13.9 vs 37.3 III: 25.6 vs 62.7 p<0.01	Resection: 46.5 vs 7.5 TACE or PEI: 32.6 vs 41.8 Conservative treatment: 20.9 vs 50.7	HCC mortality (per 100,000): 83.2 vs 131.5, RR 0.63 (95% CI 0.41-0.98), p<0.01; NR; NR
Trinchet 2011[31] Europe: France 2000-2006	US, 3 vs US, 6	640 vs 638	age: 54 vs 55 male: 69.5 vs 68.7	HBV: 12.8 vs 12.2 HCV: 44.7 vs 43.6 EtOH: 39.4 vs 39 ; hemochromatosis: 0.8 vs 2.3 other: 2.3 vs 2.6	Child A or B: 100	Milan: 79.2 vs 71.4	OLT: 18.9 vs 4.3 resection: 5.7 vs 9.7 percutaneous ablation: 37.7 vs 44.3 TACE: 17 vs 12.3	**24mo survival: 95.8 vs 93.5; 60mo survival: 84.9 vs 85.8 Total mortality: 11.3 vs 12.1, p=0.38
Wang, 2013[30] Asia: Taiwan 2006-2010	US, 4 vs US, 12	387 vs 357	Group A: 4mo Group B: 12mo age: 63.8 vs 66.6, p<0.001 male: 47.8 vs 51.8 race: NR	HBV: 30 vs 25.2 HCV: 63 vs 67.2 HBV+HCV: 7 vs 7.6	NR	BCLC: Very-early: 37.5 vs 6.7 Early: 54.2 vs 66.6 Others: 8.3 vs 26.7 , p=0.02	Curative treatment (surgical resection or RFA): 54.2 vs 20, p=0.05	1 vs 2 vs 4yr cumulative survival: Group A: 95.8 vs 78.8 vs 57.4 Group B: 80 vs 64 vs 56, p=0.399; NR

*China Liver Disease Study Group classification. I-subclinical or early stage (no signs/sx, tumor usually <5 cm). II-moderate stage, intermediate between I and III. III-late stage (obvious cachexia, jaundice, ascites, or distant metastases) Confounders adjusted for in analysis: **EtOH, HCV, age, platelet count, bilirubin, AST, ALT, alk phos, GGT, albumin, PT and AFP.

Table 2. Assessment of study methods for potential sources of bias in randomized trials of hepatocellular carcinoma screening

Author, Year, Setting Years of enrollment	Sequence generation	Allocation concealment	Blinding (patients, personnel, outcome assessors)	Incomplete outcome data	Selective outcome reporting	Other sources of bias	Overall risk of bias
Chen, 2003[29] Asia: China (Qidong county, Jiangsu Province) RCT 1989-1995	Unclear: NR	Unclear: NR	Yes - personnel staging cancers, Probably no - all others Low for mortality outcome	Mortality data likely available for everyone. Mean duration f/u similar in both groups Low	Low	Low: Baseline characteristics similar, but only age, ALT and AFP levels reported.	Unclear
Zhang, 2004[28] Asia: China (Shanghai) RCT 1993-1995	Unclear: NR	Unclear: NR	Unclear: NR	High Unclear for what proportion survival data were available.	High Vital status data reportedly available, but all-cause mortality not reported.	High Sparse baseline data available to compare both groups. No statistical analysis done to account for effects of clustering.	High
Trinchet 2011[31] Europe: France RCT screening intervals 2000-2006	Low	Low	Low (no mention of blinding, but low risk of bias for mortality outcomes)	Low	Low, intention-to-screen analysis	Low - groups similar at baseline	Low
Wang, 2013[30] Asia: Taiwan RCT screening intervals 2006-2010	Low	Low	Unclear, probably no blinding. Patient survival followed by public health nurses for all patients, so probably low risk of bias for mortality outcome.	Unclear: NR Unclear how many patients were lost to follow-up and there was no mention of death registry to ensure complete follow-up of mortality outcomes.	High: ITT analysis probably done, but not specifically mentioned. Clustered trial and no mention of adjustment for clustering.	Low Groups comparable at baseline, other than higher age and bilirubin in control group (though similar on other liver disease severity markers). Demographic characteristics among those with HCC similar in both groups.	Unclear

Table 3. Cohort studies of screening for hepatocellular carcinoma in patients with chronic liver disease

Author, Year, Setting, Years of enrollment	Screening modality, frequency (months); N screening vs no screening	Demographics (age; % male; race) Etiology, %	Etiology, %	Liver disease severity, %	Stage at Diagnosis, %	Treatment received, %	Observed mortality, screening vs no screening	Adjusted mortality, screening vs no screening
Bolondi, 2001[32] Europe: Italy 1989-1991	US+AFP, 6 313 vs 104	age: 61.8 vs 63.8 male: 70.5 vs 67.3	screening group only: HBV: 17.6 HCV: 64.2 Alcohol: 25.2 Primary biliary cirrhosis: 3.2	Child-Pugh: A: 41.0 vs 38.5 B: 47.5 vs 49.0 C: 11.5 vs 12.5	Unifocal HCC: 80 vs 53, p<0.001 Diffuse/ infiltrative HCC: 10 vs 29, p<0.01	Resection: 9 vs 8 OLT: 26 vs 13, p<0.01 PEI: 24 vs 23 TACE+PEI: 10 vs 10 TACE: 31 vs 46, p<0.05	Median survival (m) 30 vs 15 (p<0.02) Survival (%) at 3yr: 45 vs 31.7	*
Chen, 2002[33] Asia: Taiwan 1991-1998	US, 3-12 4385 vs 458	age ≥ 50: 45.0 vs 43.3 male: 78.7 vs 59.8	HBV: 65.9 vs 67.0 HCV: 18.2 vs 14.9	NR, but only 7 had cirrhosis	NR	NR	Unadjusted HR 0.76 (95% CI 0.38-1.52)	Adjusted[a] HR 0.59 (95% CI 0.29-1.20)
Davila, 2007[34] U.S - 3 VAMCs (Houston, Tennessee Valley, Kansas City) 1998-2003	AFP, US, or CT, within 36mo of HCC diagnosis 44 vs 113	age <65: 77.3 vs 55.8 (p=0.01) age ≥ 65: 22.7 vs 44.3 white: 68.1 vs 55.8	HBV: 6.8 vs 8.0 HCV: 72.7 vs 47.8 ETOH: 40.9 vs 14.2	Child-Pugh: A: 15.9 vs 26.5 B: 52.3 vs 35.4 C: 31.8 vs 38.1	One mass: 52.3 vs 38.1 2-3 masses: 22.7 vs 27.4 >3 masses: 18.2 vs 22.1	treatment n=54: Resection: 18.5 RFA: 11.1 PEI: 1.9 TACE: 35.2 chemotherapy: 31.5	Survival (%) at 1yr: 39 vs 31 3yr: 30 vs 21 (p=0.07)	*
El-Serag, 2011[35] U.S. (national VA HCV registry) 1998-2007	US and/or AFP, within 24mo of HCC diagnosis 1148 vs 332	age: 58.1 male: 99.3 white: 55.6	HCV:100	NR (but measured)	NR	NR	Unadjusted HR (95% CI) from date of HCC diagnosis, by timeframe screened during 24m prior to HCC diagnosis: 7-24m: 0.84 (0.69-1.01) 0-6m: 0.80 (0.68-0.94) Both periods: 0.71 (0.62-0.82) Median survival (days) from date of HCC diagnosis among pts screened in both periods vs neither: 368 vs 130 (p<0.01) Unadjusted HR (95% CI) from date of HCV diagnosis: 7-24m: 0.86 (0.72-1.04) 0-6m: 0.90 (0.77-1.06) Both periods: 0.82 (0.72-0.95) Median survival (days) from date of HCV diagnosis among pts screened in both periods vs neither: 1951 vs 1782	Adjusted[b] HR (95% CI) by timeframe screened during 24m prior to HCC diagnosis: 7-24m: 0.93 (0.77-1.13) 0-6m: 0.93 (0.79-1.09) Both periods: 0.84 (0.72-0.98) Adjusted HR corrected for lead time, assuming HCC sojourn time of 140 days: 7-24m: 1.04 (0.87-1.26) 0-6m: 1.00 (0.85-1.17) Both periods: 0.88 (0.76-1.02)
Giannini, 2000[36] Europe: Italy 1993-1998	AFP+US, 6 34 vs 27	age: 67 vs 68	HCV: 100	Mean Child-Pugh: 6 vs 8	One mass: 58.8 vs 51.9 >2 masses: 41.2 vs 48.5	Resection: 11.8 vs 7.4 OLT: 2.9 vs 0 PEI: 52.9 vs 33.3 TACE: 29.4 vs 25.9 None: 2.9 vs 33.3	Median survival (m) 23 vs 15 (p=0.03)	Adjusted[c] HR 0.38 (95% CI 0.17-0.87)

Author, Year, Setting, Years of enrollment	Screening modality, frequency (months); N screening vs no screening	Demographics (age; % male; race) Etiology, %	Etiology, %	Liver disease severity, %	Stage at Diagnosis, %	Treatment received, %	Observed mortality, screening vs no screening	Adjusted mortality, screening vs no screening
Kemp, 2005[37] Hospital, Victoria, Australia 1994-2002	US, 6-12 +AFP, 6 41 vs 55	age: 65 vs 68 male: 88.0 vs 78.2 Asian: 14.6 vs 16.7	HBV: 26.8 vs 12.9 HCV: 39.0 vs 29.6 Alcohol use: 43.9 vs 37.0	Child-Pugh: A: 63 vs 42 B: 27 vs 33 C: 10 vs 25	TNM I/II: 61.1 vs 21.7 III/IV: 38.9 vs 78.3, p<0.001	Resection: 11.8 vs 6.8 PEI or RFA: 52.9 vs 6.8 TACE: 33.0 vs 13.0	Median survival (m) 29.0 vs 3.3 (p<0.001)	Adjusted[d] HR 0.24 (p<0.0005)
Kuo, 2010[38] Asia: Taiwan 2002-2004	AFP+US, 12 318 vs 1118	age: 59.7 vs 59.4 male: 67.6 vs 76.4 (p=0.002)	HBV: 48.7 vs 47.1 HCV: 38.1 vs 33.4 HBV + HCV: 9.1 vs 7.8 Other: 4.1 vs 11.7	Child-Pugh: A: 73.3 vs 62.4 B: 23.9 vs 30.4 C: 2.8 vs 7.2 (p<0.001)	BCLC, p<0.001: Very early: 8.2 vs 6.5 Early: 60.4 vs 23.1 Intermediate: 21.7 vs 35.2 Advanced: 6.9 vs 30.9 Terminal: 2.8 vs 7.1	Resection: 23.9 vs 17.0 RFA: 12.6 vs 3.2 PEI: 9.1 vs 2.5 TACE: 47.2 vs 38.2 chemotherapy or radiation: 1.6 vs 12.3 None: 5.6 vs 26.7 (p<0.001)	Unadjusted HR 0.43 (95% CI 0.37-0.52) Median survival (m) 48.1 vs 12.7 Survival (%) at 3yr: 59.1 vs 29.3 (p<0.001)	Adjusted[e] HR 0.83 (95% CI 0.67-1.0)
Leykum, 2007[39] US. Michael DeBakey VAMC, Houston TX 2000-2005	2 AFP levels or one US/CT each year prior to diagnosis 16 vs 56	age: 59 vs 53.8 white: 64.2 vs 33.9	HBV: 40 vs 40 HCV: 100 ETOH: 0.68 vs 13.6	Child-Pugh: 6.3 vs 7.2	BCLC early: 100 vs 22, p<0.001	Resection: 6.3 vs 0 OLT: 6.3 vs 0 RFA: 50 vs 10.7	Unadjusted HR 0.27 (95% CI 0.13-0.60) Mean survival (m) 19.8 vs 8.5	Adjusted[f] HR 1.01 (95% CI 0.33-3.07)
Pascual, 2008[40] Europe: Spain 1996-2005	US+AFP, 6 117; NA	age: 68.8 vs 68.2 male: 66 vs 81 (p=0.002)	HBV: 3 vs 6 HCV: 61 vs 35 EtOH: 21 vs 35 EtOH + virus: 5 vs 11 (p<0.001)	Child-Pugh: A: 64 vs 33 B: 27 vs 48 C: 9 vs 19 (p<0.001)	<5cm: 60 vs 33 >5cm: 9 vs 28 multifocal: 14 vs 32 (p=0.003)	OLT: 15 vs 3 PEI: 19 vs 9 RF: 13 vs 4 TACE: 39 vs 20 none: 14 vs 64 (p<0.001)	Median survival (m) 27 vs 6 (p=0.001)	Adjusted HR[g] 0.4 (0.3-0.6), p=0.00003)
Tanaka, 2006[41] Asia: Japan 1991-2003	US+AFP, 6 182 vs 202	male: 60 vs 78	HCV: 100	Child-Pugh: A: 64 vs 58 B: 32 vs 39 C: 3 vs 3	Milan: 86 vs 50	Resection: 16 vs 12 PEI/RFA: 60 vs 34 TACE: 20 vs 42 Chemotherapy: 3 vs 9 (p<0.001)	Median survival (y) 4.7 vs 3.1 (p<0.001) Survival (%) at 3yr: 67 vs 51 5yr: 46 vs 32	Adjusted[h] RR 0.63 (95%CI 0.48–0.82). Corrected for lead time, survival was longer with screening among Child–Pugh class A patients when assumed tumor doubling time was ≤120 days: 60 days (p=0.005) 90 days (p=0.016) 120 days (p=0.048) 150 days (p=0.129) 180 days (p=0.293)

Author, Year, Setting, Years of enrollment	Screening modality, frequency (months); N screening vs no screening	Demographics (age; % male; race) Etiology, %	Etiology, %	Liver disease severity, %	Stage at Diagnosis, %	Treatment received, %	Observed mortality, screening vs no screening	Adjusted mortality, screening vs no screening
Taura, 2005[42] Asia: Japan 1991 – 2001	US, 3-12 AFP+liver function tests, 3-6 178 vs 93	age: 64.9 vs 64.3 male: 71.3 vs 85.0	HBV: 15.8 vs 15.0 HCV: 74.7 vs 69.9 HBV + HCV: 3.9 vs 1.1 Alcohol: 1.7 vs 4.3	Child-Pugh: A: 69.7 vs 74.2 B: 24.2 vs 20.4 C:6.1 vs 5.4	<3 cm: 64.6 vs 22.6 <5 cm: 94.4 vs 51.6 >3 tumors: 24.7 vs 45.2	Resection: 2.8 vs 3.2 RFA/PEI: 48.3 vs 17.2, p<0.0001 TACE:41.0 vs 59.2, (p=0.01)	Median survival overall (m): 37.3. Cumulative survival was significantly higher in screening vs no screening, NOS (p=0.01)	*
Tong, 2010[43] U.S. Pasadena, CA 1991-2008	US+AFP, 6 (cirrhosis, chronic liver disease) US+AFP, 12 (inactive carriers) 26 vs 52	age: 61.5 vs 52.9 (p=0.009) male: 80.8 vs 82.6	HBV: 100	Child-Turcott-Pugh: A: 65 vs 72.1 B: 25 vs 23.3 C: 10 vs 4.70	Milan: 61.5 vs 19.6, p=0.0004 UCSF: 76.9 vs 27.5, p<0.0001 tumors: Single: 81 vs 52 Multiple/diffuse: 19 vs 48 Metastasis: 7.7 vs 19.2 (p=0.02)	No screening vs screening: Resection: 19.2 vs 17.3 OLT: 30.1 vs 5.8 RFA and/or TACE: 26.9 vs 23.1 Chemotherapy: 0 vs 9.6 Supportive care: 23.1 vs 44.2 (p=0.012)	Survival (%) at 1yr: 100 vs 76.9 3yr: 62.5 vs 36.6 5yr: 35.7 vs 16.3 (p=0.007)	Adjusted[i] HR was non-significant, NOS. A lead time bias interval was added to the survival time of patients who presented with HCC, with tumor doubling time assumed to be 216 days.
Trevisani, 2002[44] Europe: Italy 1988-1998	US+AFP, 6 Group 1: semiannual screening, Group 2: annual screening Group 3: symptoms or incidental diagnosis 215 (group 1) vs 155 (group 2) vs 451 (group 3)	male: 70.7 vs 71 vs 78.7 (p=0.03)	HBV: 13.6 vs 20.4 vs 20.5 HCV: 66.6 vs 62.5 vs 55.9 HBV+HCV: 9.9 vs 9.9 vs 8.4 EtOH:8.5 vs 7.2 vs 13.8	Child-Pugh: A: 63.7 vs 70.9 vs 54 B: 30.7 vs 23.7 vs 33.8 C: 5.6 vs 5.4 vs 12.2 (p=0.001)	Milan: 68.7 vs 60.4 vs 31 (p<0.001)	OLT: 3.9 vs 0.2 resection: 11.6 vs 8.2 PEI: 26 vs 18.7 TACE: 33.4 vs 27.3 (p<0.001)	Median survival (m) 36 vs 34 vs 14 (p<0.001)	Adjusted[j] RR for Child-Pugh A subgroup: 0.59 (95% CI 0.45-0.78). Survival corrected for lead time was NS higher with screening in Child-Pugh B (p=0.051) and C subgroups (p=0.49).
Trevisani, 2004[45] Europe: Italy 1988-2001	Group 1: US+AFP, 6-12 Group 2: incidental diagnosis Group 3: symptoms 158 (group 1 vs 138(group 2) vs 67 (group 3)	age: 73.9 vs 74.9 vs 74.6 male: 60.8 vs 68.8 vs 76.1 (p=0.04)	HBV: 9.5 vs 6.5 vs 11.9 HCV: 67.1 vs 58.0 vs 53.7 HBV+HCV: 2.5 vs 3.6 vs 7.5 EtOH:5.7 vs 12.3 vs 10.4 EtOH+viral: 10.8 vs 10.9 vs 7.5	Child-Pugh: A: 76.8 vs 68.7 vs 42.4 B: 18.8 vs 29.8 vs 43.9 C: 4.6 vs 1.5 vs 13.6 (p<0.001)	Milan: 70.3 vs 39.1 vs 25.4 (p<0.001)	Resection: 8.4 vs 2.9 vs 0 PEI: 35.7 vs 36.8 vs 10.8 TACE: 28.6 vs 17.6 vs 20 Other/palliation: 27.3 vs 42.6 vs 69.2 (p<0.001)	Median survival (m) 30 vs 21(p=0.006) v 7 (p<0.001)	*

Author, Year, Setting, Years of enrollment	Screening modality, frequency (months); N screening vs no screening	Demographics (age; % male; race) Etiology, %	Etiology, %	Liver disease severity, %	Stage at Diagnosis, %	Treatment received, %	Observed mortality, screening vs no screening	Adjusted mortality, screening vs no screening
Wong, 2008[46] Asia: China (Hong Kong) 2003-2005	AFP, 6 US, 12-24 79 vs 393	age: 59.5 vs 58.7 male: 70 vs 88	overall HBV: 91 HCV: 10	Mean child-Pugh: 6.0 vs 6.4 (p=0.02)	Mean tumor, n: 2.6 vs 3.8 (p=0.03) Median tumor diameter (cm): 4.2 vs 7.7 (p<0.001) Extrahepatic metastasis: 8 vs 23 (p=0.002) Portal vein thrombosis: 11 vs 30 (p=0.001) Bilobal involvement: 14 vs 31 (p=0.01)	Resection: 20 vs 10, p=0.01 Transplant: 1 vs 1 Chemotx:13 vs 15 Local ablative tx: 46 vs 19, p<0.001	Median survival (wk) 88 vs 26 (p<0.001) Survival (%) at 1yr: 65.6 vs 35.5 2yr: 49.4 vs 21.1	Adjusted[k] HR 0.66 (95% CI 0.48-0.92) Survival (%) at 2yr: 49.4 in the screening group; correcting for lead-time bias in the non-screening group, by tumor doubling time: 26.7 (p=0.0035) 60-day 28.6 (p=0.035) 90-day 32.2 (p=0.18) 120-day
Yu, 2004[47] Asia: Taiwan 1996-1997	US, NR 164 vs 516	age % ≥50: 73.8 vs 65.9 male: 73.2 vs 79.3	HBV: 67.7 vs 53.57 HCV: 43.9 vs 31.3	Cirrhosis: 91.9 vs 68.2, Ascites: 10.1 vs 21.9	TNMS I: 66.2 vs 19.3 II: 27.2 vs 37.2 III: 3.7 vs 28.9 IV: 2.9 vs 14.6 (p<0.0001)	Hepatic resection: 53.5 vs 34 (p<0.0001) TACE: 35.1 vs 29.9	Unadjusted OR (95% CI) of survival at 1yr: 3.57 (5.26–2.38) 2yr: 3.70 (5.26–2.56) 3yr: 3.57 (5.26–2.44)	Adjusted[l] OR (95% CI) of survival at 1yr: 1.72 (2.86–1.03) 2yr: 2.22 (3.70-1.35) 3yr: 2.27 (3.85–1.37)

Abbreviations: (m) = months; NOS = not otherwise specified; NS = nonsignificant(ly).
* Potentially confounding variables were examined but an adjusted hazard ratio was not reported.
Confounders adjusted for in analysis:
[a] Age, sex, HBV, AST, AFP
[b] Screening test in the 3-6 years before HCC, year of diagnosis, age, race, MELD, psychosis, ascites, varices, encephalopathy
[c] Receipt of therapy, number of lesions, Child-Pugh
[d] Disease severity, cause, renal function, alcohol use, stage
[e] Etiology of disease, AFP level, solitary tumor, absence of portal vein thrombus, stage, surgical resection
[f] Psychiatric disease, PCP at tertiary center, hepatology assessment before diagnosis, early stage, receipt of potentially curative treatment.
[g] Child–Pugh status, tumor characteristics, treatment received
[h] AFP, Child-Pugh
[i] Single tumors, UCSF criteria, CTP class A, platelets per log10 increase, AST per log10 increase
[j] Sex, HBV, AFP
[k] Age, sex, and Child-Pugh
[l] Age, HBV, HCV, cirrhosis, ascites, ALT, AFP, and lead time adjustment.

Table 4. Assessment of study methods for potential sources of bias in cohort studies of screening for hepatocellular carcinoma in patients with chronic liver disease

Author, Year, Setting Years of enrollment	How was the screening group distinguished from non-screening?	Was this definition objective and replicable?	Loss to follow-up, difference in loss to follow-up between screened and unscreened?	Selection bias - are screening and nonscreening groups drawn from similar populations?	Ascertainment of outcomes adequately described and similar between surveilled and non-surveilled groups?
Bolondi, 2001[32] Europe: Italy	Screening group were patients prospectively enrolled in a screening program. Non-screening group was referred - possible that some of these patients were screened, but no data	Yes, for the screening group, not for the non-screening group.	Data for screening group only: Mean follow-up months: 56 7.7% lost to follow-up	Compared patients at an institution to referral patients.	Unclear
Chen, 2002[33] Asia: Taiwan	Those undergoing screening vs those who refused	Objective, but not valid.	NR	No - non-screening group were those that refused intervention.	Probably yes (national death registry)
El-Serag, 2011[35] U.S. (national VA HCV registry) 1998-2007	Used lab data and CPT codes to determine receipt of AFP or ultrasounds. Used an algorithm to determine whether AFP or US were performed for HCC screening based on lab data and ICD-CM codes.	Yes	NR, but unlikely that there was differential or high loss to follow-up as included all VA patients and they conducted sensitivity analyses using Medicare data for older patients	Yes	Yes - the date of death was obtained from the VA vital status file.
Davila, 2007[34] U.S - 3 VAMCs (Houston, Tennessee Valley, Kansas City)	Receipt of screening defined as having AFP, US, or CT within 3 years prior to HCC diagnosis. Detailed chart review used to assess intent of test. Tests performed for acute symptoms, during hospitalization, or to assess a mass were not considered screening.	Yes	NR	Yes	Yes
Giannini, 2000[36] Europe: Italy	Screening group defined as those who were receiving follow-up for cirrhosis. Control patients were referred patients or had tests done at "non-scheduled intervals"	No	NR	Unclear - control patients had tests done at "non-scheduled intervals" but it was unclear whether this meant they were enrolled in cirrhosis clinic but failed to present for testing or were not enrolled in a screening program.	No
Kemp, 2005[37] 1994-2002	Screening group were those treated by gastroenterology unit, which used regular screening. Unclear how unit of treatment was determined	No - it is not clear how patients were chosen for treatment by gastroenterology unit	Unclear	No - groups defined by treating unit which may treat different patient populations.	Unclear
Kuo, 2010[38] Asia: Taiwan	Screening group had AFP and US done as part of screening program and repeated within one year. Control group had HCC diagnosed because of symptoms or as part of another work-up, but it is not clear how they differentiated groups based on chart review	No	NR	Unclear - not enough detail about both groups. Unclear whether control patients were referred from outside institutions and why they would not have received screening.	Yes - national mortality dataset
Leykum, 2007[39] US. Michael DeBakey VAMC, Houston TX	Chart review. Screening group were those who received AFP or imaging in year prior to diagnosis and no alternative reason for testing was apparent from chart review.	Yes	NR, but unlikely that there was differential or high loss to follow-up as included all VA patients	Yes	Yes - VA patients

Author, Year, Setting Years of enrollment	How was the screening group distinguished from non-screening?	Was this definition objective and replicable?	Loss to follow-up, difference in loss to follow-up between screened and unscreened?	Selection bias - are screening and nonscreening groups drawn from similar populations?	Ascertainment of outcomes adequately described and similar between surveilled and non-surveilled groups?
Pascual, 2008[40] Europe: Spain 1996-2005	Screening group were patients seen in Liver Unit and diagnosed as part of their regular screening program. Non-screening group were either patients with cirrhosis diagnosed with HCC because of symptoms, or diagnosed with HCC at the time of cirrhosis diagnosis	No, it is unclear why some patients attending a screening program and others didn't. Also unclear procedures for cirrhosis work-up.	28 out of 290 patients were lost to follow-up but did not differentiate between screened and unscreened	No - some non-screening patients were referred from outside institutions and others at the institution did not attend screening program for unclear reasons.	Yes, through registry
Tanaka, 2006[41] Asia: Japan 1991-2003	Unclear - screening group patients were part of a screening program. Non-screening patients had HCC detected because of symptoms (12%), as a result of initial screening (11%), incidentally during other work-up (20%), and referred from outside hospitals (57%)	No - unclear how symptomatic detection was determined retrospectively and unclear how referral patients were surveilled.	None	Unclear - probably not, the majority of non-screening patients were referred from outside institutions with little detail about care at these institutions.	Unclear
Taura, 2005[42] Asia: Japan 1991 - 2001	Unclear - non-screening group presented with symptoms, but unclear how this was determined in retrospective review	No	Loss to follow-up - unclear Median follow-up months: 41.3 vs 29.6	Unclear - does not specify whether these were consecutive patients with HCC. All were from single institution, but unclear why some patients received routine screening while others did not.	Unclear
Tong, 2010[43] U.S. Pasadena, CA 1991-2008	Unclear: Non-screening group was referred from elsewhere. No info on screening among non-screening group.	NR	NR	Unclear - non-screening patients presented to clinic with HCC. No information about their prior care.	Source of death data NR
Trevisani, 2002[44] Europe: Italy 1988-1998	Unclear how symptomatic presentation was defined.	No	5 vs 4 vs 9	No - Most patients treated at study center were part of screening program, while referral patients were not.	unclear
2004[45] Europe: Italy 1988-2001	Unclear - no details about how symptomatic or incidental HCC diagnoses were categorized in the registry.	No	0 vs 2 vs 2	No - Most patients treated at study center were part of screening program, while referral patients were not. In fact, treating center was independently associated with survival.	unclear
Wong, 2008[46] Asia: China (Hong Kong) 2003-2005	Screening group pts enrolled in a screening program. Non-screening group was referred - possible that some of these patients were screened, but no data. "We assumed that these patients did not receive regular follow-up or screening with AFP or USG while the HCC was an incidental finding."	No	NR Data for screening group only: median duration of follow-up 184 weeks (range 61–363 weeks).	No - non-screening group defined as being all referral patients	Unclear
Yu, 2004[47] Asia: Taiwan 1996-1997	No details reported. Screening group: tumors were found during routine follow-up US, no details on frequency. The nonscreening group consisted of the opportunistic and symptomatic groups. Opportunistic group: tumors were found by incidental health checkup or other nonhepatic reasons without liver-associated symptoms Symptomatic group - visited hospital because of liver-associated symptoms.	No	NR	Unclear - not enough information about how groups were defined	Yes - linked to Taiwan mortality data

Table 5. Randomized controlled trials comparing TACE to supportive care in patients with hepatocellular carcinoma

Study Country Setting Years of Enrollment	N subjects: T vs C	Inclusion and exclusion criteria	Sample characteristics; liver disease etiology (% T vs C)	Liver disease stage (% T vs C)	Survival (%T vs C)	Adverse events
Groupe d'Etude et de Traitment du Carcinome Hepatocellulaire, 1995[53] France, Belgium, Canada 24 centers 1990-1992	50 vs 46	HCC with AFP >250ng/ml, excluded patients who were candidates for surgery, previous treatment, severe liver disease, vascular contraindications to chemoembo, increased creatinine, extrahepatic mets.	Mean age 63 vs 65 Male 96 vs 96 EtOH 76 vs 73 HBV 4 vs 7 HCV 9 vs 10 Primary hemochromatosis 11 vs 10	Okuda I 94 vs 84.8 Okuda II 6 vs 15.2	Unadjusted RR of death: 0.7, (95% CI 0.45-1.11, P=0.13) 1-year: 62 vs 43.5 2-year: 37.8 vs 26 Adjusted RR of death: 0.77 (95% CI 0.48-1.25, P=0.31) adjusted for Karnofsky score, ascites, bilirubin, albumin, tumor type, tumor mass, portal obstruction AFP, chemoembolization	Trial stopped due to deaths in both groups (liver failure, GI hemorrhage, SBP). Chemoembolization led to <50% increase in survival after 8 months, therefore trial stopped. Abdominal pain 80% Vomiting 80% Fever 76% Death 2% Ascties 10% Encephalopathy 2% GI hemorrhage 8% Cholecystitis 4% Elevated AST/ALT ≥ 5x ULN 3 days after treatment 54% Increase in serum bilirubin ≥ 0.9mg/dL 58% Other complications 18%
Lo, 2002[55] Hong Kong Single-center 1996-1997	40 vs 39	Patients with unresectable HCC. Excluded: poor hepatic function, elevated creatinine, history of prior tumor treatment of acute tumor rupture, presence of extrahepatic metastasis or vascular contrainidcations to chemoembolization, poor performance status	Mean age 62 vs 63 Male 90 vs 87 HBsAg pos 85 vs 74	Okuda I 47.5 vs 46.1, Okuda II 52.5 vs 53.9	Unadjusted: 1-year : 57 vs 32 2-year: 31 vs 11 3-year: 26 vs 3 RR of death 0.50 (95%CI 0.31-0.81, p=0.005) Adjusted RR of death: 0.49 (95% CI 0.29-0.81, p=0.006), adjusted for symptoms, portal vein obstruction, Tumor size, Okuda, treatment with TACE	38 patients had treatment stopped because of progressive disease (12 patients), death (7 patients), poor liver function (6 patients), adverse effects (6 patients), patient refusal (3 patients), arteriovenous shunting (2 patients), and hepatic artery thrombosis (2 patients). The most common clinical adverse effect was a self-limiting syndrome consisting of fever, abdominal pain, and vomiting.
Pelletier, 1990[54] France 10 hospitals 1985-1988	21 vs 21	Consecutive patients with HCC were included. Excluded: resectable HCC, patients with spontaneous encephalopathy with associated poor survival rates, non-embolizable HCC due to portal vein thrombosis, or previous porto-caval anastamosis.	Age 64 vs 66 Male 91 vs 86 EtOH 71.4 vs 66.7 Non-EtOH 28.6 vs 33.3	Okuda I 28.6 vs 23.8 Okuda II 53.4 vs 52.4 Okuda III 19 vs 23.8	Unadjusted: 6 month 33 vs 52 1 year 24 vs 31 (no statistical difference)	Two severe complications of chemoemoblization: death from acute renal failure in one patient, and a gastrointestinal hemorrhage from acute gastroduodenal ulcerations

Table 6. Assessment of study methods for potential sources of bias in randomized trials of TACE in patients with early-stage hepatocellular carcinoma

Author, Year, Geographic setting, Years of enrollment	Sequence generation	Allocation concealment	Blinding of participants, personnel and outcome assessor	Incomplete outcome data	Selective outcome reporting	Other sources of bias	Summary assessment: Risk of Bias
Groupe d'Etude et de Traitment du Carcinome Hepatocellulaire, 1995[53] France, Belgium, Canada	Yes, randomized	Yes, centralized telephone center	Can't answer, no mention of blinding	Yes	Yes	Yes	Low
Lo, 2002[55] China	Unclear	No – sealed, but not necessarily opaque envelopes, not centralized	Can't answer, blinding not discussed	Yes, ITT. Censored patients lost to F/U	Yes	Yes	Unclear
Pelletier, 1990[54] France	Unclear	Probably yes, randomization and assignment done centrally at one center	Can't answer, no mention of blinding of radiologists and others that could have been blinded	can't answer, No mention of loss to follow-up	Yes	Yes	Unclear

Table 7. Cohort studies comparing resection, RFA, TACE, and OLT to supportive care in patients with hepatocellular carcinoma

Study Country Study design Years of enrollment	N subjects T vs C	Active Treatment modalities	Cohort definition	Sample characteristics; liver disease etiology (% T vs C)	Liver disease stage, Child-Pugh, or MELD (% T vs C)	Stage at Diagnosis (% T vs C)	Survival (% T vs C)	Adverse events
DuBay, 2011[64] Canada Retrospec-tive cohort 1999-2007	77 vs 93	RFA	All patients with diagnosis of HCC within Milan criteria on the liver transplant waiting list or listed patients who developed HCC while waiting liver transplant at a single transplant center in Toronto. Patients were stratified into RFA (n = 77) and No Treatment groups (n = 93).	Age 56 vs 55 Male 86 vs 81 Female 11 vs 18 HBV: 22 vs 19 HCV: 64 vs 56 EtOH: 12 vs 26 NASH/Cryptogenic 4 vs 4 Other: 1 vs 5	MELD (14 vs 15)	Mean number 1.33 vs 1.35 Max size 2.5 vs 2.4	Among non-transplanted patients (waiting list drop-off events) Unadjusted: 1-year: 87 vs 71 3-year: 76 vs 39 5-year: 55 vs 30 (P=0.009) Adjusted RR not reported	n=77, No major events, 2 minor events (L portal vein thrombosis, vasovagal reaction)
Farinati, 2012[80] Italy 1987-2006	25 27 22 68 41	OLT Resection RFA TACE Supportive/ other medical therapy	Consecutive patients with HCC at 10 institutions forming the ITA.LI.CA (Italian Liver Cancer) group, of whom 228 were eligible for OLT.	OLT eligible (n=228): Male 77.6	Child-Pugh class: A 52.2 B 47.8	Single lesion: 62.2 Up to 3 nodules: 37.7	Among pts eligible for OLT, median survival in months: OLT (mean) 143.7 Resection 56 RFA 44 TACE 34 Supportive 23 (p=0.001) Adjusted HR not for each modality not reported.	No
Lee, 2012[62] Korea Retrospective cohort 2000-2003	86 vs 22 (TNM I, II) Overall n=257	Resection; RFA; TACE; systemic chemotherapy	All patients diagnosed with HCC at a single center in Korea. Excluded patients with inadequate data, prior initial treatments for HCC at other hospitals, or interruption to follow up. The survival of the patients was analyzed on the basis of the initial treatment adopted in patients with Child-Pugh class A or B. For initial treatment, 17 patients (6.6%) underwent surgical resection, 19 (7.4%) underwent RFA, 135 (52.5%) underwent TACE, 2 (0.8%) received systemic chemotherapy, and 84 (32.7%) received supportive care.	Age <50 16, ≥ 50 84 Male 77, Female 23 Serum AFP levels > 400 ng/mL 41.2 HBV 66 HCV 5 HBV/HCV 1 EtOH 19 Unknown 9	Childs A 41 Childs B 40 Childs C 19	TNM I 7 TNM II 37 TNM III 31 TNM IV-a 16 TNM IV-b 9	Unadjusted survival in patients with TNM I & II disease: Resection vs RFA vs TACE vs Conservative: 1-year: 100 vs 81.8 vs 73 vs 25 3-year: 91.7 vs 36.4 vs 33 vs 8.3 5-year: 75 vs 27.3 vs 19 vs 8.3 (P<0.01) Adjusted RR not reported	No

Study Country Study design Years of enrollment	N subjects T vs C	Active Treatment modalities	Cohort definition	Sample characteristics; liver disease etiology (% T vs C)	Liver disease stage, Child-Pugh, or MELD (% T vs C)	Stage at Diagnosis (% T vs C)	Survival (% T vs C)	Adverse events
Liu, 2004[57] USA Retrospective cohort 1988-1998	229 vs 188	Resection	All histologically confirmed HCC, patients considered resection candidates with a ≤5cm solitary lesion confined to a single lobe of the liver and no medical contraindications to surgery (e.g. cirrhosis), based on SEER data. Excluded patients with contraindications to surgery, and patients who received local therapy (e.g., cryoablation) or underwent transplantation.	Age 60.9 vs 66.8 White 51.5 vs 47.9 Black 4.8 vs 13.8 Asian 30.1 vs 28.2 Hispanic 13.5 vs 10.1 Not reported	Not reported, no cirrhosis in this cohort	Mean tumor size 3.0 vs 3.7	Unadjusted: 1-year: 72.7 vs 40.9 5-year: 32.5 vs 7.3 Median survival 47.1 vs 17.9 month, p<0.001 Adjusted HR 0.45 (95% CI 0.34-0.59, p<0.01), adjusted for resection, age, size, gender, race	No
Mahady, 2010[81] Australia Prospective cohort 1998-2007	128 vs 132	Locoregion-al therapy (RFA, TACE, PEI)	All patients diagnosed with HCC at a single center. Patients were divided into those who received locoregional therapies and those who received supportive care.	M/F 81/19 vs 74/26 Age (mean) 60 vs 58 Caucasian 59 vs 76 Asian 32 vs 17 Other 9 vs 7 HCV 50 vs 46 HBV 25 vs 22 Combined 2 vs 0 EtOH 14 vs 20 Other 9 vs 7	Childs A 57.0 vs 23.5 Childs B 30.5 vs 24.2 Childs C 10.1 vs 17.4 non-cirrhotic 3 vs 3 Ascites 29 vs 53 Tumor symptoms 23 vs 38	CLIP 0: 16 vs 5 CLIP 1-2: 73 vs 51 CLIP 3-6: 10 vs 34 Tumor extending >50% of liver 5 vs 17 Portal vein thrombosis 5 vs 29	Unadjusted HR for death 0.48 (95%CI 0.35-0.65, p=0.001 Adjusted HR for death 0.59, 95% CI 0.41-0.83, p=0.03, adjusted for CLIP score, AFP, Alk Phos, Bilirubin	No
Tong, 2010[58] USA Retrospective cohort 2000-2007	236 vs 42	OLT; OLT + other; resection; resection + other; RFA only; TACE only; RFA + TACE; Chemother-apy; Supportive care	Asian American patients with HCC who were referred to a single tertiary Liver Cancer Center during a 7-year period	Mean age 61.5 (SD 11.7) Males 78.1 Ethnicity: Chinese 52.5 Korean 17.3 Vietnamese 14.0 Japanese 13.3 Other 2.9 Hepatitis B 57.9 Hepatitis C 33.1 HBV and HCV 1.4 Hemochromatosis 1.1 Alcoholic liver disease 0.7 Nonalcoholic steatohepatits 0.4 Von Gierke Disease 0.4 Unknown etiology 4.7	Child Turcotte Pugh A 70.3 B 19.1 C 2.9 Mean MELD score 15.6 (SD 7.8)	Within Milan criteria 56.8 Macrovascular invasion 11.2 Metastasis (11 lung, 3 bone) 5	Unadjusted 1/3/5 year survival: OLT 65/53/53 OLT and TACE or RFA 96/58/50 Resection 66/59/- RFA only 87/63/49 TACE only 49/19/- RFA and TACE 96/48/21 Chemotherapy 17/-/- Supportive 12/12/- Adjusted RR not reported	No

* Stage I: tumor size <50%, no ascites, albumin >3 g/dL, and bilirubin <3 mg/dL; Stage II: moderately advanced (one or 2 of the signs of advanced disease are present; Stage III: very advanced.

Table 8. Assessment of study methods for potential sources of bias in cohort studies of resection, OLT, RFA, and TACE in patients with hepatocellular carcinoma

Author, Year, Geographic setting, Years of enrollment	Comparability of groups? Confounding by indication? Selection of the non-exposed cohort drawn from the same community as the exposed cohort?	Outcome assessment bias? Difference in loss to follow-up between treated and controls?	Adequate adjustment for potential confounders?
DuBay, 2011[64] Canada 1999-2007	Confounding by indication an issue as no details were given as to why certain OLT candidates would receive RFA and which would not, groups are generally comparable	No bias in outcome assessment. No difference in loss to follow-up	Unclear which variables were modeled in the multivariable analysis of overall survival with RFA versus control
Farinati, 2012[80]	Selected all patients who would be potentially eligible for OLT on the basis of age, tumor stage, and liver disease severity but did not account for other factors that would determine surgical candidacy so confounding by indication likely present.	No discussion of loss to follow-up	Yes
Lee, 2012[62] Korea 2000-2003	Confounding by indication present, unable to assess the characteristics of treatment group as compared to the control group, other than by stage	No description of loss to follow-up	Unclear multivariable analysis for survival
Liu, 2004[57] USA 1988-1998	Chose all patients who would be surgical candidates, groups were comparable. Registry data on surgical contraindications originated from chart review, but unable to account for patient comorbidities that may have influenced decision to perform surgery.	No discussion of loss to follow-up	No liver disease variable, but did not include cirrhotic
Mahady, 2010[81] Australia	Baseline groups were not similar, confounding by indication present	No	Yes
Tong, 2010[58] USA 2000-2007	unknown, groups drawn from the same community, but confounding by indication present	Loss to follow-up not discussed	Yes adjusted for confounders

Table 9. Non-comparative observational studies of OLT, RFA, and TACE in patients with hepatocellular carcinoma

Study Setting Time period of enrollment	N Liver disease etiology %	Treatment modality	Long-term survival %	Harms and other findings
Burra, 2013[60] Europe, ELTR database 1988-2010	5626 HBV 26 HCV 71 HBV/HCV 3	OLT	5-year: 61-72 10-year: 45-66	NR
Ioannou, 2008[61] US, UNOS database 1997-2007	5776 HBV 8 HCV 61 EtOH 9	OLT	4-year: 67-73	NR
Onaca, 2009[59] International, ITR registry 1983-2005	902 NR	OLT	5-year: 56	NR
Chen 2011[66]	121 HBV 45.5	RFA	• 1-year: 92.5 2-year: 78.5 3-year: 67.2	9.9% of patients experienced major complications, namely hemothorax, pneumoperitoneum, persistent intrahepatic biliary dilatation, branch portal vein thrombosis, and peritoneal seeding.
Livraghi, 2008[63] Italy 1995-2006	218 HCV 83.9 HBV 7.3 HCV-HBV 4.1 Alcohol 3.2 Unknown 1.3	RFA	3-year: 76 6-year: 55	1.8% experienced major complications regarded as treatment-related: peritoneal bleeding, hemothorax, neoplastic seeding, hyperbilirubinemia lasting for 1 month.
Eltawil, 2012[82] Canada 2005-2010	48 HCV 35.4 HBV 4.2 NASH 8.3 Alcohol 33.3 other 10.4	TACE	1-year: 72 3-year: 28 4-year: 12	Post-embolization syndrome: 40-50% Hepatic abscess: 2% Transient decompensation with ascites: 8% Quality of life was measured (WHOQOL-BREF) at baseline and after the treatment period. No statistically significant temporal trends were detected for any of the 4 health domain QOL measures (physical, psychological, social relationships, and environmental well-being).

Table 10. Summary of AASLD, APASL, and EASL-EORTC guidelines for screening for hepatocellular carcinoma

Organization	Population for whom screening is recommended	Screening modality	Timeframe for screening	Levels of evidence used in guidelines	Strength of recommendation levels used in guidelines
AASLD	Patients with HBV; Patients with cirrhosis (evidence level I; recommendation NR) Patients awaiting transplant (evidence level III; recommendation NR)	Ultrasound (evidence level II; recommendation NR)	6 month intervals (evidence level II; recommendation NR) The screening interval does not need to be shortened for patients at higher risk of HCC (evidence level III; recommendation NR)	Levels of evidence were assigned according to study design: I Randomized controlled trials II-1 Controlled trials without randomization II-2 Cohort or case control analytic studies II-3 Multiple time series; dramatic uncontrolled experiments III Opinion of respected experts; descriptive epidemiology	NR
APASL	Patients with HBV or HCV and cirrhosis (evidence 2a, recommendation B)	Ultrasound and a-fetoprotein (evidence 2a; recommendation B)	6 month intervals (evidence 2a; recommendation B)	The quality of existing evidence was ranked 1 (highest) to 5 (lowest) according to the Oxford system of evidence-based approach for developing the consensus statements.	The strength of recommendations ranked from A (strongest) to D (weakest) according to the Oxford system of evidence-based approach for developing the consensus statements.
EASL-EORTC	Patients with HBV and active hepatitis or family history of HCC (evidence 1B; recommendation A1 for Asian patients; evidence 3D; recommendation C1 for Western patients); Patients with chronic hepatitis C and advanced fibrosis (evidence 3D; recommendation B1 for Asian patients; evidence 3D; recommendation B2 for Western patients); Patients with cirrhosis (evidence 3A; recommendation B1); Patients awaiting transplant (evidence 3D; recommendation 1B)	Ultrasound performed by experienced personnel (evidence 2D; recommendation 1B)	6 month intervals (evidence 2D; recommendation 1B)	(adapted from National Cancer Institute*) Level 1: Randomized controlled clinical trials or meta-analyses of randomized studies* (i) Double-blinded (ii) Non-blinded treatment delivery Level 2: Non-randomized controlled clinical trials Level 3: Case series (i) Population-based, consecutive series (ii) Consecutive cases (not population-based) (iii) Non-consecutive cases Strength of evidence according to end-points: A. Total mortality (or overall survival from a defined time) B. Cause-specific mortality (or cause-specific mortality from a defined time) C. Carefully assessed quality of life D. Indirect surrogates (i) Event-free survival (ii) Disease-free survival (iii) Progression-free survival (iv) Tumor response rate	(adapted from the GRADE system) Grading of evidence A -High quality: Further research is very unlikely to change our confidence in the estimate of effect B -Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate C- Low or very low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate. Any estimate of effect is uncertain. Grading recommendation 1-Strong recommendation warranted: Factors influencing the strength of the recommendation included the quality of the evidence, presumed patient-important outcomes, and cost 2-Weaker recommendation: Variability in preferences and values, or more uncertainty: more likely a weak recommendation is warranted. Recommendation is made with less certainty: higher cost or resource consumption

*National Cancer Institute. PDQ_ levels of evidence for adult and pediatric cancer treatment studies. Bethesda, MD: National Cancer Institute. http://www.cancer.gov/cancertopics/pdq/levels-evidence-adult-treatment/healthprofessional/

Abbreviations: AASLD = American Association for the Study of Liver Diseases; APASL = Asian Pacific Association for the Study of the Liver; EASL-EORTC = European Association for the Study of the Liver/European Organisation for Research and Treatment of Cancer; NR = not reported

Table 11. AGREE II quality assessment of guidelines for screening for hepatocellular carcinoma

AGREE II Quality Assessment Item		AASLD	APASL	EASL-EORTC
Overall: Rate the overall quality of this guideline		**2**	**3**	**3**
		1 (strongly disagree) – 7 (strongly agree)		
Domain 1: Scope and Purpose				
1.	The overall objective(s) of the guideline is (are) specifically described.	6	6	7
2.	The health question(s) covered by the guideline is (are) specifically described.	6	6	7
3.	The health question(s) covered by the guideline is (are) specifically described.*	6	6	7
Domain 2: Stakeholder Involvement				
4.	The guideline development group includes individuals from all relevant professional groups.*	2	5	3
5.	The views and preferences of the target population (patients, public, etc.) have been sought.*	1	1	1
6.	The target users of the guideline are clearly defined.*	4	4	6
Domain 3: Rigour of Development				
7.	Systematic methods were used to search for evidence.*	2	3	4
8.	The criteria for selecting the evidence are clearly described.	1	1	1
9.	The strengths and limitations of the body of evidence are clearly described.	1	2	3
10.	The methods for formulating the recommendations are clearly described.*	2	3	3
11.	The health benefits, side effects, and risks have been considered in formulating the recommendations.*	2	2	2
12.	There is an explicit link between the recommendations and the supporting evidence.*	2	3	4
13.	The guideline has been externally reviewed by experts prior to its publication.*	2	5	1
14.	A procedure for updating the guideline is provided.*	1	1	1
Domain 4: Clarity of Presentation				
15.	The recommendations are specific and unambiguous.	5	5	5
16.	The different options for management of the condition or health issue are clearly presented.*	4	4	5
17.	Key recommendations are easily identifiable.*	4	5	6
Domain 5: Applicability				
18.	The guideline describes facilitators and barriers to its application.	2	2	2
19.	The guideline provides advice and/or tools on how the recommendations can be put into practice.	2	2	2
20.	The potential resource implications of applying the recommendations have been considered.	2	2	2
21.	The guideline presents monitoring and/or auditing criteria.	1	1	1
Domain 6: Editorial Independence				
22.	The views of the funding body have not influenced the content of the guideline.*	1	1	1
23.	Competing interests of guideline development group members have been recorded and addressed.*	3	1	1

Abbreviations = AASLD: American Association for the Study of Liver Diseases; APASL: Asian Pacific Association for the Study of the Liver; EASL-EORTC: European Association for the Study of Liver, European Organisation for Research and Treatment of Cancer

APPENDIX E. PEER REVIEW COMMENTS AND RESPONSES

	Reviewer	Comment	Response
	Question 1: Are the objectives, scope, and methods for this review clearly described?		
1.	1	Yes. Very impressive.	
2.	2	Yes. The methods section pertaining to inclusion of systematic reviews of comparative effectiveness studies is a little unclear—at first it says these studies were excluded, but then it says you included reviews of comparative effectiveness studies (if I'm understanding correctly). I didn't really see anything in the results comparing one active treatment to another so it seems to me that it would be clearer to just say that studies that compared active treatments were excluded—unless there was some specific outcome or reason that you included them, and then just say "we included systematic reviews of studies that compared active treatments and reported xx outcomes" or something like that—which I think would be more straightforward.	We agree. We initially had looked at these studies to gather more information about treatment-specific harms, but we agree that the majority of harms data we report are from cohort studies. We've clarified that studies that compared active treatments were excluded.
3.	3	Yes (no comment)	
4.	4	Yes (no comment)	
5.	5	Yes. Effects of HCC surveillance on mortality in pts with chronic liver disease	
6.	6	Yes (no comment)	
7.	8	Yes. Although the objectives, scope, and methods for this review are clearly described, they do not specifically address issues related to the situation in the Veterans Health Administration. In an e-mail dated 19 February 2013 to the ESP, I had indicated that this was an issue of major interest to my office.	The scope of the review was broad enough to have captured studies both directly and indirectly relevant to VA. We did include a section (in Results) specifically focused on VA studies. In our edits, we added more detail to our discussion of the 3 VA observational studies and more about current screening practices in VA. We agree that the background should have included VA-specific information – we've added some information about change in prevalence of HCV and HCC diagnoses in VA over time.
8.	9	Yes	
	2. Is there any indication of bias in our synthesis of the evidence?		
9.	1	No (no comment)	
10.	2	No (no comment)	
11.	3	No (no comment)	
12.	4	No (no comment)	
13.	5	No. Just the exclusion of articles not in English	
14.	6	No (no comment)	
15.	8	Yes. The failure to consider the natural history of HCC introduces a serious methodologic flaw that biases the results towards the finding that there is not evidence to support treatment of early HCC . The implicit assumption that HCC is similar in its biologic behavior to malignancies such as prostate cancer – where patients may survive for decades without treatment – is incorrect (please see comments 3 and 4 *infra*). If applied to non-small cell lung cancer, the approach used here would likely conclude that there is insufficient evidence to support resection of stage 0 or stage 1A NSCLC.	We agree that the natural history of HCC is important contextual information. In the initial draft, we did consider natural history in the discussion, albeit briefly. In our edits, we have expanded our discussion of natural history and the implications of natural history on the overall balance of risk/benefit for HCC screening. Much of the information about the natural history of early-stage HCC is several decades old. What information there is suggests that the natural history of early-stage HCC varies.
16.	8	In addition, the failure to include trials that compare 2 different modalities systematically excluded evidence that supports treatment of early HCC.	Throughout the report, we have clarified that we evaluated treatment studies only as a way to indirectly evaluate the effects of screening. Because this review was focused on screening, we looked for evidence of the effects of treating screen-detected (or early-stage HCC) compared to no treatment. Studies comparing 2 or more active treatments would not have provided evidence about the effects of treating screen-detected disease.

	Reviewer	Comment	Response
17.	8	In addition, the discussion of harms from evaluation of liver lesions detected via surveillance failed to note that the diagnosis of HCC is most often made by imaging rather than biopsy. This consideration was communicated in my e-mail of 19 February 2013, but not addressed in this document.	The initial draft's discussion section did include a statement about diagnoses most often occurring as a result of imaging rather than biopsy. Nevertheless, we have re-written the screening harms section to clarify that, in considering harms, we considered the harms of the initial screening tests themselves as well as the harms of additional testing done in response to initial positive screening test results. We agree that liver biopsy is not commonly performed for the diagnosis of HCC, but it is performed sometimes and the harms are important to consider. In our re-drafted section, we clarify that few studies reported testing actually performed. The 2 studies that reported rates of liver biopsy used in HCC diagnosis reported 33 and 80% of patients had liver biopsy performed. We also clarified in the summary of evidence table that there was limited directness of information about harms (since most had to do with needle track seeding and biopsy is not often used in current practice).
18.	9	Yes. The search methodology is well explained and most of the relevant direct evidence has been captured. There are however several problematic issues. Based on the review, there is a large body of observational studies that consistently show an association between HCC surveillance and HCC diagnosis at an early stage, increase receipt of resection or transplant, and increased survival. This is understandably not grade 1 evidence but it is grade 2. Expressing the data as "there is no evidence" is inaccurate. Rather, there is evidence of grade 2 level that consistently indicate xxx etc.	We disagree that we characterized the data as "there is no evidence". We graded the evidence as "insufficient" because it is impossible to know whether the improved survival observed in these studies among screen-detected patients was related to a true effect of screening or, rather, to lead-, length-, and selection biases common to the studies. There are also important inconsistencies among studies – 3 of 4 studies that attempted to correct for lead-time found survival advantages attenuated with longer tumor doubling time assumptions. Nevertheless, we agree that readers could misinterpret the term "insufficient" as "there is no evidence". Therefore, we elected to use GRADE terminology to summarize the strength of evidence rather than the AHRQ terminology we originally used. Using GRADE, we believe the strength of evidence would be "very low". We have used this term in place of insufficient in our edited draft and hope this will better communicate that there are indeed studies, but that any conclusions drawn from this body of evidence are very tenuous.
19.	9	There seems be "kitchen sink" approach of piling disadvantages of the two RCT, but not enough follow through as to the possible consequences of the disadvantages. A couple of limitation like low screening rates, and inclusion of non cirrhotics would bias the results toward the null (not the opposite).	We agree that this section could have been more thoughtfully presented. We've re-written to include a more precise discussion of how various deficiencies might affect the results.
20.	9	For TACE, they seem to ignore the meta analyses of RCT (and observational studies) which demonstrate statistically significant benefit in survival. For sorafenib, there is no mention! It is the only RCT proven efficacious intervention for palliative therapy of HCC.	These studies were not included because we were focused on screening efficacy and, therefore, focused only on studies evaluating the effects of treating early-stage HCC (since the effects and rationale for screening is most likely to increase the detection of earlier stage disease). The TACE meta-analysis included the 3 trials we included plus a fourth (the Llovet Lancet trial) that we excluded because it included only patients with advanced stage disease. Likewise, sorafenib has only been evaluated inpatients with advanced stage disease. We have clarified throughout the report that we were focused on this specific subgroup of treatment studies. We agree that, though this language was in the initial draft, readers might have missed that we were not evaluating HCC treatment in general. We also include an additional statement in the limitations section that this review applies only to early-stage disease and not advanced stage, and we acknowledge that some patients undergoing screening could have newly discovered advanced stage disease.
21.	9	The search did not get into the many studies that report survival of patients who were listed but not transplanted compared to those listed and transplanted (at the same stage).	We included comparative and non-comparative observational studies of OLT if they included patients with early-stage disease.
22.	9	In the summary, there is no numerical emphasis on worst case scenario against HCC surveillance (for example mentioning the point estimate of the meta analysis for tumor seeding but not the 95% CI or the more recent study described in the body of the document). Long term survival (which is quiet good) with resection and transplant is presented without numbers.	We have included the long-term survival numbers, the CI information, and we also note in the table the most recent needle-track seeding study.

	Reviewer	Comment	Response
23.	9	Certainly the summary of the document is not accurate or supported by the data: consistent evidence of grade 2 is not the same as insufficient evidence to draw conclusions.	See response to comment #18.
24.	9	Feasibility and ethics of RCT need also to be addressed in light of existing studies that indicate that patients are unlikely to accept such a trial when (nothing is one of the arms).	We have included this as an important area for future research.
25.	9	Some references 1: Poustchi H, Farrell GC, Strasser SI, Lee AU, McCaughan GW, George J. Feasibility of conducting a randomized control trial for liver cancer screening: is a randomized controlled trial for liver cancer screening feasible or still needed? Hepatology. 2011 Dec;54(6):1998-2004. doi: 10.1002/hep.24581. PubMed PMID: 21800340. 2: Llovet JM, Bustamante J, Castells A, Vilana R, Ayuso Mdel C, Sala M, Brú C, Rodés J, Bruix J. Natural history of untreated nonsurgical hepatocellular carcinoma: rationale for the design and evaluation of therapeutic trials. Hepatology. 1999 Jan;29(1):62-7. PubMed PMID: 9862851. 3: Davila JA, Duan Z, McGlynn KA, El-Serag HB. Utilization and outcomes of palliative therapy for hepatocellular carcinoma: a population-based study in the United States. J Clin Gastroenterol. 2012 Jan;46(1):71-7. doi: 10.1097/MCG.0b013e318224d669. PubMed PMID: 22157221; PubMed Central PMCID:PMC3832893. 4: Altekruse SF, McGlynn KA, Reichman ME. Hepatocellular carcinoma incidence, mortality, and survival trends in the United States from 1975 to 2005. J Clin Oncol. 2009 Mar 20;27(9):1485-91. doi: 10.1200/JCO.2008.20.7753. Epub 2009 Feb 17. PubMed PMID: 19224838; PubMed Central PMCID: PMC2668555. 5: Schwarz RE, Smith DD. Trends in local therapy for hepatocellular carcinoma and survival outcomes in the US population. Am J Surg. 2008 Jun;195(6):829-36. doi: 10.1016/j.amjsurg.2007.10.010. Epub 2008 Apr 23. Review. PubMed PMID: 18436176. 6: El-Serag HB, Siegel AB, Davila JA, Shaib YH, Cayton-Woody M, McBride R, McGlynn KA. Treatment and outcomes of treating of hepatocellular carcinoma among Medicare recipients in the United States: a population-based study. J Hepatol. 2006 Jan;44(1):158-66. Epub 2005 Nov 2. PubMed PMID: 16290309. 7: El-Serag HB, Mason AC, Key C. Trends in survival of patients with hepatocellular carcinoma between 1977 and 1996 in the United States. Hepatology. 2001 Jan;33(1):62-5. PubMed PMID: 11124821. 8: Davila JA, El-Serag HB. Racial differences in survival of hepatocellular carcinoma in the United States: a population-based study. Clin Gastroenterol Hepatol. 2006 Jan;4(1):104-10; quiz 4-5. PubMed PMID: 16431312. 9: Leung JY, Zhu AX, Gordon FD, Pratt DS, Mithoefer A, Garrigan K, Terella A, Hertl M, Cosimi AB, Chung RT. Liver transplantation outcomes for early-stage hepatocellular carcinoma: results of a multicenter study. Liver Transpl. 2004 Nov;10(11):1343-54. PubMed PMID: 15497158.	Most of these provide background or contextual information. We have reviewed all these studies and included them in background or discussion as appropriate. We have re-written the background to include more information on changing HCC incidence (including the Altekreuse reference). We did examine the Leung study for inclusion as a comparative observational study. However, it compared HCC patients receiving OLT to patients receiving OLT for non-malignant disease. Because there were several large national OLT cohorts, we only examined noncomparative OLT studies with > 500 patients. The 5-year survival reported in this smaller cohort was quite similar to the survival we report from the larger cohorts.
	3. Are there any <u>published</u> or <u>unpublished</u> studies that we may have overlooked?		
26.	1	No (no comment)	
27.	2	No. Not that I'm aware of.	
28.	3	No. Not that I know of.	
29.	4	No (no comment)	
30.	5	No. None that I am aware of	

	Reviewer	Comment	Response
31.	6	No	
32.	8	Yes. Note: These are only a small portion of the literature that could be cited that contradicts the findings in this document, particularly with regard to treatment of early HCC. 1. Cho YK, Kim JK, Kim MY, Rhim H, Han JK. Systematic review of randomized trials for hepatocellular carcinoma treated with percutaneous ablation therapies. Hepatology. 2009 Feb;49(2):453-9. 2. El-Serag HB, Siegel AB, Davila JA, Shaib YH, Cayton-Woody M, McBride R, McGlynn KA. Treatment and outcomes of treating of hepatocellular carcinoma among Medicare recipients in the United States: a population-based study. J Hepatol. 2006 Jan;44(1):158-66. 3. Llovet JM, Real MI, Montaña X, Planas R, Coll S, Aponte J, Ayuso C, Sala M, Muchart J, Solà R, Rodés J, Bruix J; Barcelona Liver Cancer Group. Arterial embolisation or chemoembolisation versus symptomatic treatment in patients with unresectable hepatocellular carcinoma: a randomised controlled trial. Lancet. 2002 May 18;359(9319):1734-9. 4. McMahon BJ, Bulkow L, Harpster A, Snowball M, Lanier A, Sacco F, Dunaway E, Williams J. Screening for hepatocellular carcinoma in Alaska natives infected with chronic hepatitis B: a 16-year population-based study. Hepatology. 2000 Oct;32(4 Pt 1):842-6. 5. Orlando A, Leandro G, Olivo M, Andriulli A, Cottone M. Radiofrequency thermal ablation vs. percutaneous ethanol injection for small hepatocellular carcinoma in cirrhosis: meta-analysis of randomized controlled trials. Am J Gastroenterol. 2009 Feb;104(2):514-24. 6. Schwarz RE, Smith DD. Trends in local therapy for hepatocellular carcinoma and survival outcomes in the US population. Am J Surg. 2008 Jun;195(6):829-36.	We had reviewed most of these studies. They were excluded for following reasons: Cho – comparative effectiveness studies El-Serag – no outcomes specific to patients with early-stage disease. Llovet – trial included only patients with advanced stage disease. McMahon – we excluded observational screening studies with historic controls (because there were plenty of studies with contemporary controls and the use of historic controls would have introduced one more source of bias). Orlanda – comparative effectiveness Schwarz – no outcomes information specific to patients with early-stage disease.
33.	9	Yes. Population based US data on survival of patients with HCC who were transplanted. There are several publications based on UNOS data that show excellent (70% plus 5 year recurrence free survival) survival among HCC patients who received liver transplants.	We included several large OLT cohorts including a large UNOS cohort. We agree long-term survival was good and we reported this.
34.	9	There are multiple publications from population based US based cohorts (plus the publicly available SEER data) that show stage specific survival of patients with HCC, and again clearly showing remarkable improvement with transplant and resection compared to patients with similar stages who did not get transplanted/resected. These are observational studies but the magnitude of differences is dramatic.	In initial draft, we did not include the comparative numbers in the summary of evidence (only in the text) – we've now added this to the table. We reported that survival in patients selected for surgery is quite good and markedly better than patients not selected for surgery. We have added more to the discussion about the potential explanations for such findings – unclear whether this is a true treatment effect or whether this difference reflects careful patient selection and confounding by indication.
35.	9	There are also publications of US based population based cohorts that show unfirmly grim prognosis of untreated patients with HCC (irrespective of stage if one looks at 5 year survival). Compare with 5 year survival of transplant/resection.	See above
36.	9	The harms is remarkably deficient and biased. There is mention of harms of biopsies for liver cancer, and harms of different treatments, but nothing about ultrasound of the liver or blood draws (for AFP). This is akin to talking about the harms of colorectal resection or chemotherapy but not colonoscopy or FIT (for CRC screening) or harms of breast biopsy (but not mammography). Both ultrasound and blood draws (which is the method used to obtain AFP) for all kinds of indications have been around for decades and there is considerable safety data on both.	We have re-written the screening harms section to clarify that, in considering harms, we considered the harms of the initial screening tests themselves as well as the harms of additional testing done in response to initial positive screening test results. We acknowledge that ultrasound and blood draws are, themselves, likely quite safe. Nevertheless, it is still important to consider the harms of testing done in response to initial positive screening test.

	Reviewer	Comment	Response
37.	9	There is a lot of discussion on TACE (which is appropriate) but there is a remarkable avoidance of the one palliative therapy that has grade 1 evidence of efficacy from multiple very well done modern RCTs, namely sorafenib. This agent can only be applied (and has only been tested in those with compensated cirrhosis (mostly Child A) and mostly asymptomatic/mildly symptomatic patients. Surveillance even if it does not detect early cancer that is treated with potentially curative therapy, will detect asymptomatic cancer which could be treated with efficacious palliative treatment. The days of removing patients who are not candidates for liver transplant from surveillance consideration are gone.	See response to #20. While it is true that some patients undergoing screening will have asymptomatic advanced stage disease identified, the intent and results of screening are to increase the detection of early-stage disease. Trials evaluating screening intervals found, in both groups, that screening mostly increased the number of early-stage small tumors. While we acknowledge that our review does not cover the detection and treatment of advanced stage disease, we believe a focus on early-stage disease is appropriate for the discussion of screening.
38.	9	There are actually four RCT of TACE/TAE (check recent meta analyses)	See above #20.
	4. Please write additional suggestions or comments below. If applicable, please indicate the page and line numbers from the draft report.		
39.	1	It is only the cost of CT and MR that precluded their inclusion in the AASLD PG. I was on that committee when the issue came up. Everyone I know uses AFP to screen despite the nonsensical "firing of the AFP" by that PG. I have seen MANY advanced HCCs that were missed on u/s and obvious on CT. CT as performed in the US for HCC screening is dreadfully insensitive.	Noted. We searched for any screening studies using any of these potential screening modalities. Most studies included patients who were detected with U/S and/or AFP.
40.	2	a) In a number of places the report refers to radiofrequency ablation but doesn't mention TACE as another "active" treatment. These are not the same thing so both should be mentioned.	We cover both RFA and TACE.
41.	2	b) In the section describing the RCT's of screening vs. no screening there is some discussion of the Chinese staging system which might warrant some revision. The report states that the Chinese system doesn't consider liver disease severity which is not quite true, as jaundice/ascites/cachexia etc are markers of disease severity (and to my understanding such clinical markers are now included in the Barcelona and other staging systems). I think the more accurate critique would be to note that the Chinese studies used a different staging system than currently used in the U.S. and Europe that didn't include factors such as tumor size, number of lesions, and location of lesions, and focused on clinical markers of disease severity (if I'm stating this accurately) and leave it at that.	We have re-written the trial results section to be more precise about the effects of each flaw, including the staging system.
42.	2	c) In the section on screening vs. no screening it would be helpful to at least report some summary of the difference in survival reported in the observational studies. I would focus on adjusted estimates only and report the median difference with the range. Right now there is really no quantitative report of the results so it's hard for readers to know what to make of it. Even if the data are unreliable providing some numbers can give readers some sense of the potential magnitude of effect, whether confounding is likely to explain much or all of the results, inconsistency, etc.	We have re-drafted the screening observational studies table to make it easier to find the survival and HR data. We have also created a new figure displaying comparative survival information. Finally, we have included a summary of this information in the relevant parts of the results section.
43.	2	d) Same for the section on active treatments vs. conservative treatments—for the RCT's I'd suggest reporting the actual results from the two studies that didn't report any significant effect on survival and for the observational studies reporting some information regarding the differences between treatment and no treatment.	This information is included in the treatment studies table.
44.	3	No comments – well done review.	
45.	4	The review addresses internal validity (risk of bias) of the trials quite well. I wonder if a statement about external validity (applicability/generalizability) might not also be useful. The reason is that many Hep C providers use the Chinese trial of patients with hep B as the basis for recommending screening/surveillance for HCC in hep C patients. Clearly that trial is not sufficient for recommending screening, even for hep B patients, but there may be an additional issue with extrapolating the results (however interpreted) to hep C patients.	We included such a statement in our revisions.

	Reviewer	Comment	Response
46.	5	Under data extraction, in addition to mode and frequency of surveillance it would be interested to review technical limitations of surveillance. A main limitation of ultrasound surveillance is the poor visualization of the liver parenchyma in patients with more advanced cirrhosis and fatty change. In these patient populations, the detection of early HCC, and particularly smaller HCC lesions, is limited. These limitation would likely have an impact on surveillance efficacy.	This was beyond the scope of our review. However, an AHRQ review is currently being completed that covers just this topic.
47.	6	a) Overall, a great job going through a large number of studies. The draft report is clearly written and well structured to present results at different levels of detail, depending on the reader's specific needs.	
48.	6	b) In oncology, surveillance means looking for recurrence of a prior cancer; thus, the title and report should not use that term but rather screening for HCC. Screening = looking for cancer in those without a prior diagnosis of cancer. This is more than semantics as the continued misuse of the term surveillance tends to separate HCC screening from a wealth of knowledge about cancer screening which seems to be ignored in this context. This sentence seems to highlight the problem: "However, recommendations for HCC surveillance remain controversial in part because of concerns over the quality and paucity of existing evidence, and because there have been concerns raised about overdiagnosis and patient harms in other cancer screening programs". It is too easy for some to ignore this sentence because HCC surveillance is not a cancer screening program. At the very least you should include a definiton of terms and clear statement that HCC "surveillance" is in fact a type of cancer screening and not cancer surveillance as would be done after surgical resection of HCC.	We agree – we have changed the terminology to screening from surveillance.
49.	6	c) In the executive summary, it would be helpful to separate the review of the two RTCs from the non-RTCs. In particular, the latter RTC seems to have flaws that were not considered by the guideline writing groups and may not be well known to the end users of this ESP report who read only the Executive Summary.	We have done so.
50.	6	d) Please consider including overdiagnosis bias as another confounder for non RTCs of HCC screening, and perhaps comment on the magnitude of overdiagnosis in the two RTCs.	We have included more in the discussion about natural history and about uncertainty regarding the potential for overdiagnosis in screen-detected HCC. We could not estimate the magnitude of overdiagnosis in the RCTs, but we do make note of the trials comparing screening intervals in which more early-stage patients were found and treated in the intensive screening group without an improvement in survival.
51.	7	While I think the paper probably reaches the right conclusion, the whole paper seems to be somewhat disjointed and a bit frustrating to read. A reader gets the sense of "diffuse anxiety" about how studies were not well-done. While that may be generally true, I wonder if the authors could be more constructive to the field with some more detailed critique and assessment and suggestion. For example: a. What is the 'best evidence' and how good is it? On p9, you don't really explain what you are looking for, regarding "quality of evidence" or "magnitude of effect." Later you will describe weaknesses of many studies, but the weaknesses are described for several studies at a time rather than individual studies, and a reader was left wondering "is there a baby somewhere in this bathwater; is there ANY study that qualifies at 'best' and is that 'good enough'". My hunch is that the answer is no; for example both the Chinese RCTs have major deficiencies. But if that's the case then maybe be clear about it.	We agree. As reported above, we revised the results section to be clearer about the flaws and their implications. We also reconfigured the observational studies results section and discuss the 3 "best" studies earlier.

	Reviewer	Comment	Response
52.	7.	b. Summaries On p13, right above observational studies, you've just finished a 2-page review of RCTs. But you write no summary of what you think of the 2 pages you have written. There is no indication that you've synthesized or evaluated this entire group of studies. (This is not clearly done in the Discussion either). You simply end with critique of whatever trial happened to be last. In contrast, a reader would like to know "What is your bottom line about this whole class of studies, about what you have just told me? Do any come close to being satisfactory to draw a conclusion? Is the quality of all unsatisfactory? Can we take away any idea about magnitude of any impact, or is quality so bad (or magnitude so low) that we can't?	We agree. We have included a summary paragraph at the beginning of both the screening and treatment sections.
53.	7.	Most of the following sections have the same problem: You write paragraphs about individual studies or groups of studies, but you don't summarize at the end of each section what you have described and what you think it means. For example look at the last full paragraph on p14; there's no summary, just diffuse anxiety (again probably warranted); but the whole process looks like you are just throwing up your hands. Ditto for sections on lead-time bias, harms, treatment.	See above.
54.	7.	c. Other organization On p13, in observational studies, you need to remind the reader "what question were you looking at" - benefit, harm, other?	We have revised the section headings accordingly
55.	7.	d. Current guidelines The current guidelines section has the potential to be interesting, to the extent that other guidelines recommend FOR surveillance. When they do, can you: 1) describe what is their recommendation (it's in the Table); 2) what it that based on (what evidence; what studies, what statements about benefit vs risk). Some of this may be buried in Table 12, but can you distill the essence and explain what you think the problem is? 3) how you judge (2). Right now you seem to rely on saying the guidelines aren't any good because (p19, first para) you "identified several methodological flaws…". Can you elaborate more on the details in Appendix D, Table 12: How strong or weak is an overall quality grade of 2 or 3? How serious are the generally low grades in rigor of development? Enough to be disqualifying? Can you say any detail about what specific studies about the evidence each guideline seems to rely on - which study, what magnitude of benefit and of harm? Do they rely on the Chinese RCTs, now largely-discredited? Or other? Right now the whole thing seems somewhat an abstract exercise; can you give it a little more detail and life, in interpreting/summarizing the data in the Tables and in your text summary?	We have revised this section accordingly.
56.	7.	e. Future studies On p22, 2nd para in Future Studies, you could provide much more help to the field by being more specific about what you think should be done in the future, based on what you have learned from your reviews. The recommendations for "consecutive patients" and "prospectively collected …. about adverse effects" seems somewhat generic. In earlier text you've discussed lead-time bias, trying to adjust for it, and other sources of problems. Can you, then, in this section, try to tell future-researchers how to improve what they are doing.	We agree. We now included a future studies table that pairs study suggestions with the evidence gap they are meant to address.

	Reviewer	Comment	Response
57.	8	Background 1. Page 1 - The statement that the 5 year survival for HCC is 16.5% is incorrect. The cited statistics from references 1-2 refer to the SEER category of "liver and intraductal bile duct cancers," not HCC. In addition, reference 1 specifically notes incomplete reporting of VA cancer registry data to state cancer registries, a factor that makes relevance of the cited data to VA unclear. A population-based study of 2,963 HCC patients based on the linked SEER-Medicare dataset found a median survival of 104 days with a 3 year survival rate of 5.7%, not the 16.5% rate cited in this review (El-Serag *et al.* 2006).	There are issues with all long-term survival estimates we found. The SEER data is the most current though we agree it is flawed in that there are some non-HCC cases included. However, HCC are likely to be the majority. We added an additional reference and present the mortality as a range. We believe this shows there is a range, while underscoring the original point which was that long-term survival is poor. The SEER-Medicare data includes mostly (91%) patients > 65 so it is not surprising the long-term survival reported here was lower.
58.	8	2. Page 1 - The statement that "the National Cancer Institute recommends against surveillance" is incorrect. The source relied on for this statement, reference 8, explicitly states that "The summary reflects an independent review of the literature and does not represent a policy statement of NCI or the National Institutes of Health (NIH).	We have taken this out.
59.	8	Page 1 - The statement that "recommendations for HCC surveillance remain controversial" is based solely on reference 12, an opinion piece (not an original study) written by two VA authors who failed to disclose that they were attempting to secure VA funding to support a placebo-controlled study of surveillance.	We have consolidated the background section and added additional references.
60.	8	Methods 1. Page 8 - No statistical justification is provided for the arbitrary sample size of 500 patients required for inclusion of studies involving OLT.	We did not use a sample size limit for observational studies comparing treatment to no treatment. The sample size of 500 patients applied to noncomparative observational studies – as such, there is no statistical testing. Rather, we included such studies both as a way to understand harms of treatment as well as long-term survival. The sample size of 500 was chosen because there were several, large noncomparative observational studies of OLT and we felt these would provide more generalizable data regarding long-term survival and harms. The smaller, noncomparative observational studies we examined had similar findings and would not have changed the results – good long-term survival and poor reporting of harms.
61.	8	2. Page 8 – The rationale for not including studies that compared two or more active treatments was not provided. Active-controlled studies are well recognized as a valid mechanism for establishing efficacy of an intervention, particularly when ethical considerations preclude a placebo or no-treatment arm. Of note, the PCORI methodology relied on by the review (reference 23) explicitly endorses active comparator studies; PCORI standard RQ-5 states that "non-use (or no specific treatment) comparator groups should be avoided unless no specific treatment is a likely option in standard care."	The PCORI causal inference standards to which we refer apply to observational studies. The active-controlled studies guidance applies to situations in which the benefits of treatment have been established. The strength of evidence for HCC screening depends in part on the balance of benefits and harms of treating screen-detected disease. Screening is likely to detect additional small, early-stage HCC. The efficacy of treating such screen-detected tumors compared to watchful waiting is the question that applies most directly to the HCC screening and we were tasked – after discussion with a group of topic-specific and screening-methodology experts - with finding and examining such studies. Examining active comparator studies would not have contributed substantially to the strength of evidence for HCC screening. We have clarified that our approach to the treatment literature was meant to examine the relative benefits of treating screen-detected disease and that our review does not examine the issue of HCC treatment as a whole.
62.	8	3. Page 9 – The statement that there is no widely accepted standard for quality assessment of observational studies" is incorrect. The Strobe (Strengthening the Reporting of Observational Studies in Epidemiology; www.strobe-statement.org) is referenced by the Uniform Requirements for Manuscripts Submitted to Biomedical Journals by the International Committee of Medical Journal Editors and by the Cochrane Collaboration, and endorsed by the *Annals of Internal Medicine*, *British Medical Journal*, and *Lancet.*	The STROBE statement summarizes the elements an observational study manuscript should report – it is not a standard for quality assessment of observational studies. It is akin to the CONSORT statement for trials. There is no widely accepted quality assessment tool for observational studies (see AHRQ methods guide for effectiveness and comparative effectiveness reviews, Chapter 4). There are various tools for assessing the quality of observational studies and we drew from several as appropriate to this topic.

	Reviewer	Comment	Response
63.	8	Results 1. Pp. 11-15: The discussion of surveillance did not include the population study of McMahon *et al.* (2000), which demonstrated a survival benefit in Alaska natives with chronic hepatitis B.	We examined the McMahon study, but it was excluded because the use of historic controls was a pre-specified exclusion criteria. This was an exclusion criterion because of the potential for additional confounding from secular trends in overall management of liver disease patients. We included a large body of observational studies and many, as we describe, found that screen-detected patients had longer survival. Whether screening truly prolonged survival or screen-detected patients appeared to live longer because of lead-time and other biases is not clear. The McMahon study would have provided findings and methodologic issues similar to the many other observational studies we examined. We did revise our presentation of observational study results in the table and narrative portions to make the precise survival data and hazard ratios easier to follow.
64.	8	2. Pp. 15-8 – The conclusions regarding the effects of OLT, resection, RFA, and TACE do not reflect the following data, which were not included in the review: a. El-Serag *et al.* (2006) used data on HCC patients from SEER and Medicare claims to construct a Cox proportional hazards model to identify predictors of outcome; the model incorporated tumor size and extent. The type of therapy received was a significant predictor of survival.	See #32
65.	8	b. A similar study by Schwartz and Smith (2008), adjusting for disease extent and vascular invasion among other factors, found risk ratios of 0.56 for transplantation and 1.53 for ablation. As noted below, ablation is associated with increased survival compared to other modalities.	See #32
66.	8	c. Although cited in this document, the data from reference 18 (Mazzaferro *et al.* 1996) were not compared with historic controls for early stage HCC; the survival rates in this study far exceed those observed in patients with untreated early disease (see references listed under Comment 3).	See #62
67.	8	d. Although cited in this document, the results from reference 49 (Liu *et al.* 2004), which found that HCC resection was associated with a mortality reduction of 55%, are dismissed with statements about "confounding by indication" and concerns about performance status as a confounding. There is no evidence that the natural history of HCC is significantly affected by etiology (assuming that that was what the authors meant). Liu *et al.* excluded patients from their study who had medical contraindications to surgery, which makes confounding by performance status unlikely.	We have rewritten this part. We agree that the Liu study is probably the best of the comparative observational studies and we describe the results as low-strength evidence. As Liu et al themselves acknowledge, they did not have information on comorbidities and other confounding factors.
68.	8	e. Cho *et al.* (2009) conducted a meta-analysis of trials comparing ablation with percutaneous ethanol injection in patients with early HCC and found a significant survival advantage for ablation. Similar results were obtained by Orlando *et al.* (2009)	See #32
69.	8	f. Llovet *et al.* (2002) conducted a landmark prospective RCT comparing TACE to TAE to conservative therapy in patients with early stage HCC. TACE was associated with a significant survival benefit.	See #32
70.	8	3. Page 21 – Discussion The statement that "Our focus on studies comparing active treatment to conservative management admittedly may have missed important effects of current treatments for HCC since many have been evaluated in the context of comparative effectiveness studies" is an understatement. If there was a question about whether treatment of early HCC is effective, inclusion of controlled trials with a superiority design would have been absolutely appropriate. These do not represent "comparative effectiveness" studies, as suggested in the text.	See #16