SUPPLEMENTARY MATERIALS

I. GENERAL STATISTICAL AND MATCHED SURVIVAL ANALYSIS

II. UNMATCHED SURVIVAL ANALYSIS

III. EMULATED TARGET TRIAL PROTOCOL

IV. EMULATED TARGET TRIAL COHORT SIZES AFTER APPLYING SELECTION CRITERIA

V. EMULATED TARGET TRIAL BALANCE DIAGNOSTICS

VI. REFERENCES CITED IN SUPPLEMENTARY MATERIALS

I. GENERAL STATISTICAL AND MATCHED SURVIVAL ANALYSIS

Differences between continuous variables were tested using ANOVA or the Kruskall-Wallis test. Chi-square with continuity correction tested differences in proportions. Kaplan-Meier survival curves were assessed by log-rank tests. Hazard ratio (HRs) and 95% confidence intervals (95% CI) were estimated using Cox proportional hazards regression. Multivariable Cox analysis was performed with and without 2-way interactions, with difference in model fit tested by the likelihood ratio test. Proportionality was confirmed using Schoenfeld residuals. Statistical significance was based on two-sided $p < 0.05$. Analyses were conducted using SAS Studio 3.8 (SAS Institute Inc., Cary, NC) and R 3.6.0 (R Foundation). The R packages 'matchit' and 'cobalt' were used for propensity-score matching and balance assessment, respectively[1-3].

II. UNMATCHED SURVIVAL ANALYSIS

Kaplan-Meier curves comparing OS between the TARE and systemically treated patients in the unmatched raw cohort (n=1514) are shown in Figure S1. Median survival was 9.7 (95% CI 8.4 – 12.3) months for the TARE-treated group versus 6.5 (95% CI 6.1 – 7.2) months for the systemically treated group ($p < 0.0001$). However, treatment groups in this unmatched cohort showed a lack of balance on multiple covariates (manuscript Figure 3) and the results from a well-balanced propensity-score matched pragmatic analysis were more modest yet still significant. Although in this study the results did not differ greatly between matched and unmatched analyses, the potential for bias inherent to a poorly matched observational cohort analysis limits confidence in the precision of results from the latter, and in other cases such analyses have resulted in exaggerated estimates of the treatment effect[4-6]. The use of propensity-score balancing in this study raises confidence in the precision of the observed effect on outcome.

III. EMULATED TARGET TRIAL PROTOCOL

1. Overview

   This emulated target trial (ETT) was designed to approximate clinical trial design elements of an industry-sponsored phase III prospective randomized clinical trial titled "90Yttrium trans-arterial radio-Embolization (TARE) vs. Standard of care (chemotherapy) for the treatment of advanced Hepatocellular Carcinoma (HCC) with Portal vein thrombosis" (YES-P, ID NCT01887717). Since this trial was closed before completion, no results or data from this trial are available in the clinicaltrials.gov record.

2. Protocol Design

   The ETT is a two arm observational data analysis using propensity-score matched cohorts drawn from the US National Cancer Database (NCDB). The actual target trial protocol (ATTP) that will serve as the template for this ETT was based on NCT01887717, downloaded from clinicaltrials.gov (YES-P Protocol v3.0 HCC Phase III Protocol TS-104, dated 2014/08/08, file name Prot_000.pdf, access date November 27, 2020).

3. ETT Study Objective

   To assess the efficacy of TARE in comparison to standard of care (chemotherapy) for the treatment of patients with unresectable hepatocellular carcinoma associated with major vascular invasion (MVI).

4. Primary Endpoint

   Overall survival (OS) from date of diagnosis

5. Treatment Arms

   The ATTP was an industry-sponsored multi-center Phase III RCT that required specific drug agents for the treatment and control arms. The specified agent of the experimental arm was TheraSphere (yttrium-90 glass microspheres) and the specified agent of the control arm was sorafenib (Nexavar, generic not available). In contrast, the ETT is a nationwide observational data analysis comparing two classes of therapeutic

agents, TARE (of which ThereSphere is a specific agent) and chemotherapy (of which sorafenib is a specific agent). Both are currently being used in clinical practice to treat advanced HCC. Proprietary or specific therapeutic agent, either approved or investigational, are not specified for the ETT.

The arms of the ETT are defined as follows:

The control arm of the ETT is single-agent systemic therapy, the standard of care for advanced-stage HCC, as recommended by AASLD/EASL guidelines. In the NCDB, this control arm is identified based on treatment codes signifying single agent chemotherapy (1,2) as the first course of treatment.

The treatment arm of the ETT is defined as intraarterial radioisotopic injection for the purpose of embolizing the tumor or tumor affected hepatic segment/lobe to deliver interstitial brachytherapy. The Commission on Cancer accredited institutions participating in the NCDB employed multiple codes under the radiation treatment modality to identify TARE (brachytherapy, low-dose rate interstitial or NOS (7, 10), or radioisotope (13)).

6. Treatment Cross-Over

With the exception of patients who progress clinically while on sorafenib, the ATTP allowed patients assigned to one arm to cross-over to the other arm if clinically warranted (eg. patient did not tolerate their assigned treatment). In keeping with the intention to treat (ITT) principle, these patients remained in their original assigned arms for analysis. Data recorded in the NCDB allows this cross-over rule to be applied in the ETT (see part 14).

7. Treatment Timing

   The ATTP specified that the control arm be treated within 2 weeks of randomization, and the treatment arm be treated within 3 weeks. Since the date of diagnosis serves as a surrogate for the date of randomization in the ETT, it would not be feasible to apply such treatment time windows in the ETT analysis. A series of conditional landmark analyses will be performed covering 30, 60, 90 days from the start of follow-up.

8. Eligibility Criteria

   Eligibility criteria from the ATTP and their corresponding translations for the ETT are shown in Table S1. Identical criteria are not absolutely required for target trial emulation. Criteria deemed not essential or not translatable by variables available from the NCDB are indicated as [No corresponding criteria]. None of the unmapped criteria were determined to be crucial for meeting the study objectives of the ETT.

9. Method for Assigning Patients to Treatment Groups

The ATTP involved trial site randomization at a 1:1 ratio with stratification based on alpha-fetoprotein level > 400 ng/mL. The ETT will take advantage of an almost 3-fold larger number of chemotherapy-treated patients by performing pseudo-randomization (ie. propensity score matching) at up to 3:1 ratio to increase statistical power. In lieu of stratification, propensity-score matching will include covariate balancing based on alpha-fetoprotein level >= 400 ng/mL.

10. Blinding

The ATTP did not involve blinding.

11. Determination of Sample Size

The ATTP states that the study was designed to detect a 4 month increase in median OS time from 9 months in the control arm to 13 months in the treatment arm (ie. HR 0.69) using a log rank test. An estimated 250 subjects will yield 80% power to detect this target difference with a two-sided alpha of 0.05.

12. Evaluations and Study Visits

Not applicable to the ETT as it will only evaluate the primary endpoint of OS.

13. Time Points

The ATTP defined the first day of follow-up to be the date of randomization. As there is no date of randomization for an ETT, the default date of diagnosis as recorded in the

NCDB participant user file will serve as the baseline date for calculating time to event (ie. first day of follow-up). When necessary to convert units of months into days, the following formula will be used: Days = Months * 365.25 / 12.

14. Causal Estimands

The ATTP implements an ITT basis but includes both ITT and per-protocol analyses. This approach is also feasible for the ETT: For the chemotherapy arm of the ETT, the ITT principle is implicit as the natural discontinuation rate of chemotherapy will be reflected in the NCDB time-to-event data given the manner by which the first course of treatment is recorded. Because the ATTP selection criteria was intended to screen out patients who would be disqualified from chemotherapy, patients entered into the chemotherapy arm are highly likely to receive treatment as planned (as was also observed in the SIRveNIB and SARAH trials), but this is not an absolute requirement for an ITT analysis. The measured time-to-treatment interval is also relatively short in the chemotherapy arm, with the data expected to reflect a neglible amount of fallout due to disease progression.

For the TARE arm of the ETT, discontinuation following initiation of treatment is not relevant because TARE is delivered as a same-day treatment, usually all at one time (on occasion, two times). A more serious issue in regards to emulating the ITT principle is the higher likelihood of TARE being cancelled as compared to chemotherapy. Therefore, in order to adhere to ITT principles, patients in whom radiation treatment was

recommended/planned but not ultimately performed will be identified and included in the TARE arm.

Since the clinical alternative to TARE would be chemotherapy, and because the ATTP did allow for treatment cross-over from the TARE group, it is appropriate to identify such patients from among the sorafenib treated group drawn from the NCDB. These patients can be identified using the REASON_FOR_NO_RADIATION variable recorded in the participant user file. To simulate ITT, these patients will be excluded from the chemotherapy group and added to the TARE group prior to propensity score matching and outcomes analysis.

15. Efficacy

The primary efficacy variable is overall survival (OS), defined as the time from the baseline date until the date of death due to any cause, calculable using DX_LASTCONTACT_DEATH_MONTHS. No secondary efficacy variables will be evaluated by the ETT.

16. Adverse Events and Quality of Life

Since only the primary outcome is being evaluated, no adverse event or quality of life data will be analyzed by the ETT.

17. Statistical Plan

The statistical analysis plan is described in the Section I of this document and the primary manuscript.

IV. EMULATED TARGET TRIAL COHORT SIZES AFTER APPLYING SELECTION CRITERIA

The effects of applying specific selection criteria on emulated target trial cohort size are shown in Table S2.

V. EMULATED TARGET TRIAL BALANCE DIAGNOSTICS

After propensity-score matching, the emulated target trial cohort became well balanced (standardized mean differences of < 0.01 for all covariates), emulating the effects of randomization. The covariate balance plot for this cohort is shown in Figure S2.

VI. REFERENCES CITED IN SUPPLEMENTARY MATERIALS

1. Zhang Z, Kim HJ, Lonjon G, Zhu Y, written on behalf of AMEB-DCTCG. Balance diagnostics after propensity score matching. Ann Transl Med 2019;7(1):16. DOI: 10.21037/atm.2018.12.10.
2. Ho DI, K: King, G; Stuart, E. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. Political Analysis 2007;15(3):199-236.
3. Greifer N. cobalt: Covariance Balance Tables and Plots. (https://CRAN.R-project.org/package=cobalt).

4. Dickerman BA, Garcia-Albeniz X, Logan RW, Denaxas S, Hernan MA. Avoidable flaws in observational analyses: an application to statins and cancer. Nat Med 2019;25(10):1601-1606. DOI: 10.1038/s41591-019-0597-x.

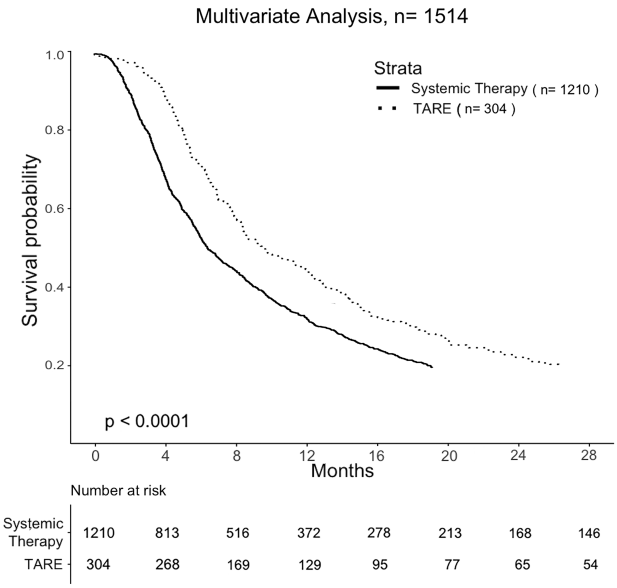
5. Vaughan AS, Kelley CF, Luisi N, del Rio C, Sullivan PS, Rosenberg ES. An application of propensity score weighting to quantify the causal effect of rectal sexually transmitted infections on incident HIV among men who have sex with men. BMC Med Res Methodol 2015;15:25. DOI: 10.1186/s12874-015-0017-y.

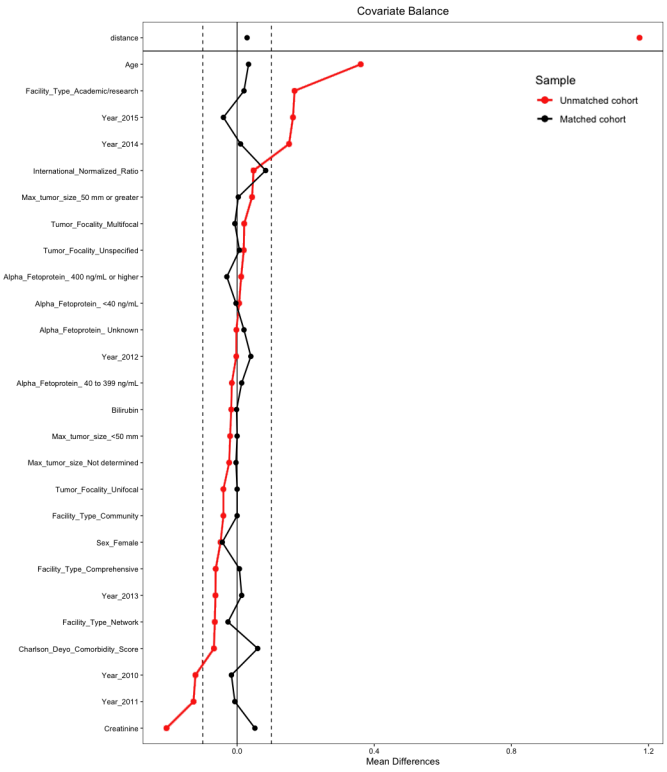
6. Hernan MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. Epidemiology 2008;19(6):766-79. DOI: 10.1097/EDE.0b013e3181875e61.

SUPPLEMENTARY FIGURE CAPTIONS

Figure S1: Kaplan-Meier curves comparing OS between the TARE and systemically treated patients in the unmatched raw cohort. The unmatched cohort showed a lack of balance on multiple covariates (manuscript Figure 3) and the results from a well-balanced propensity-score matched pragmatic analysis were more modest.

Figure S2: Standardized mean differences of the covariates before and after propensity-score matching. Vertical hash-lines demarcate 0.1 standardized mean difference as the threshold for covariate balance. The plot also shows a substantial reduction in the overall measure of imbalance ('distance').


Multivariate Analysis, n= 1514
Strata
Systemic Therapy (n= 1210)
TARE (n= 304)
Survival probability
1.0
0.8
0.6
0.4
0.2
p < 0.0001
0
4
8
12
16
20
24
28
Months
Number at risk
Systemic Therapy
1210
813
516
372
278
213
168
146
TARE
304
268
169
129
95
77
65
54


Covariate Balance
distance
Age
Facility_Type_Academic/research
Year_2015
Year_2014
International_Normalized_Ratio
Max_tumor_size_50 mm or greater
Tumor_Focality_Multifocal
Tumor_Focality_Unspecified
Alpha_Fetoprotein_ 400 ng/mL or higher
Alpha_Fetoprotein_ <40 ng/mL
Alpha_Fetoprotein_ Unknown
Year_2012
Alpha_Fetoprotein_ 40 to 399 ng/mL
Bilirubin
Max_tumor_size_<50 mm
Max_tumor_size_Not determined
Tumor_Focality_Unifocal
Facility_Type_Community
Sex_Female
Facility_Type_Comprehensive
Year_2013
Facility_Type_Network
Charlson_Deyo_Comorbidity_Score
Year_2010
Year_2011
Creatinine
Sample
Unmatched cohort
Matched cohort
0.0
0.4
0.8
1.2
Mean Differences

ATTP Eligibility Criteria	ETT Eligibility Criteria
Age > 18 years, regardless of race and gender	Age > 18 years, regardless of race and gender
HCC confirmed histology or non-invasive criteria (EASL/AASLD)	Histology code 8170, ICD-10 site code C22.0
Advanced stage with portal vein thrombosis, hepatic vein invasion is excluded	AJCC Stage 3B with major vascular invasion defined as invasion of branches of the main portal vein or one or more of the three hepatic veins
Treatment naïve or recurrent HCC after curative treatment	Treatment naïve HCC
Child-Pugh A	Total bilirubin < 2.0 mg/dL, INR < 1.7, Exclude CDCS 2 and 3
ECOG 0-1	CDCS 0-1
Creatinine < 2.0 mg/dL	Creatinine < 2.0 mg/dL
No moderate or severe comorbid conditions	Exclude CDCS 2 and 3
No confirmed extrahepatic metastases	No recorded extrahepatic metastases
Unilobular disease	No multilobular disease based on NCDB CS_EXTENSION code
Negative serum pregnancy test in females of child-bearing potential	[No corresponding criteria]
Not breastfeeding	[No corresponding criteria]
Not on liver transplantation list	[No corresponding criteria]
No history of organ allograft	[No corresponding criteria]
No evidence of adverse effect of prior therapy	[No corresponding criteria]
Platelet Count > 50,000/uL	[No corresponding criteria]
WBC > 1500/uL	[No corresponding criteria]
AST/ALT below 5 times upper limit	[No corresponding criteria]
Tumor volume < 70% of total liver volume	[No corresponding criteria]
No indication for any curative treatment after multidisciplinary assessment	[No corresponding criteria]
No participation in concurrent interventional clinical trial	[No corresponding criteria]
Signed consent form	Not applicable.

Table S1: Emulated target trial eligibility criteria

Criteria	Trial cohort size (Systemic therapy group/TARE group)
Starting cohort	717/177 (was 750/144 before ITT emulation)
Age > 18	717/177
Charlson-Deyo Comorbidity Score 0 or 1	539/141
Creatinine < 2.0 mg/dL	446/124
Bilirubin < 2.0 mg/dL	338/95
INR < 1.7	309/91
Tumor extent = Unifocal	169/67
ITT cohort	169/67

Table S2: Effects of selection criteria on pre-match emulated target trial cohort size