HALT-C Ancillary Study Proposal

Proposal Name: Failure rates in a formalized screening program for hepatocellular carcinoma among patients in the HALT-C Trial

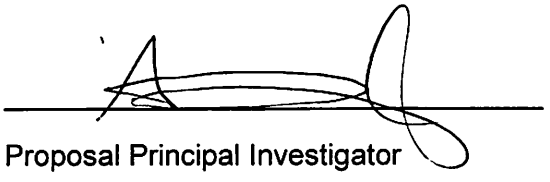
Proposal PI: Amit G. Singal

Co-Investigators: William M. Lee, Jorge A. Marrero, and Anna S. Lok

HALT-C PI: William M. Lee

Funding Agency and Review Body: UT Southwestern

I agree to follow HALT-C Policies and Procedures when conducting this study. I acknowledge that the data obtained from this study will belong to the NIH and will be placed in the HALT-C database for use by other investigators. I understand that I cannot begin experiments using HALT-C specimens/data until I receive approval from the HALT-C Ancillary Studies Committee and funding from the Scientific Review Body for my proposal. I also understand that the data analysis for this proposal will be performed by NERI (unless otherwise approved by the HALT-C study) and that Protocols approved by the HALT-C Ancillary Studies Committee will be placed on the HALT-C Restricted Website.



12/14/2010

Proposal Principal Investigator — Date

12/14/2010

HALT-C Principal Investigator — Date

1. Specific Aims: HCC is the 8th leading cause of cancer-related death in the United States and a leading cause of death among patients with cirrhosis. Screening for HCC in patients with cirrhosis using ultrasound and alpha-fetoprotein (AFP) every 6-12 months is effective at detecting tumors at an early stage and has been shown to reduce mortality. Currently, only 40% of HCC nationally are diagnosed at an early stage when curative options exist. The occurrence of late stage cancers implies a breakdown in the screening process. Systematically identifying these screening process failures is essential to designing effective interventions to improve screening rates. The Quality in the Continuum of Cancer Care (QCCC) conceptual model categorizes screening process failures as an a) absence of screening, b) absence of follow-up for abnormal tests, or c) failure of detection despite completing screening testing [1]. It has been successfully used to examine factors associated with failures in breast and cervical cancer screening [2] but a similar analysis has never been performed for HCC screening.

The HALT-C Trial provides an ideal platform to explore failure rates in a formalized HCC surveillance program and identify factors associated with screening process failures among a highly selected group of patients prospectively followed for up to 8.7 years. The goal of HALT-C was to determine the efficacy of maintenance PEG-IFN in preventing disease progression in HCV patients with advanced fibrosis who were non-responders to PEG-IFN and ribavirin and one of its primary aims was to evaluate the incidence of HCC. All patients were enrolled in a standardized HCC surveillance protocol, including AFP testing every 3-6 months and ultrasounds every 6-12 months. Patients with an elevated or increasing AFP or a suspicious lesion on ultrasound were to be evaluated with CT or MRI. A panel of investigators adjudicated each HCC diagnosis including the date of presentation, date of diagnosis, date of last negative imaging, and tumor stage at diagnosis. After a median follow-up of 6.1 years, 88 patients have developed HCC. We propose to use data from HALT-C to evaluate the following two specific aims:

Aim 1: To characterize screening process failure rates among a large cohort of patients with bridging fibrosis and cirrhosis. Despite an ideal prospective clinical trial setting executed in academic centers, we hypothesize that at least 40% of patients will have a failure in the screening process. We anticipate that absence of screening will account for more screening process failures than absence of follow-up or failure of detection.

Aim 2: To identify factors associated with failures in the HCC screening process. We anticipate that previously described factors, including patient age, race, and gender, will not be associated with an absence of screening in an ideal clinical trial setting.

2. Background HCC is a leading cause of death among patients with cirrhosis [3]. Patients detected at an early stage can achieve 70% five-year survival rates with resection or transplantation, in contrast to a median survival of 17 months with advanced HCC. Despite ultrasound and AFP having a sensitivity of 70% for detecting HCC at an early stage [4], over 60% of patients with HCC are diagnosed at advanced stages [3]. The occurrence of late stage cancers implies a breakdown in the screening process.

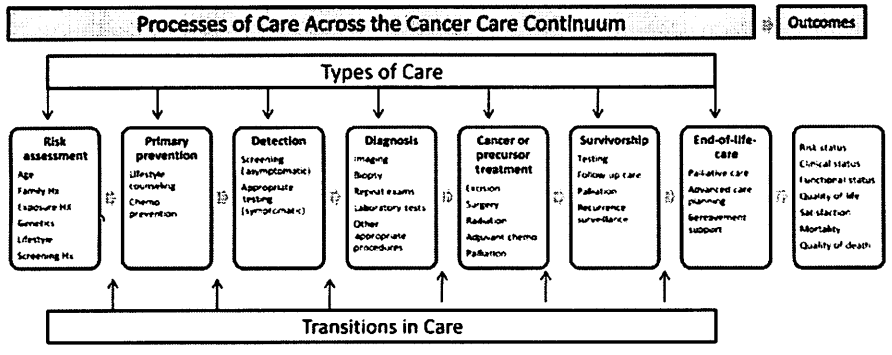


Figure 1: The Quality in the Continuum of Cancer Care model

The QCCC conceptual model provides a framework for improving cancer care along its entire spectrum, from prevention to end-of-life care (Figure 1) [1]. It emphasizes the

relationship between steps of care and health outcomes, identifies the potential for failure at each step, and suggests interventions to improve performance. An effective HCC screening process depends on providers recognizing at-risk patients, referring these patients for ultrasound and AFP (screening), accurate tests (detection), and timely evaluation after a positive screening test (follow-up). Although all patients enrolled in HALT-C were recognized as being at risk for HCC and were enrolled in an HCC surveillance program, the possibility of subsequent screening failures still existed. Prior studies that applied the QCCC to breast and cervical cancer demonstrated that an absence of screening or follow-up accounts for more late-stage cancers than does the quality of the screening tests [2]. An analysis of failure rates in the HCC screening process has yet to be performed.

Many high-risk patients with cirrhosis fail to receive HCC screening in clinical practice. In a study of 541 Medicare patients with known cirrhosis and HCC, only 30% of patients had screening *once* during the preceding three years and less than 15% received screening consistent with previously established practice guidelines [5]. Patients who were younger, female, Asian, or lived in areas of higher income and education were more likely to have received prior HCC screening. Appropriate follow-up of abnormal screening tests is also crucial. Several studies have documented an absence of follow-up for abnormal screening tests for breast, cervical, and colon cancer screening, which can often be traced back to physicians not knowing about the test result or lack of time or knowledge [2]. Failure rates for follow-up of abnormal HCC screening tests have never been studied.

Several patient and system level factors influence failure rates for cancer screening and follow-up and must be considered when designing interventions to improve the cancer screening process [6]. Patient factors of note include age, gender, race/ethnicity, and socioeconomic status. Provider factors include number of years in practice, time constraints in clinic, and knowledge of guidelines. System factors include availability of screening tests, wait times for testing, reminder systems, and availability of treatment options.

3. Relationship to Aims of HALT-C: The goal of HALT-C was to determine the impact of maintenance PEG-IFN therapy on long-term outcomes in HCV patients. The study has produced over 65 publications evaluating many aspects of disease progression, including the incidence of HCC and the accuracy of HCC biomarkers for the detection of early HCC. Despite close follow-up at academic medical centers, 25% of HCC cases in HALT-C were diagnosed at advanced stages. Once again, the occurrence of late stage tumors suggests failures in the screening process despite optimal conditions. The proposed study will identify possible intervention targets that could shed light on ways to further improve HCC surveillance in patients with advanced liver disease.

4. Study Design and Experimental Groups: For Aim 1, we will examine screening failure rates among all 1050 randomized patients in the HALT-C Trial. Given that surveillance testing and dates were recorded for each patient, HALT-C allows us to determine screening process failure rates for those who developed HCC as well as those who did not develop HCC. Process failures, even in patients without HCC, are important intervention targets given their potential to lead to outcome failures (i.e. late stage tumors) when uncorrected over time. For Aim 2, we will identify factors associated with any screening failure in the entire cohort of 1050 patients. Given the limited number of advanced HCC cases, we will compare screening failure rates between patients with early HCC and those with advanced HCC as a secondary exploratory analysis.

5. Methods and Data Usage

Aim 1: We will review HCC screening data for all patients to determine failure rates in the HCC screening process. Screening failure could be related to many causes including patients not coming to clinic visits, physicians not ordering an ultrasound and/or AFP, patients not complying

with screening tests, screening tests not detecting HCC, or physicians failing to follow-up on positive screening tests. Therefore, each failure in the screening process will be categorized into one of three mutually exclusive categories: absence of screening, absence of follow-up, and failure of detection. The primary outcome will be the percent of patients with an absence of screening, which will be obtained by dividing the total number of patients with an absence of screening by the total number of patients in the cohort at each period in time. The percent of patients with either an absence of follow-up or failure of detection will be secondary outcomes.

Absence of screening will be defined as a lack of ultrasound and AFP performed within each twelve-month period. Absence of screening will be assessed every 12 months during follow-up for each patient to determine if an ultrasound and AFP were done over that preceding twelve-month period. We chose a one-year cut-off based on recommendations in the HALT-C surveillance protocol. Patients with one test performed in isolation (i.e., ultrasound without AFP or vice-versa) will be coded as an absence of screening. We will unfortunately not be able to reliably discriminate between cases in which screening was not ordered and those in which patients did not comply with a scheduled screening visit.

Absence of follow-up will be defined as a lack of cross-sectional imaging within three months of any positive screening test. Patients with a suspicious mass on ultrasound or AFP > 20 ng/mL should be evaluated with cross-sectional imaging (CT scan or MRI) [7]. We chose a 3-month cut-off based on tumor doubling time as well as the frequency of clinic visits during HALT-C.

Failure of detection will be defined as cases of late stage HCC (beyond stage T2) that occur despite completion of screening and follow-up (as defined above). Failure of detection is a diagnosis of exclusion and will not be coded in cases with an absence of screening or an absence of follow-up. Staging of HCC will be defined using TNM staging, with late stage tumors defined as those tumors beyond T2.

Data Required:
We are requesting dates and results for all clinic visits, AFP testing, ultrasounds, and CT/MRIs for each patient during the entirety of their follow-up. For imaging results, we request data regarding the type of imaging (ultrasound, CT, or MRI), presence vs. absence of a liver mass, if the mass was well-defined vs. ill-defined, any comparison to prior films (stable vs. increased vs. deceased), and if further evaluation was felt to be necessary (including reasons if follow-up not needed). We also request available tumor data including date of diagnosis, method of diagnosis, definite vs. probable HCC, number of nodules, maximum diameter, tumor location, vascular invasion, extrahepatic spread, T stage, N stage, and M stage. We will require dates of death, transplantation, and last clinic visit for all patients in order to censor patients at these times. We also request baseline data to describe the patient cohort (including patient age, race, gender, BMI, presence of diabetes, Child Pugh score, HCV genotype, platelet count, AST, ALT, alkaline phosphatase, bilirubin, albumin, INR, the stage of fibrosis, presence of varices, lifetime alcohol and smoking history, and treatment assignment).

Aim 2: We will identify factors associated with an absence of screening during follow-up. Absence of screening will be defined as a vector with a binary outcome (presence vs. absence of screening) at each time point of interest for each patient. Independent variables of interest are listed below. A secondary exploratory analysis for Aim 2 will be a case-control study to compare screening process failures in those with early HCC and those with late stage HCC.

Data Required: We will use patient factors including age, race/ethnicity, gender, income, education, clinical characteristics (duration of HCV infection, cirrhosis vs. bridging fibrosis on biopsy (baseline, year 1.5 and year 3.5), Child Pugh score at baseline and end-of-study, presence of varices at baseline, and lifetime alcohol and tobacco history), other measures of

compliance (compliance with clinic visits), and trial characteristics (IFN vs. placebo arm, initial 3.5 years of the study vs. the subsequent follow-up period, and study center).

6. Anticipated Results

Aim 1: Although patients in a clinical trial setting will have lower rates of screening failure than those seen in routine clinical practice, we hypothesize that at least 40% of patients will have a failure in the screening process, with most of these patients experiencing an absence of screening. Although some patients will have a failure of detection or absence of follow-up for abnormal test results, these will account for a minority of breakdowns in the screening process.

Aim 2: We believe that previously described factors, including patient age, race, gender, socioeconomic status, and education will not be associated with an absence of screening in patients receiving care within a clinical trial.

7. Statistical Analysis

Aim 1 The primary outcome for Aim 1 is the percent of patients with a failure in the screening process due to an absence of screening. Absence of screening will be defined as a lack of screening with ultrasound and AFP within each twelve-month period. We will obtain a point estimate by dividing the total number of patients with an absence of screening by the total number of patients in the cohort at each period in time. The Clopper-Pearson exact confidence interval method will be used to obtain 95% confidence interval estimates. If the screening absence rate is 40% as anticipated, we can calculate the screening absence rate with a precision of ±4.9% given our sample size of 1050 patients.

Secondary outcomes for Aim 1 are the percent of patients with a failure in the screening process related to an absence of follow-up or a failure of detection. Given the anticipated small number of positive screening tests, we anticipate a limited number of screening failures related to absence of follow-up or failure of detection. We will determine point estimates and 95% confidence intervals for both outcomes using the same methods as above but expect very wide confidence intervals given the limited sample size.

Aim 2: Our primary outcome for Aim 2 is factors associated with an absence of screening. We will define the outcome variable by a vector with a binary outcome (1=absence of screening, and 0=presence of screening) at each time point of interest. We will analyze the data with a *Generalized Linear Model* using the logit link function for binary response. The marginal model approach will be used to model the marginal expectation of the response on the explanatory variables. The marginal expectation represents the population average probability of having an absence of screening per year. For an outcome measure at a single time point, this is essentially a logistic regression model. However, since we have repeated measures of the binary outcome, we will model these repeated measures as a function of the explanatory variables and determine their association with the marginal expectation (probability of having an absence of screening per year). To estimate the regression coefficients for this model with repeated measures, we will use the *alternating logistic regression implementation of generalized estimating equations (GEE)*, which overcomes the problem of GEE by explicitly modeling the association between the screening outcomes at various time points. The regression coefficients can be interpreted as the weighted averages of the cross-sectional coefficients for each visit.

For model selection, the initial model will include all the explanatory factors and subsequently a modified backward selection procedure will be used. A set of predictors of primary interest are forced into the model, as well as confounding variables thought to be important for face validity. The remaining variables are ranked in the order of importance and variables in this set are deleted in ascending order of importance until the first variable that meets the criteria of $P<0.2$ is encountered. Then the backward selection procedure is stopped and the final model is selected. The final model will identify the multiple important predictors associated with the outcome of

absence of screening. In our case, we will include patient age, gender, race, and presence of cirrhosis as a priori predictors of interest.

8. HALT-C samples to be used: We are requesting patient data as detailed above but do not need any tissue, blood, or DNA samples.

9. Financial Issues: Any expenses will be paid from Dr. Singal's discretionary funds available under the Dedman Scholar program. Statistical support is available free-of-cost to Dr. Singal through the Clinical Scholars program at UT Southwestern.

References

1. Zapka, J.G., et al., *A framework for improving the quality of cancer care: the case of breast and cervical cancer screening.* Cancer Epidemiol Biomarkers Prev, 2003. **12**(1): p. 4-13.
2. Leyden, W.A., et al., *Cervical cancer in women with comprehensive health care access: attributable factors in the screening process.* J Natl Cancer Inst, 2005. **97**(9): p. 675-83.
3. Altekruse, S.F., K.A. McGlynn, and M.E. Reichman, *Hepatocellular carcinoma incidence, mortality, and survival trends in the United States from 1975 to 2005.* J Clin Oncol, 2009. **27**(9): p. 1485-91.
4. Singal, A., et al., *Meta-analysis: surveillance with ultrasound for early-stage hepatocellular carcinoma in patients with cirrhosis.* Aliment Pharmacol Ther, 2009. **30**(1): p. 37-47.
5. Davila, J.A., et al., *Use of surveillance for hepatocellular carcinoma among patients with cirrhosis in the United States.* Hepatology, 2010. **52**(1): p. 132-41.
6. Taplin, S.H. and A.B. Rodgers, *Toward improving the quality of cancer care: addressing the interfaces of primary and oncology-related subspecialty care.* J Natl Cancer Inst Monogr, 2010. **2010**(40): p. 3-10.
7. Bruix, J. and M. Sherman, *Management of Hepatocellular Carcinoma: An Update.* Hepatology, 2010.

Protocol Part III: Sample Requirements

Visit	Liver # patients, mm*	Blood # patients, ml	DNA # patients, ug	Liver Biopsy Slides # patients, slides/patient	Other (describe) # pts, amount
Screen 1					
Screen 2					
Baseline					
Lead in Week 4					
Week 8					
Week 12					
W16					
Week 20					
Week 24					
Randomized Month 9					
Month 12					
Month 15					
Month 18					
Month 21					
Month 24					
Month 27					
Month 30					
Month 33					
Month 36					
Month 39					
Month 42					
Month 45					
Month 48					
Post-treatment					
Responders W30					
W36					
W42					
W48					
W60					
W72					

Data needed (please specify):
For Aim 1, we are requesting dates and results for all clinic visits, AFP testing, ultrasounds, and CT/MRIs for each patient during the entirety of their follow-up. For imaging results, we request data regarding the type of imaging (ultrasound, CT, or MRI), presence vs. absence of a liver mass, if the mass was well-defined vs. ill-defined, any comparison to prior films (stable vs. increased vs. deceased), and if further evaluation was felt to be necessary (including reasons if follow-up not needed). We also request available tumor data including date of diagnosis, method of diagnosis, definite vs. probable HCC, number of nodules, maximum diameter, tumor location, vascular invasion, extrahepatic spread, T stage, N stage, and M stage. We will require dates of death, transplantation, and last clinic visit for all patients in order to censor patients at

these times. We also request baseline data to describe the patient cohort (including patient age, race, gender, BMI, presence of diabetes, Child Pugh score, HCV genotype, platelet count, AST, ALT, alkaline phosphatase, bilirubin, albumin, and INR).

For Aim 2, we will use patient factors including age, race/ethnicity, gender, insurance status, income, education, clinical characteristics (duration of HCV infection, cirrhosis vs. bridging fibrosis on biopsy (baseline, year 1.5 and year 3.5), Child Pugh score at baseline and end-of-study, presence of varices at baseline, and lifetime alcohol and tobacco history), other measures of compliance (compliance with clinic visits), and trial characteristics (IFN vs. placebo arm, initial 3.5 years of the study vs. the subsequent follow-up period, and study center).