Statistical Analysis Plan (SAP)

Handling of Missing Data

There are two types of missing data, at the respondent level (unit non-response) and at the specific question level (item non-response). At the respondent level, missing data occurred because of patient death, withdrawal from study, or inability to be reached via phone for an interview. For the SGRQ, values of 100 (worst QOL) were substituted for an assessment in which a patient was not available because of death. Comparisons of missingness of change in SGRQ score (from baseline to 6 months) by several patient characteristics were performed with Fisher's exact or Chi-Square tests to evaluate potential bias on the treatment effect. A sensitivity analysis was performed using a mixed-effects generalized linear model with all available SGRQ data (baseline, 3 months and 6 months post-discharge) for each patient. This analysis compared the baseline to 6-month change with robust variance estimates, clustering by hospital unit, patient as random effect, and treatment arm as fixed effect.

 For counts of rehospitalizations and ED visits, visits within a measurement period when a death occurred were neither counted nor imputed. Visits for some patients took place at non-Hopkins sites. Information on these visits could not always be confirmed adequately for inclusion in the cumulative counts. To evaluate the potential effect of missing events, patients with no non-Hopkins COPD-related events were compared to patients with at least 1 non-Hopkins event by age, hospital unit, and prior hospitalization (within last year).

At the item level, efforts were made to ensure that no or very limited amounts of missing data existed on all data, and especially on primary study outcomes and patient characteristics from the baseline assessment considered as covariates for analysis. No data were missing for the covariates such as age, gender, home oxygen use, and prior hospitalization. Missing data for

30   other patient characteristics were minimal. Since they were not critical to the current analyses,

31   imputation was not considered.

32   In preparing the outcomes that were based on items as part of a scale or domain statistic, the

33   handling of missing items was consistent with the outcome's literature to determine if, and

34   how, the outcome could be calculated or if it was to be considered missing. For example, the

35   SGRQ scores were prepared in spreadsheet calculators, provided by the instrument's authors,

36   and accounted for any assignments of missing scores based on the availability of item

37   responses.

38

39   Statistical Analysis

40

41   Statistical summaries and distributions of patient characteristics were reviewed across hospital

42   enrollment units. Statistical summaries and distributions of patient and clinical characteristics

43   were also reviewed across study arms. Such periodic assessments of the data provided

44   opportunities to evaluate missing data and provided input for additional data review and

45   quality assurance. Randomization of the patients to the study arms expects that on average the

46   arms were balanced on characteristics that might affect the study results, such as socio-

47   demographic and health status. Thus, no statistical tests to compare the arms on patient

48   characteristics were performed (Assmann et al 2000).

49

50   Exploratory analyses were performed cross-sectionally at baseline, 1-week, 1, 3 and 6 months

51   post-discharge for each of the outcomes across the study arms. This provided an assessment of

52   the outcomes' distributions, missingness patterns and the need for additional data review and

53   quality assurance.

54

55   The main analysis approach consisted of performing unadjusted analyses of the treatment

56   effect under intention to treat (ITT), followed by analyses adjusted for at least hospital

57   enrollment unit. The type of general linear model analysis performed depended on the

58   distribution of the outcome variable, and included gaussian, logistic, poisson and negative

59   binomial with robust estimators of variance and accounting for within hospital-unit correlation.

60   The outcomes were either point-in-time, change from baseline, or cumulative counts. Analyses

61   of change score outcomes also included the outcome's baseline score as a covariate. With

62   respect to SGRQ change from baseline to 6 months, two types of treatment effect analyses

63   were performed. One analysis used a general linear model to evaluate change in SGRQ score

64   (1-100), adjusted by the baseline score. A second sensitivity analysis was based on a 3-category

65   outcome of SGRQ change – better: $\leq$ -4, worse: $\geq$ 4, or no change: > - 4 and < 4 . The

66   distribution of the three categories was compared between the arms with a Chi-square test.

67

68   Analyses of cumulative rehospitalizations and ED visits, separate or combined, were analyzed as

69   all-cause and COPD-related. Estimates of means and incidence rates with 95% confidence

70   intervals (CIs) for the two arms were calculated based on the analysis models. Adjusted models

71   included age category, home oxygen use, prior hospitalization and hospital unit. Distributions of

72   each type of cumulative count were plotted for the two arms and compared by Fisher's exact

73   test. A further evaluation of the treatment effect for each type of cumulative count, where

74   there are a large number of zeroes, was performed. The visit count was dichotomized to zero

75   visits and at least one visit, and unadjusted and adjusted logistic regressions were performed.

76   Between arm comparisons for questions or measures taken at one-time point, or number of

77   deaths were performed using Fisher's exact test, Chi-square test, t-test or Mann-Whitney test

78   as appropriate. Unadjusted survival analyses using Kaplan-Meier and log rank tests were

79   performed for 'time to death or first COPD-related hospitalization or ED visit'. Analyses of

80   caregiver outcomes were performed with nonparametric tests for correlated measures.

81   Statistical significance was considered for $p < 0.05$.

82   **References:**

83   Assmann SF, Pocock SJ, Enos LE, Kasten LE (2000), "Subgroup Analysis and Other (Mis)Uses of

84   Baseline Data in Clinical Trials", Lancet, 355, 1064-1069

85