Helping Improve Pediatric Patient Outcomes (HIPPO)

ClinicaTrials.gov ID: NCT03471221

Unique Protocol ID: 1R21HD091877

Statistical Analysis Plan

January 2, 2019

Updated to included Addenda: October 24, 2022

Statistical Analysis Plan for HIPPO Randomized Trial

January 2, 2019

Updated to included Addenda: October 24, 2022

Overview

Animal-assisted interventions are increasingly common but evidence about their safety and efficacy is limited. Pediatric oncology patients are an important population in which to study animal-assisted interventions. Children with cancer experience significant distress, and though they are often resilient, strategies to improve their symptoms, hospital experience, and health-related quality of life are needed. As part of the Helping Improve Pediatric Patient Outcomes (HIPPO) Study, we are conducting a randomized controlled trial to assess the efficacy and safety of therapy dog visits for children with cancer. Our specific objectives are as follows:

1) To assess the effect of therapy dog visits on psychosocial outcomes and satisfaction with care among pediatric oncology inpatients.
2) To determine whether therapy dog visits increase microbial levels on children's hands.

We hypothesize that therapy dog visits will reduce patient distress, lower treatment-related anxiety, increase happiness, and improve satisfaction with hospital care. We further hypothesize that therapy dog visits (including standard hand sanitization with alcohol-based gel) will not increase microbial levels on children's hands.

Trial registration: NCT03471221

Trial design

The study is an individually-randomized, two-arm, parallel-group, controlled trial to examine the effectiveness and safety of dog therapy visits versus usual care.

Participants

Participants will be recruited from among pediatric oncology patients at Seattle Children's Hospital (SCH). The full set of eligibility criteria are defined in the Protocol (Section 3).

Trial arms

Patients will be randomized to either the intervention (therapy dog visits) or to control (usual care).

> **Intervention:** Participants randomized to the intervention arm will be offered a visit or visits from a therapy dog (goal: ~1x/week if dog team capacity allows) during their hospitalization and during any possible subsequent hospitalizations that occur during the study period (which is up to 4 weeks after enrollment), as long as they remain medically eligible.
>
> **Control:** Control arm patients will receive usual care.

The therapy dog visits will be observed and notes will be taken summarizing the visit (including which, if any, hands of the patients touched the dog during the visit).

Data, including questionnaires and hand sampling (to measure microbial load), will be collected in the same way from participants from both groups (See **Appendix 1** for timing of instruments).

Study outcomes

Primary outcomes

The primary Aim 1 outcome is the PedsQL™ Present Functioning Visual Analog Scales (VAS) [1] total score (which quantifies "distress"), measured immediately after the first visit (either intervention or control). This is a continuous measure ranging from 0 (lowest distress) to 100 (highest distress). It is computed as the average of 6 scales (afraid/scared, sad/blue, angry, worry, tired, pain/hurt). Higher scales represent higher distress. We hypothesize that patients randomized to the intervention arm will have lower levels of distress following a visit (either intervention or control) as compared to patients randomized to the control arm.

The primary Aim 2 outcome is the total microbial load (colony forming units per milliliter [cfu/mL], on the log10 scale as is common in the literature[2-4]), measured on each hand after the first visit (either intervention or control). We hypothesize that, after therapy dog visits (which include hand cleaning at the end of the visit), microbial levels on intervention patients' hands will not be meaningfully different than levels for usual care patients.

Secondary Aim 1 outcomes

- PedsQL Present Functioning Scales[1] emotional summary distress score [0-100], measured immediately after the first visit. Lower scores indicate a better outcome.
    - We hypothesize that patients randomized to the intervention arm will have lower levels of emotional distress than patients randomized to the control arm
- PedsQL Present Functioning Scales[1] individual scales (afraid/scared, sad/blue, angry, worry, tired, pain/hurt) [0-100], measured immediately after the first visit. Lower scores indicate a better outcome.
    - We hypothesize that patients randomized to the intervention arm will have lower levels of worry, tiredness, fear, sadness, and pain than patients randomized to the control arm (we do not hypothesize to see differences in the anger scale). We hypothesize that the largest changes will be in the worry and fatigue scales.
- PedsQL Cancer Module Treatment Anxiety [0-100], measured at discharge, follow-up 1 (3 days post-hospital discharge), and follow-up 2 (9-weeks post-hospital discharge). Higher scores indicate lower treatment anxiety.
    - We hypothesize that the intervention will reduce treatment anxiety, although we do not have a prior hypothesis on the specific timing at which this effect would be likely to occur.
- Patient's feeling when thinking about being in the hospital, measured at discharge, follow-up 1, and follow-up 2. Four scales: afraid/scared, sad/blue, angry; worried. Type: 4-point Likert scale for each.
    - We hypothesize that patients randomized to the intervention arm will have lower levels of fear, sadness, anger, and worry than patients randomized to the control arm. We do not have a prior hypothesis on the specific timing at which this effect would be likely to occur.
- Parent measure of child's positive affect measured by the Positive and Negative Affect Schedule (PANAS) [5-25], measured at discharge (primary), follow-up 1, and follow-up 2. Low scores represent worse outcomes.
    - We hypothesize that parents of patients randomized to the intervention arm will report a greater positive affect in their child than parents of patients randomized to the control arm.

- Parent measure of child's negative affect measured by the Positive and Negative Affect Schedule (PANAS) [5-25], measured at discharge (primary), follow-up 1, and follow-up 2. Low scores represent better outcomes.
    - We hypothesize that parents of patients randomized to the intervention arm will report a less negative affect in their child than parents of patients randomized to the control arm.
- Parent: Total score from six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI) [continuous, 20-80], measured immediately after the first visit. Higher scores represent worse outcomes.
    - We hypothesize that parents of patients randomized to the intervention arm will have higher anxiety levels than parents of patients randomized to the control arm.

      ADDENDUM – this text had a typo and should have said "lower"
- Parent: Hospital rating [0-10], measured at discharge (primary), follow-up 1, and follow-up 2. Lower scores represent worse hospital ratings.
    - We hypothesize that parents of patients randomized to the intervention arm will give the hospital a higher rating than parents of patients randomized to the control arm.

Secondary Aim 2 outcomes

- Presence of 9 distinct clinically important microorganisms (see section 9 of protocol for list), collected from hands sampled immediately after the first visit (see subsection below on "repeated visits")
    - We hypothesize that the presence of clinically important microorganisms will not differ in patients randomized to the intervention arm vs. control arm.

Exploratory Aim 1 and 2 outcomes

Other study outcomes are described in the Protocol and detailed below in the statistical analysis sections for Aims 1 and 2.

Sample size and power calculations

A sample size of 40 patients (20 in each group) provided adequate (≥80%) power to achieve our study aims. A potential sample size of up to 50 patients total was approved by the IRB.

For the effectiveness aim (**Aim 1**), with a sample size of 40 patients (20 in each group) and a type 1 error rate (two-sided) of 0.05, we will have ≥80% power to detect an absolute mean difference between groups in our primary outcome of 13.0, a difference smaller than the change of 14.2 observed in our pilot study at SCH.[5] Calculations[6-8] assumed normally distributed outcomes with standard deviation (SD) equal to 14.7, which was value obtained for the pre-visit scores in our pilot study.

For the safety aim (**Aim 2**), with a sample size of 40 patients (20 in each group), we will have ≥80% power to detect noninferiority with a margin of Δ=0.4 log10 cfu/mL when the SD of the outcome is 0.5 and with a margin of Δ=1.3 log10 cfu/mL when the SD is 1.4, for a 1-sided alpha level of 0.025, if there is truly no difference in the mean microbial load between the two groups. The minimal detectable noninferiority margin (Δ) for testing the difference in mean bacterial load between the intervention and control groups was calculated[9, 10] assuming normally distributed outcomes with a common SD in the two groups, where we selected values for the SD ranging from 0.5 to 1.4 log10 cfu/mL based on the published literature on hand cleaning methods in non-pediatric populations,[2-4] as we are not aware of microbial load values for pediatric populations in the literature.

Randomization

Participants will be assigned (1:1) using a computer-generated list of random numbers to the intervention and control groups. We will conduct a constrained randomization, in which we stratify participants by age (≥13 vs <13 years). Within the two strata, patients will be randomized in permuted (variable) blocks of 2, 4, and 6 patients.

Allocation concealment: The sequence of each block was generated prior to recruitment by the study biostatistician and will be concealed by the REDCap database.

Other design and statistical considerations

Eligibility to receive dog therapy visits

Among patients randomized to intervention, only those who are eligible to receive a dog therapy visit (e.g., not in isolation) on each day therapy dog sessions are offered a visit. Study staff will review patient medical records to determine whether they are eligible. Study staff will similarly review patient medical records of control patients at the first visit to determine if they would be eligible for an intervention visit (e.g., not in isolation) had they been randomized to the intervention group. Data will be collected even if patients (in either group) are not medically eligible for their first visit.

Repeated visits

Patients have variable lengths of stay in the hospital and some patients who are enrolled in the study may be readmitted at later dates. Analyses of primary outcomes will be applied to the first visit (either intervention or control) of a participant within the study.

Multiple testing

We are not planning to do multiple comparison adjustment for testing the primary outcomes for Aims 1 and 2, given that these are testing two separate scientific questions operating under two distinct mechanisms: effectiveness (Aim 1) and safety (Aim 2). Results of secondary analyses will be interpreted cautiously, in light of multiple comparison issues, whereby chance of finding a false positive is greater than the nominal level when multiple tests are conducted.

Intention-to-treat

Unless otherwise specified, analyses follow an intention-to-treat approach, in which individuals are analyzed by randomized group regardless of participation in any therapy dog visits.

Changes to methods after trial commencement

To increase study enrollment, we lowered the eligibility age to 5 years old. The primary outcome has been validated on children as young as 5 years old. The modification was approved on 8/14/2018.

ADDENDUM – The first five-year old was enrolled as the 11th participant in September 2018.

Descriptive Analyses

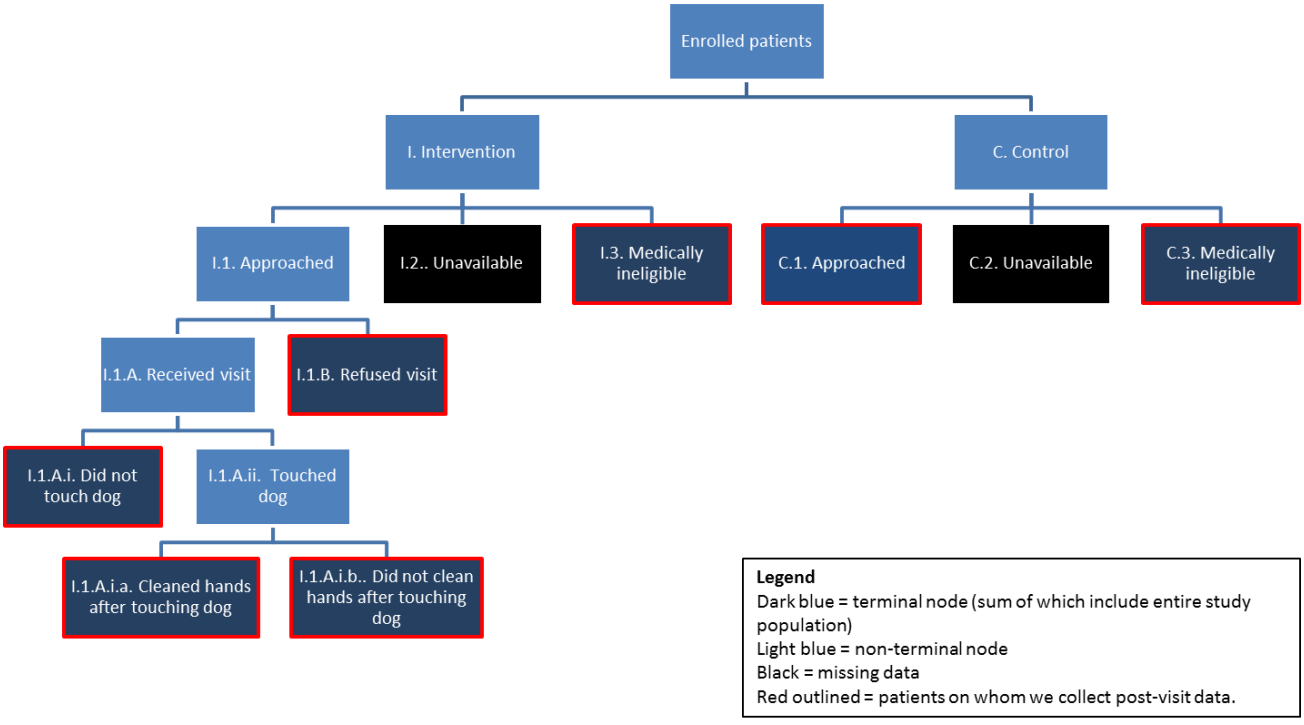
All data collected will be screened for outliers and nonsensical values. Descriptive statistics and graphical procedures will be used to inspect distributions to ensure that they meet assumptions of statistical tests and estimation procedures.

Descriptive statistics will be used to compare baseline covariates between intervention and control groups, though statistical tests comparing differences across these groups will not be conducted as recommended by CONSORT guidelines.[11]

We will describe the study flow via a CONSORT diagram, which will characterize relevant features of the study population, including the number of patients consented, randomized to each group, eligible for a first visit, and who received a first visit. Among those who received a first visit, we will further describe the number for which the first visit included a therapy dog or not and whether hands were cleaned after touching the dog (see Figure below for different groups that will be characterized).

Among individuals randomized to the intervention group, we will describe the distribution of the number of dog therapy visits received, as well as frequency of dog therapy visits per week.

> ADDENDUM – descriptive analyses on the frequency of dog therapy visits per week were not conducted due to COVID-19 delays, which resulted in too little time to complete all analyses.
>
> ADDENDUM – the following flow diagram was not used; based on discussion with the study team, the flow diagram was modified to more clearly present participant flow.



For Aim 2, in which microbial load will be measured on participants' hands, although we are planning to sample both hands after the visit, there is a chance we will be unable to do so (e.g., if a patient is called away mid-visit). Descriptive analyses will compare the proportion of patients, overall and by treatment group, who had 0, 1, or 2 hands sampled immediately following their first visit (intervention or control). In addition, for each patient with a therapy dog at their first visit, we will characterize the number of

hands (0, 1, or 2) that touched the dog, and of these, the number that touched the dog with or without subsequent hand cleaning.

Aim 1 statistical analyses

Primary Aim 1 outcome analysis

We will compare distress (quantified by the PedsQL total score) immediately following the first visit in patients randomized to the intervention group as compared to patients randomized to the control group. We plan to fit a linear regression model that regresses the PedsQL total score on an indicator for assigned treatment group, and that adjusts for the stratification variable (age ≥13 vs <13 years) and the baseline (pre-randomization) value of the outcome,

$$E(Y_i) = \beta_0 + \beta_1 int_i + \beta_2 age_i + \beta_3 baseline_i$$

where Y_i is the PedsQL total score immediate following the first visit for person *i*, int_i is the assigned intervention group, age_i is an indicator variable for whether the person was 13 years or older at baseline, and $baseline_i$ is the pre-randomization value of the outcome. We will test the null hypothesis that $\beta_1 = 0$ by using a two-sided Wald test with a type 1 error rate of 0.05. Wald 95% confidence intervals for $\beta_1$will also be constructed.

Secondary analysis of primary Aim 1 outcome

We plan to conduct a per-protocol analysis, in which we compare distress immediately following the first visit in patients who receive a dog visit to distress immediately following the first visit among controls who would have been medically eligible for a visit.

Secondary Aim 1 outcome analyses

Other PedsQL measures following the first visit

We will apply the same modeling approach as for the primary outcome analysis for the PedsQL emotional distress summary score, as well as for each of the PedsQL individual items.

PedsQL cancer module treatment anxiety

We will evaluate whether the intervention reduces treatment anxiety by fitting a repeated measures linear regression model of the PedsQL cancer module treatment anxiety measure. We will model the outcome at all time points simultaneously and include interaction terms between the randomization group and indicators for the time point (discharge, follow-up 1, follow-up 2). Specifically, our model will be of the form

$$E(Y_{it}) = \beta_0 + \beta_1^d int_i \cdot I(t = discharge) + \beta_1^{f1} int_i \cdot I(t = follow\ up\ 1) + \beta_1^{f2} int_i \cdot I(t = follow\ up\ 2) + \beta_2 age_i + \beta_3 baseline_i$$

where Y_{it} is the treatment anxiety score for person i at time point t, β_1^d corresponds to the intervention effect at the discharge time point, β_1^{f1} is the intervention effect at follow-up time point 1, and β_1^{f2} is the intervention effect at follow-up time point 2. To account for possible dependency of measures within an individual, we will fit the model using generalized estimating equations (Liang and Zeger 1986), assuming a working independent correlation structure; because our anticipated sample size is not large, we plan to use a jackknife estimate of the standard errors.[12] Separate tests will be conducted for each of the time-point specific intervention effect estimates (β_1^t).

ADDENDUM – Due to the high proportion of non-response at discharge and follow-up time points 1 and 2, outcomes were not modeled but were instead presented descriptively at each time point by study arm.

Patients' feeling when thinking about being in the hospital

Separate analyses for each of the four scales will be conducted, following a similar modeling approach as for the PedsQL cancer module treatment anxiety (described above). The main difference will be that we will use a proportional odds model since the outcomes are ordinal (ranging from 1-4). Additionally, the model will not adjust for the baseline measure of the outcome as this is not being collected, but will instead adjust for the baseline measure of the corresponding PedsQL scale.

ADDENDUM – Due to the high proportion of non-response at discharge and follow-up time points 1 and 2, outcomes were not modeled but were instead presented descriptively at each time point by study arm.

PANAS affect (positive/negative)

Separate analyses for each of the positive affect and negative affect scores will be conducted, following the same modeling approach as for the PedsQL cancer module treatment anxiety (described above).

ADDENDUM – Due to the high proportion of non-response at discharge and follow-up time points 1 and 2, outcomes were not modeled but were instead presented descriptively at each time point by study arm.

Spielberger State-Trait Anxiety Inventory (STAI)

Analysis will follow same approach as the primary outcome analysis.

Hospital rating (parent-report)

The analysis of hospital rating [0-10] will follow the same modeling approach as for the PedsQL cancer module treatment anxiety (described above).

ADDENDUM – Due to the high proportion of non-response at discharge and follow-up time points 1 and 2, outcomes were not modeled but were instead presented descriptively at each time point by study arm.

Exploratory Aim 1 analyses

Pre versus post analysis

We will describe the change in PedsQL Present Functioning scores comparing before versus immediately after the first visit among patients randomized to the intervention group who received a therapy dog visit. This will be done for the total distress score, the emotional summary score, and for each individual item. For this analysis, we will apply a linear regression for each outcome where we adjust for the stratification variable (age ≥13 vs <13 years) and the baseline (pre-randomization) value of the outcome. We hypothesize that total distress, emotional distress, and individual items except anger will be lower after the dog visit than before the visit.

Effect modification by age group

We will evaluate whether the ITT effect differs by age (≥13 vs <13 years) by expanding the regression model used for the primary Aim 1 outcome analysis to include an interaction term between age and the intervention group. We hypothesize that the effect will be greater among patients <13 years old.

PedsQL cancer module total score

Analysis of the PedsQL cancer module total score will apply the same approach as described above for the treatment anxiety sub-score. The only difference will be that we will adjust for age group in 3 categories (6-7, 8-12 and ≥13 years at randomization), as parents respond to the questionnaire for children ages 6-7 years. We hypothesize that patients randomized to the intervention arm will have lower scores than patients randomized to the control arm.

> ADDENDUM – Due to the high proportion of non-response at discharge and follow-up time points 1 and 2, outcomes were not modeled but were instead presented descriptively at each time point by study arm.

Dose-response analysis

Depending on the distribution of the number of dog therapy visits among patients assigned to the intervention group, we will consider conducting analyses, among patients in the intervention group who had at least 1 dog therapy visit, to examine if patients who have more dog therapy visits have better outcomes (measured at discharge and follow-up time points 1 and 2) than patients with fewer dog therapy visits.

Aim 2 statistical analyses

Primary Aim 2 outcome analyses

We will apply a noninferiority test, where our null hypothesis is that average microbial load in the intervention group exceeds that of the usual care group by pre-specified margin Δ, chosen to reflect a maximal "acceptable" difference. Levels of acceptable microbial load following hand sanitization in a pediatric inpatient population have not been characterized in the literature. Therefore, we plan to set Δ to be one SD of the outcome value at baseline (pre-randomization).

In safety aims, intent-to-treat (ITT) analyses may not be conservative, as patients who are randomized to the intervention arm but do not receive a dog therapy visit (and therefore do not have the potential for microbes to be transferred to their hands from the dog) are analyzed in the intervention group. We therefore plan to conduct both ITT and per protocol analyses and to reject the null-hypothesis of inferiority if the intervention of dog therapy visits is found to be noninferior under both analyses (Mauri and D'Agostino 2017).

For the **ITT analysis**, we will compare total microbial load levels (log10 transformed) on patients' hands immediately following the first visit in patients randomized to the intervention group as compared to patients randomized to the control group. For the **per-protocol analysis**, we will compare microbial levels immediately following the first visit in patients who receive a therapy dog visit (all hands will be included regardless of whether they touched the dog) to microbial levels immediately following the first visit in controls who would have been medically eligible for a visit; patients not meeting either of these two conditions will be excluded from the per-protocol analysis.

We will conduct a hand-level analysis (i.e., with hand as the unit of analysis), adjusting for possible correlation of repeated measures within individuals using generalized estimating equations (GEE).[13] We plan to fit the following linear regression model,

$$E(Y_{ij}) = \beta_0 + \beta_1 int_i + \beta_2 age_i + \beta_3 baseline_{ij}$$

where Y_{ij} is the log10 microbial load on hand *j* of patient *i,* int_i is the intervention group (either ITT or per-protocol), age_i is an indicator variable for whether the person was 13 years or older at baseline, and $baseline_{ij}$ is the pre-randomization value of the outcome for that hand. Because our anticipated

sample size is not large, we plan to use a jackknife estimate of the standard errors.[12] We will apply a 1-sided test of noninferiority with a type 1 error rate of 0.025 (i.e., we will reject the null hypothesis if the upper limit of the 2-sided 95% confidence interval for the coefficient β_1 from the GEE model is smaller than Δ, defined above).

Secondary analyses of primary Aim 2 outcome

Adjusting for additional covariates

As a sensitivity analysis, we plan to adjust for additional covariates: an indicator for whether the hand is the dominant hand and the participant's gender. It is hypothesized by the investigators that these variables may be associated with microbial load, though it is unknown whether they may be associated after adjusting for the baseline value of the outcome.

> ADDENDUM – Given the smaller than expected sample size additional covariate modeling was not done

Comparing hands of intervention patients that touched the dog to hands of control patients

As an additional analysis, we will compare the microbial load immediately following the first visit among hands of patients assigned to the intervention arm that received a therapy dog visit and that touched the dog to the microbial load immediately following the first visit among hands of patients assigned to the control arm who were medically eligible for a therapy dog visit. For this analysis, we will fit the same model as described above, restricted to the hands of patients that belong to one of these two comparator groups.

Secondary Aim 2 outcome analyses

Presence of clinically important microorganisms

For each of the 9 clinically important microorganisms, we will apply a similar approach as in the primary Aim 2 outcome analysis. In particular, we will fit a modified Poisson regression (Zou 2004) to estimate a relative risk (RR) where the outcome is an indicator variable for whether the microorganism is present on the hand, and the model will adjust for the intervention group (either ITT or per-protocol) and age. Evidence from simulation studies demonstrates poor performance of regression modeling of binary outcomes when there are few outcome events, especially with fewer than 5-10 events per predictor included in the model (Peduzzi et al. 1996; Vittinghoff and McCulloch 2006). Given that this outcome is expected to be rare (based on pilot [Phase 1] data collected prior to the trial), we therefore only plan to conduct a formal regression analysis of these outcomes if we observe at least 10 events (across both trial arms). Additionally, we do not plan to adjust for the baseline value of the outcome, in order to reduce the number of predictors included in the model.

The noninferiority margin for the RR (exponentiated coefficient on the intervention group indicator) will be selected to correspond to a 20% increase in the probability of the microorganism being present among intervention patients' hands versus control patients' hands (RR = 1.2). This margin was selected as a clinically meaningful difference, while recognizing that we may not have power to conclude noninferiority given the small sample size of the study. As above, both ITT and per-per protocol analyses will be conducted.

> ADDENDUM – Given the smaller the small number of outcome events, outcome rates were presented descriptively rather than modeled per the above analytic plan. Given that hand-level outcome rates are less interpretable, we instead presented outcome rates at a patient level (i.e., percentage of patients with a clinically meaningful microorganism present on either hand).

Exploratory Aim 2 analyses

Pre versus post analysis

We will describe the change in total microbial load comparing before versus immediately after the first visit among hands of patients randomized to the intervention group that touched the dog during the dog therapy visit. For this analysis, we will apply a similar modeling approach to the primary Aim 2 outcome, except the model will not include the intervention group indicator (since only intervention participants are included in this analysis). This exploratory analysis will be conducted only among participants who got a dog visit and only the hand(s) that touched the dog.

> ADDENDUM – this exploratory analysis was not conducted due to COVID-19 delays which resulted in too little time to complete all analyses.

Infections

We will describe, across treatment groups, the number of patients with a clinically meaningful infection, as well as the average number of infections. The analysis of the presence of an infection at 30 days post-study participation (where participation concludes at 4-weeks post-enrollment) will follow the same approach as for analyses of clinically important microorganisms (described above). The noninferiority margin will similarly correspond to a RR of 1.2 based on a clinically meaningful difference. Analyses are considered exploratory given concerns about not having adequate baseline measures of disease severity to adjust for potential chance imbalance across treatment groups given the small sample size of this study.

Missing data

We will investigate whether the proportion of patients with missing outcome information differs between the control and intervention arms. We will evaluate if any covariates are associated with missing outcome information. We will examine the missingness patterns of outcomes measured at multiple time points (discharge, follow-up 1, follow-up 2). In secondary analyses, we will include any baseline predictors of missing outcome information as baseline variables in the regression models above to remove measured bias due to loss to follow-up (Groenwold et al. 2011). If loss-to-follow-up is high (which we do not anticipate based on the low burden of this study), or if we find that post-baseline covariates significantly predict missing outcome information (particularly for those outcomes measured at multiple time points), we will consider using multiple imputation (MI) methods to increase our statistical power and further reduce bias (Little and Rubin 2014).

As described above, analyses will be adjusted for the baseline value of the outcome. Although we do not expect to have missing information on the baseline measures, if this occurs we plan to use single imputation for missing baseline measures (unless the analysis is using MI, in which case these measures will be imputed as part of the MI process). For continuous baseline measures we will impute the mean across all participants; for categorical or binary measures we will impute the mode.

Appendix 1. HIPPO Surveys by Age at Enrollment and Timepoint

Age at enrollment (years)	Baseline	Visit 1	Discharge and follow-up
6-7 [ADDENDUM: Changed to 5-7 in August 2018]	Child: - PedsQL Present Functioning VAS Parent (all): - PANAS-C-P - PedsQL Cancer - Parent Report for Young Children Parent (enrolled): - STAI	Child, pre: - PedsQL Present Functioning VAS Child, post: - PedsQL Present Functioning VAS Parent (enrolled), post: - STAI	Child (but sent to parent at f/u): - Questions about being in the hospital Parent (all): - PANAS-C-P - PedsQL Cancer - Parent Report for Young Children Parent (enrolled): - Modified CAHPS
8-12	Child: - PedsQL Present Functioning VAS - PedsQL Cancer – Child Report Parent (all): - PANAS-C-P Parent (enrolled): - STAI	Child, pre: - PedsQL Present Functioning VAS Child, post: - PedsQL Present Functioning VAS Parent (enrolled), post: - STAI	Child (but sent to parent at f/u): - Questions about being in the hospital - PedsQL Cancer – Child Report Parent (all): - PANAS-C-P Parent (enrolled): - Modified CAHPS
13-17	Child: - PedsQL Present Functioning VAS - PedsQL Cancer – Teen Report Parent (all): - PANAS-C-P Parent (enrolled): - STAI	Child, pre: - PedsQL Present Functioning VAS Child, post: - PedsQL Present Functioning VAS Parent, post: - STAI	Child (sent to child at f/u): - Questions about being in the hospital - PedsQL Cancer – Teen Report Parent (all): - PANAS-C-P Parent (enrolled): - Modified CAHPS

Abbreviations: CAHPS: Consumer Assessment of Healthcare Providers and Systems Child Hospital Survey hospital rating item; f/u: follow-up; PANAS-C-P: Positive and Negative Affect Schedule for Children-Parent Report; STAI: Spielberger State-Trait Anxiety Inventory state scale short form; VAS: Visual Analog Scales

References

1. Sherman, S.A., et al., *The PedsQL Present Functioning Visual Analogue Scales: preliminary reliability and validity.* Health Qual Life Outcomes, 2006. **4**: p. 75.
2. Howard, J.D., et al., *New method for assessing hand disinfection shows that pre-operative alcohol/chlorhexidine rub is as effective as a traditional surgical scrub.* J Hosp Infect, 2014. **88**(2): p. 78-83.
3. Kampf, G., et al., *Evaluation of two methods of determining the efficacies of two alcohol-based hand rubs for surgical hand antisepsis.* Appl Environ Microbiol, 2006. **72**(6): p. 3856-61.
4. Monistrol, O., et al., *Hand contamination during routine care in medical wards: the role of hand hygiene compliance.* J Med Microbiol, 2013. **62**(Pt 4): p. 623-9.
5. Chubak, J., et al., *Pilot Study of Therapy Dog Visits for Inpatient Youth With Cancer.* J Pediatr Oncol Nurs, 2017. **34**(5): p. 331-341.
6. Brown, H. and R. Prescott, *Repeated measures data*, in *Applied mixed models in medicine*. 2006, John Wiley & Sons, Ltd: Chichester. p. 231-288.
7. Liu, H. and T. Wu, *Sample size calculations and power analysis of time-averaged difference.* J Mod Appl Stat Methods, 2005. **4**(2): p. 434-45.
8. Diggle, P.J., K.Y. Liang, and S.L. Zeger, *Design considerations*, in *Analysis of longitudinal data*. 1994, Oxford University Press: New York, NY. p. 22-32.
9. Chow, S.C., J. Shao, and H. Wang, *Sample size calculations in clinical research*. 2003, New York, NY: Marcel Dekker.
10. Julious, S.A., *Sample sizes for clinical trials with normal data.* Stat Med, 2004. **23**(12): p. 1921-86.
11. Moher, D., et al., *CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials.* Bmj, 2010. **340**: p. c869.
12. Paik, M.C., *Repeated Measurement Analysis for Nonnormal Data in Small Samples.* Communications in Statistics - Simulation and Computation, 1988. **17**: p. 1155-1171.
13. Liang, K. and S.L. Zeger, *Longitudinal data analysis using generalized linear models.* Biometrika, 1986. **73**(1): p. 13-22.