Statistical Analysis Plan
Hopewell Hospitalist: A Video Game Intervention to Increase Advance Care Planning by Hospitalists [Efficacy Trial]

NCT 04557930
11 November 2021

Statistical Analysis Plan

Variable Specification

Variable type	Variable Name		Variable Format	Notes
Primary Analysis				
dependent variable	primary	acpbilled	dichotomous (0/1)	
	secondary	mortality	dichotomous (0/1)	
		length of stay	continuous	
		7-day readmission	dichotomous (0/1)	
		30-day readmission	dichotomous (0/1)	
		ICU usage	dichotomous (0/1)	
primary predictor	phase of trial at hospital		categorical [0 = pre; 1 = enrollmentand post]	We will retain 3-level coding until the point of the analysis (pre 0; enrollment 0.5; post 1)
time	date		categorical with days since start of baseline period ascategorical labels	Will capture an anticipated non-linear and differential impact of COVID across time
	block		categorical	Further captures the COVID effect by adjusting for differential time-trends by block. Will allow for heterogeneity as these were pre-formed and control for the same general region of the country.
patient covariates	patient age		categorical variable (65-74, 75-84, 85+)	Association between age and probability of ACP billing is continuous with no inflections at specific ages (based on graphical analysis). We considered using the continuous variable but thought that the categorical variable provides more clinical context (e.g., an OR of 1.02 per year vs. OR 1.32 for patients age 74-85 and OR 1.76 for patients with age 85+)
	Charlson comorbidity index		categorical (0 to ≥5)	

	serious illness (selected based on differences in cohort between pre/post trial period)	dichotomous (0/1)	Serious illness codes based on Kelley et al. "Identifying older adults with serious illness" 2019 J Pain Symptom Management
	illness severity (evidence of organ failure)	exclude for endogeneity	It may be difficult to parse the differences between organ failure and receipt of LST (for example - codes for respiratory failure may be equivalent to the receipt of mechanical ventilation in some institutions). Since we anticipate that stepped-wedge design takes care of differences in cohort, we exclude these codes.
	surprise question	dichotomous (0/1)	Include if available by the time we are ready to submit. Otherwise exclude for missingness.
	COVID status	dichotomous (0/1)	Based on the presence of an ICD10 diagnosis code or a flag created by the treating physician that the patient was confirmed positive.
hospital covariates	risk-adjusted ACP rate prior to the start of the trial.	continuous	The correlation between risk-adjusted ACP rate and change in ACP rate is 0.23 (and therefore the risk of collinearity is small); these variables capture distinct effects.
	change in ACP rate between Quarter 2 2019 and Quarter 1 2020	continuous	
	practice size	quartiles	
	region	categorical	
	COVID admissions – fraction of total admissions with COVID in the month of the date of admission	continuous	
	COVID prevalence - HRR-level COVID rate in the month of the date of admission	continuous	We will test for collinearity with COVID admissions and exclude based on the variance inflation factor
random effect	hospital	categorical	
Secondary Analysis (Mediation)			
mediator	completion	dichotomous (0/1)	Whether physician finished the Game
	adherence (minutes)	quartiles	Minutes playing the Game

	engagement	quartiles	Physician experience with the game as captured by the Narrative Engagement Scale (scored out of 20)
	change in attitudes	difference in median attitude score before and after exposure	We control for baseline score to separate level effects from change- variable effects
random effect	physician, hospital	categorical	
Secondary Analysis (moderation)			
moderator (hospital and physician- level)	contextual effect – site cooperation rate	continuous	consider as covariate and interaction with MD adherence; remove interaction-effect if not-significant at the 0.05-level
	opportunity – proportion of physicians seen by patient who enrolled in the trial.	continuous	The total number of physicians will be considered as an alternative specification. One additional analysis will be a repeated measures model at the patient level where we define a separate outcome for each physician encountered by the patient until an ACP conversation occurred.
	physician CME compliance	dichotomous	

Note: In situations where the effects of ordinal categorical predictors (e.g. quartiles) follow a clear monotonic pattern, we will seek to simplify the model by considering use of the underlying continuous variable as a predictor. Such simplifications will only be implemented if the addition of the continuous predictor renders its categorical counterpart highly non-significant (e.g., p > 0.2). The advantage of simplifying the model is that wewill potentially be able to make a stronger interpretation of the results and moderation analyses will be more easily interpreted.

Primary Analysis

Let Y_{ijt} denote the binary outcome variable (coded as 1 if an ACP conversation occurred and 0 otherwise) for patient *i* seen at hospital *j* at time *t*; $Game_{jt}$ a binary variable indicating whether hospital j has received the Game during period t ($Game_{jt} = 1$ if received by hospital j before or during period t and 0 otherwise), x_{ijt} a vector of patient-level covariates, z_j a vector of hospital-level covariates and θ_j a random effect for hospital. The mathematical specification of the statistical model is given as $Y_{ijt}|\theta_i \sim Bernoulli(\pi_{ijt})$, where

$$logit(\pi_{ijt}) = log\left(\frac{\pi_{ijt}}{1-\pi_{ijt}}\right) = \beta_0 + \beta_{1t} + \beta_2 Game_{jt} + \beta_3 x_{ijt} + \beta_4 z_j + \theta_j$$

where $\theta_j \sim Normal(0, \tau^2)$ is the distribution of the hospital-level random effects to account for the fact that the statistical significance of inferences about the effect of the game are likely to be reduced by the clustering of patients in hospitals. The model includes fixed-effects for time-period, β_{1t}, to allow for an unstructured trend across calendar time, which makes the effect of the game (the primary target of inference) to be estimated net of any time-trend. The key coefficient of interest is β_2, which captures the structural shift in the outcome of patients who were enrolled in the study when the hospital receives the iPads, net of general trends across time and other covariates. Because this is a cluster-randomized study, there is a risk that the hospitals in each step are not perfectly balanced, despite attempts to balance these during randomization by forming blocks, and that the distributions of patient characteristics of patients treated by a given hospital may vary across time. To mitigate these concerns, we will adjust for judiciously selected patient and hospital covariates that we hypothesize are reasonably likely to be associated with the outcome. We do not plan to adjust for time-varying hospital-level covariates but we will adjust for whether the hospital was in other programs (e.g., the bundled payment care initiative (BPCI) program) that might influence the culture of the hospital towards ACP; an advantage of adjusting for BPCI participation is that we may obtain more precise inferences.

The reason why physician is excluded from the above model is that a patient may receive care from multiple physicians during their hospital stay. This makes it difficult to designate a single physician as being responsible for the patient's care and thus whether or not they receive an ACP conversation. In our primary analysis we hold the hospital as a collective unit as being responsible for the patient and, therefore, exclude any involvement of physician factors or identifiers in relation to the likelihood of the patient having an ACP conversation. However, based on analyses of preliminary data, we anticipate that for 80% of hospitalizations a single physician will dominate the care of the patient. Therefore, in a sensitivity analysis, we will add a physician layer to the above model and perform a physician-level analysis. Where more than one physician treats a patient, we will assign the patient to the discharging physician, as per the practice of the staffing organization. The resulting statistical model with be a three-level model with physician as the second level (between patient and hospital) to allow patients to be nested within physicians that are in turn nested within hospitals. Because patients are not randomized to physician, we will consider adjusting for physician covariates, emulating some of the secondary analyses described below.

Secondary analyses

In secondary analyses, we will also explore whether there is evidence on an interaction effect between BPCI participation and the impact of the game on the adjusted odds that a patient has an ACP billed. We will also estimate the effect of the intervention on ACP practices, using both the chart review and the MiPS measures to estimate the sensitivity and specificity of the different methods of measuring ACP. Finally, we will test the effect of mediators on the effect of the intervention on practice patterns, including the dose of a patient's exposure to the intervention, physicians' self-reported engagement with the intervention, and physicians' prior training. A natural game exposure-dose is the number of physicians, encountered by the patient, who had played the game by the time they cared for the patient. The game-exposure measure will replace the hospital-level indicator of game intervention status as the key predictor in these analyses. In analyses in which a single physician is attributed to the patient, the indicator of whether or not that physician has played the game will become the primary predictor of interest, although we may still include other exposure variables in order to extract the independent effect of each source of exposure.

The above factors are potential mediators of the effect of the game being employed at a hospital on patient outcomes as they are on the causal pathway of the hospital-level intervention to patient outcomes; if no physicians who indicated their willingness to participate in the study end up playing the game it is difficult to imagine how the game could then impact their patients' outcomes. Likewise, the hypothesis that a patient who encounters multiple physicians who played the game will have outcomes that are more pronounced than a patient who encountered only a single physician or even no physicians who played the game a priori appears to be plausible.

In a potential extended analysis, we will adapt statistical methods for incorporating the sensitivity and specificity of the measurement of the occurrence of an ACP conversation, which is informed by the agreement between chart-review and insurance-claim (or MiPS) measurement, into the analysis. The resulting analysis can be viewed as a calibration analysis that combines the standard cluster-randomized stepped-wedge design with a bivariate outcome (a more expensive measurement in the form of chart-review and a less expensive measurement in the form of insurance-claim or MiPS) in order to evaluate the impact of the deployment of the game at a hospital on chart-based measurement of ACP occurrence. The statistical model entwining the outcomes will allow the missing values of chart-based measurement for those observations where charts are not reviewed to be learned from observations for which multiple forms of ACP measurement are made and automatically allow for uncertainty in the missing values of chart-review measurements to permeate through the analysis. A Bayesian statistical model and Bayesian computational methods may provide the least burdensome pathway to successfully implementing this analysis.

Power calculation

We arrived at our sample size using a combination of feasibility (cost) and assumptions regarding effect size, absent any pilot data about the latter. For each step, we plan to recruit between 25 to 30 physicians from each of 4 to 8 hospitals. Assuming a baseline ACP rate of 22% (rising by 1.5 percentage-points per-quarter), a hospital intra-class correlation (ICC) coefficient of 0.01-0.10, and 160 evaluable patients per physician-quarter, we can detect a 3.5

percentage-point difference between ACP practices before and after the distribution of the intervention using a two-sided test at the 0.05-level with power in excess of 99%, even under the most conservative sample-size assumptions. If we invert the problem to find the smallest effect-size at which our study has 80% power, we find that in the most conservative scenario (76,800 total patients) we can detect a 1.5 percentage-point difference and in the most optimistic scenario (192,000 total patients), we can detect a 1 percentage-point increase.

The method of computing power for this stepped-wedge design follows the commonly used strategy for cluster randomized trials of first determining the design-effect, which can be thought of as a measure of the inefficiency of the given design in comparison to a completely randomized design that is expressed in terms of a ratio of the sample-sizes needed to obtain equally precise estimates, and then applying conventional power calculations. The latter computes power for a two-population comparison using the effective-sample-sizes determined from the design-effect. We estimate the design-effect using the expression in Woertman et al (2013), that was clarified and illustrated in Hemming (2016). Because hospitals may induce correlations in the outcomes of patients who receive care from them, we perform illustrative power calculations that account for the net impact of clustering at the hospital-level. Based on our own prior research and published results of others, we decided that the ICC of hospital is highly likely to be in the range 0.01 to 0.10. The design-effects across the optimistic and pessimistic scenarios ranged between 2.88 and 3.14, implying that for all considered scenarios the stepped-wedge design is about 33% as efficient as a patient-level completely randomized design. The effective sample-sizes (ESS) per group ranged from 30,603 to 12,388 patients per group over the study period (the 5 steps and a baseline period).

The second part of the calculation is to determine the power of a two-group comparison of a binary outcome in the absence of clustering when the total sample-size per group equals the above values for the ESS. Because the sample-sizes are still reasonably large, an asymptotic normal approximation is well justified, especially at a baseline ACP rate of 22%. Because we generally err on the side of making conservative estimates about the level of information available (e.g., we may extend the baseline period in which can retrospectively acquire data to 3-months). Therefore, this approximate two-step calculation yields trustworthy estimates of power that, if anything, are expected to err on the side of being conservative.