Official Title:	Hepatitis C-Video vs. Brochure Education Delivery
NCT number:	NCT04204447
Document Type:	Statistical Analysis Plan
Date of the Document:	June 30, 2018

Statistical Methods

We will initiate the activities on this study with collection of patient demographics at all sites. As part of the preparatory work, we will conduct cognitive testing of the survey instruments to be administered to ensure reliability and accuracy.

The population will be designated into 2 groups i.e. Intervention & Control. The intervention will have the video class intervention and the control will read an HCV educational brochure. We will design both interventions to assess the level of HCV knowledge. Information provided by the brochure and the video will be assessed by the pre/post test questions. These interventions will occur after the pre-test assessment by a designed questionnaire which will include questions on different aspects of hepatitis-C infection disease processes. The last step before analysis will be post interventional assessment with the same questionnaire used for pre-intervention assessment administered 1 month (+/- 2 weeks) after the intervention. All of the intervention and assessment tools will be validated before use. The second assessment may be done over the phone if the subject has no scheduled visits to the clinic during the follow-up period.

At the end we do plan to utilize the data for analysis and manuscript preparation.

Description of the clinics: The two clinics are approximately the same in terms of the percentage of rural patients, of rural HCV seropositive patients, and in the gender distribution. Because patient characteristics differ on the basis of age, race, and ethnicity, we will perform 1:1 matching of the subjects, if possible from the two clinics on these variables.

Statistical analysis: We will conduct exploratory data analysis using box plots for comparison between distributions of continuous variables. Histograms also will provide an additional way of comparing these distributions. Continuous variables are going to be summarized by means and standard deviations or medians and interquartile range, as appropriate. Discrete, that is, categorical variables will be summarized by proportions.

To understand the impact of clinic and individual patient characteristics on the HCV seropositive individuals, we will begin with building logistic regression models. If the outcome variable is defined as “1” when an HCV seropositive individual is identified and “0” otherwise, logistic regression will provide the association of HCV seropositivity as a function of clinic and individual characteristics. Further, to understand the changes in these percentages over time, we will perform a trend analysis at the minimum.

We will test the differences between the scores obtained on the post-intervention assessment and the 1-month follow up assessment. We will assess the distribution of the scores. Modeling will be done using generalized linear mixed-effects models or an alternative appropriate approach. We will include demographic factors and other covariates as appropriate.

Power/Sample Size:

The following table shows the sample size needed per group, as a function of the difference in the proportion of patients from both clinics who appear for the first appointment and the desired power of the test. The level of significance is 0.05.

Differences in proportions	Power	
	0.80	0.90
0.10	293	392
0.15	138	184
0.20	81	108

0.25	54	72
0.30	39	51

Thus, if our difference is 20% between those who make the first appointment for an HCV evaluation in the clinic that receives the SAMHSA brochure and the full education groups, we will need 81 patients per group (total 162) to have an 80% power.