All-cause mortality rate in patients with idiopathic pulmonary fibrosis. Implications for the design and execution of clinical trials.

This is the author’s manuscript

Original Citation:
All-cause mortality rate in patients with idiopathic pulmonary fibrosis. Implications for the design and execution of clinical trials. / King TE Jr;Albera C;Bradford WZ;Costabel U;du Bois RM;Leff JA;Nathan SD;Sahn SA;Valeyre D;Noble PW. - In: AMERICAN JOURNAL OF RESPIRATORY AND CRITICAL CARE MEDICINE. - ISSN 1073-449X. - 189:7(2014), pp. 825-831.

Availability:
This version is available http://hdl.handle.net/2318/156709 since 2016-11-14T16:42:16Z

Published version:
DOI:10.1164/rccm.201311-1951OC

Terms of use:
Open Access
Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

All-cause mortality rate in patients with idiopathic pulmonary fibrosis. Implications for the design and execution of clinical trials.

King TE Jr¹, Albera C, Bradford WZ, Costabel U, du Bois RM, Leff JA, Nathan SD, Sahn SA, Valeyre D, Noble PW.

La versione definitiva è disponibile alla URL:

Rationale: FVC has emerged as a standard primary endpoint in clinical trials evaluating novel therapies for patients with idiopathic pulmonary fibrosis (IPF). However, it has recently been proposed that all-cause mortality or a composite comprised of all-cause mortality and all-cause nonelective hospitalization be adopted as the standard primary endpoint for IPF clinical trials.

Objectives: To conduct a comprehensive evaluation of mortality in three phase 3 clinical trials and evaluate the feasibility of mortality trials in patients with IPF.

Methods: The study population included 622 patients randomized to placebo in the CAPACITY studies evaluating pirfenidone (n = 347) or the INSPIRE study evaluating interferon-γ1b (n = 275). The Kaplan-Meier estimate of 2-year survival was fit to the exponential distribution and used to calculate sample size requirements for a mortality study with 90% power to detect a 25% reduction in all-cause mortality with a two-sided $\alpha$ of 0.05. Modeling analyses were used to assess the effects of selected variables on sample size and study design.

Measurements and Main Results: A total of 73 deaths occurred during the period of observation (mean duration of follow-up, 80.1 wk). The all-cause mortality rate was 6.6% at 1 year and 13.7% at 2 years. Based on the observed 2-year mortality rate, a total of 508 events would be required to detect a significant treatment benefit in a two-arm trial with 90% power to detect a 25% reduction in all-cause mortality. The estimated sample size for a trial enrolled over 3 years with a maximum follow-up period of 5 years is 2,582 patients.

Conclusions: The all-cause mortality rate is relatively low in patients with IPF with mild to moderate impairment in lung function. Accordingly, the necessary size, duration, and cost of all-cause mortality trials in this population are substantial and likely prohibitive.

KEYWORDS: idiopathic pulmonary fibrosis, mortality, clinical trial, hospitalization, endpoints
all-cause nonelective hospitalization should be adopted as the standard primary endpoint for therapeutic clinical trials in patients with IPF (6).

To explore the feasibility of mortality trials in the setting of a highly efficacious therapy, we conducted a comprehensive evaluation of mortality outcomes in three phase 3 global clinical trials in patients with IPF with mild to moderate impairment in baseline measures of physiologic function. These data were then used as the basis for power calculations and sample size estimates for a theoretical, randomized, double-blind, placebo-controlled trial designed to detect a statistically significant treatment effect on all-cause mortality in patients with IPF. This study has been presented in part at the 2013 International Conference of the American Thoracic Society (7).

**METHOD**

Study Population

All patients randomized to the placebo group in the CAPACITY studies (8) evaluating treatment with pirfenidone (n = 347) or the INSPIRE study (9) evaluating interferon-γ1b (n = 275) were included in the analysis.

Criteria for enrollment in the INSPIRE trial (GIPF-007) included a confident diagnosis of IPF according to protocol-defined criteria, evidence of clinical worsening within the preceding year, baseline percent predicted FVC greater than or equal to 55%, and baseline percent predicted diffusing capacity of carbon monoxide (DL_{CO}) greater than or equal to 35% (9). Eligibility criteria for the CAPACITY studies (Study 004 and Study 006) included a confident diagnosis of IPF according to protocol-defined criteria within the preceding year, baseline percent predicted FVC greater than or equal to 50%, baseline percent predicted DL_{CO} greater than or equal to 35%, and either baseline percent predicted FVC or percent predicted DL_{CO} less than or equal to 90% (8). Criteria mandating exclusion from all three trials included a history of clinically significant environmental exposure known to cause pulmonary fibrosis; unstable or deteriorating cardiac or pulmonary disease (other than IPF) within the previous 6 months; a diagnosis of connective tissue disease; clinical evidence of active infection; or any alternative explanation for interstitial lung disease. Eligible patients were also not permitted to be on a waiting list for lung transplantation at the time of randomization.

Statistical Analyses

A series of comprehensive analyses was undertaken to explore mortality outcomes in the pooled population of patients from the placebo groups in the INSPIRE and CAPACITY studies. Outcomes that were assessed included overall survival, cause of death, relatedness of death to the underlying respiratory disease, and the relationship between mortality and baseline lung function.

Demographic and baseline characteristics are summarized using descriptive statistics for continuous variables, and number and percentage counts of patients for categorical variables. Overall survival data are summarized as Kaplan-Meier estimates of the probability of survival in the pooled population. Cause of death was assigned by the principal investigator based on clinical judgment and coded according to system organ class as defined in the Medical Dictionary for Regulatory Activities. Additionally, deaths were characterized as IPF-related or not IPF-related by clinical investigators who were masked to treatment assignment.
To explore the feasibility of a clinical trial designed to assess the effect of an intervention on mortality in patients with IPF, the Kaplan-Meier estimate of 2-year survival was fit to the exponential distribution and used to calculate sample size requirements for a mortality study with 90% power to detect a 25% reduction in all-cause mortality (hazard ratio \([HR]\), 0.75) with a two-sided \(\alpha\) of 0.05. A modeling analysis was conducted to assess the effect of deaths that are either indirectly related or unrelated to IPF on the magnitude of effect on directly related deaths required to achieve a HR of 0.75 for all-cause mortality under a range of assumptions regarding the effect of treatment on indirectly related deaths. The estimated proportion of indirectly related deaths in the model was based on clinical judgment and informed by the proportion of deaths assessed by clinical investigators as either IPF-related or unrelated to IPF in previous phase 3 clinical trials (8, 9). Directly related death was defined as death caused by progression of disease or acute exacerbation; indirectly related death was defined as death caused by a respiratory or nonrespiratory cause or event in which IPF contributed indirectly but in a clinically significant way (e.g., pneumonia). Study sample size, randomization (1:1), enrollment, and duration of follow-up were held constant in the model.

**RESULTS**

**Patient Characteristics**

A total of 622 patients were included in the analysis. The median (range) age at baseline was 67 (40–80) years; median (range) values for percent predicted FVC and percent predicted \(Dl_{\text{CO}}\) were 71.7 (48.0–135.5) and 45.6 (17.1–90.1), respectively (Table 1). High-resolution computed tomography findings were interpreted by radiologists at the study sites as consistent with the protocol-defined criteria for a diagnosis of “definite IPF” in 89.5% of patients. High-resolution computed tomography findings were interpreted as consistent with the protocol-defined criteria for a diagnosis of “probable IPF” in 10.5% of patients; in accordance with the protocol, the diagnosis was subsequently confirmed by surgical lung biopsy in all such patients.

<table>
<thead>
<tr>
<th>tic</th>
<th>Value* ((n = 622))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, yr</td>
<td>67.0 (40–80)</td>
</tr>
<tr>
<td>Male, %</td>
<td>70.6</td>
</tr>
<tr>
<td>White, %</td>
<td>96.3</td>
</tr>
<tr>
<td>FVC, % predicted</td>
<td>71.7 (48.0–135.5)</td>
</tr>
<tr>
<td>(Dl_{\text{CO}}), % predicted</td>
<td>45.6 (17.1–90.1)</td>
</tr>
<tr>
<td>6MWT distance, m</td>
<td>400.0 (42.0–750.0)</td>
</tr>
<tr>
<td>Time since IPF diagnosis, yr</td>
<td>0.7 (0–5.0)</td>
</tr>
<tr>
<td>HRCT “definite IPF,” %</td>
<td>89.5</td>
</tr>
<tr>
<td>Supplemental O(_2) use, %</td>
<td>18.5</td>
</tr>
</tbody>
</table>

*Definition of abbreviations: 6MWT = 6-minute-walk test; \(Dl_{\text{CO}}\) = carbon monoxide diffusing capacity; HRCT = high-resolution computed tomography; IPF = idiopathic pulmonary fibrosis.

*Values are expressed as the median (range) unless otherwise indicated.

Demographics and Baseline Characteristics.
Mortality Outcomes

Vital status was ascertained in 612 of 622 (98.4%) patients. A total of 73 deaths occurred during the observation period (mean duration of follow-up, 80.1 wk); of these, 61 (83%) were judged by clinical investigators to be IPF-related. The all-cause mortality rate was 6.6% at 1 year and 13.7% at 2 years (Figure 1).

Figure 1

![Kaplan-Meier estimate of overall survival in the pooled placebo populations from the INSPIRE and CAPACITY studies.](image)

**Figure 1.** Kaplan-Meier estimate of overall survival in the pooled placebo populations from the INSPIRE and CAPACITY studies.

The absolute risk of mortality by baseline physiologic impairment is summarized in Figure 2. Consistent with prior research (10–12), graded relationships were observed between baseline percent predicted FVC and the crude risk of both all-cause and IPF-related mortality. Death from any cause occurred in 21% of patients with a baseline percent predicted FVC value less than 65%, compared with 14.7% in patients with baseline percent predicted FVC values between 65 and 80%, and 6.1% in patients with baseline FVC values greater than or equal to 80%.

Figure 2
Figure 2. Risk of mortality by baseline physiologic impairment. *Assessed by clinical investigators who remained blinded to treatment assignment. IPF = idiopathic pulmonary fibrosis.

Risk of mortality by baseline physiologic impairment. [More]

Cause of death is summarized in Table 2. The primary cause of death was reported as respiratory, thoracic, and mediastinal disorders in 47 (7.6% of the total study population) patients. Cardiac disorders (7 [1.1%] patients), general disorders (7 [1.1%] patients), and infections and infestations (6 [1.0%] patients) were the only other causes of death with an incidence greater than or equal to 1%.

Table 2: Cause of Death

<table>
<thead>
<tr>
<th>System Organ Class*</th>
<th>N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respiratory, thoracic, and mediastinal disorders</td>
<td>47 (7.6)</td>
</tr>
<tr>
<td>Cardiac disorders</td>
<td>7 (1.1)</td>
</tr>
<tr>
<td>General disorders/administration site conditions</td>
<td>7 (1.1)</td>
</tr>
<tr>
<td>Infections and infestations</td>
<td>6 (1.0)</td>
</tr>
<tr>
<td>Neoplasms, benign, malignant, and unspecified</td>
<td>5 (0.7)</td>
</tr>
<tr>
<td>Gastrointestinal disorders</td>
<td>1 (0.2)</td>
</tr>
</tbody>
</table>

*As defined in the Medical Dictionary for Regulatory Activities.

1Includes two cases of “disease progression,” one case of “condition aggravated,” and one case of “sudden death.”

Cause of Death.
Sample Size Calculation: IPF Mortality Trial

The observed 2-year mortality rate of 13.7% was used to calculate the estimated sample size for a two-arm study (1:1 randomization) with 90% power to detect a 25% reduction in all-cause mortality (HR, 0.75). Assuming vital status is ascertained in greater than or equal to 95% of study participants, a total of 508 all-cause mortality events would be required to reject the null
hypothesis. Based on an enrollment period of 3 years and a maximum follow-up period of 5 years, such a trial would require a sample size of 2,582 patients (Figure 3A). A substantially larger sample size would be required to detect a more modest treatment effect; based on the same assumptions regarding study power (90%) and 2-year mortality rates (13.7%), a sample size of more than 17,000 patients would be required to detect a 10% reduction (HR, 0.90) in all-cause mortality (Figure 3B).

**Figure 3.**

Power calculations and sample size estimates for mortality trials in idiopathic pulmonary fibrosis. (A) Sample size requirements for a trial with a 3-year enrollment period, 2 years of follow-up, and a hazard ratio of 0.75 (log-rank test for survival in two groups followed for fixed time with a fixed hazard ratio). (B) Sample size requirements for a two-arm trial with a 3-year enrollment period, 2 years of follow-up, and 90% power to detect a treatment effect with a two-sided α of 0.05 (log-rank test for survival in two groups followed for fixed time with fixed power and a constant hazard ratio over time).

Power calculations and sample size estimates for mortality trials in idiopathic pulmonary fibrosis.

To examine the effect of deaths that are not directly related to the underlying disease biology on the required magnitude of treatment effect on deaths that are presumably amenable to
antifibrotic therapy, HRs for directly related deaths were calculated for a range of scenarios (Table 3). Calculations were based on an overall 25% reduction in all-cause mortality among patients randomized to active treatment. If treatment reduced indirectly related deaths by 30% (HR, 0.70), a similar reduction in directly related deaths (HR, 0.70) would be required to achieve a 25% reduction in all-cause mortality. Under a scenario in which treatment has no effect on indirectly related deaths, a nearly 50% reduction in directly related deaths would be required to achieve a 25% reduction in all-cause mortality (HR, 0.75).

DISCUSSION

Recently, an expert working group convened to deliberate on the selection of clinically meaningful endpoints in phase 3 clinical trials in IPF concluded that all-cause mortality and all-cause nonelective hospitalization represent the most scientifically robust and clinically meaningful primary endpoints for therapeutic trials in patients with IPF (6). Based on this conclusion, the group proposed that, in the absence of a valid and reliable surrogate for mortality, a composite endpoint comprised of all-cause mortality and all-cause nonelective hospitalization should be adopted as the standard primary endpoint for clinical trials in patients with IPF. Although all-cause mortality inarguably satisfies the criterion of a scientifically rigorous and clinically meaningful endpoint in patients with a fatal disease, the practicality and cost of executing a properly powered mortality trial in patients with IPF have not been rigorously examined. In the present analysis, we conducted a comprehensive evaluation of mortality in a large and well-defined cohort of patients with IPF with mild to moderate impairment in lung function who were randomized to treatment with placebo in one of three large phase 3 clinical trials and followed prospectively for a mean duration of 80 weeks.

Our findings show that the all-cause mortality rate is relatively low in patients with IPF with mild to moderate impairment in baseline measures of physiologic function and, importantly, approximately one in five die from causes that are unrelated to IPF. Accordingly, the necessary size, duration, and cost of all-cause mortality trials in this population are substantial and likely prohibitive. Based on the results of our analysis, we estimate that a properly powered trial (1-β = 0.9) designed to detect a 25% reduction in all-cause mortality among patients with IPF with mild to moderate physiologic impairment would require a sample size of nearly 2,600 patients and cost approximately $250 million (13, 14). Moreover, the estimated time from study initiation to regulatory approval of the study drug, assuming the study results provide statistically persuasive evidence of a meaningful therapeutic benefit, would be nearly 7 years.

It is tempting to speculate based on the observed graded relationship between baseline lung function and mortality that the challenges associated with the design and execution of mortality trials in patients with IPF with mild to moderate physiologic impairment might be addressed in part by enrolling patients with moderate to severe functional impairment. Indeed, mortality studies in patients with advanced disease offer the theoretical advantages of higher event rates, smaller sample sizes, and shorter trial duration. However, the disease is not well characterized in this population and insufficient data exist to reliably inform estimates of sample size and study duration. Moreover, such comorbidities as pulmonary hypertension and end-stage irreversible fibrosis are markedly more prevalent in this population, further reducing the proportion of deaths that are likely to be amenable to antifibrotic therapy.

An additional challenge in the design of mortality trials in patients with IPF arises from the need to rely on mortality outcomes in underpowered studies to guide assumptions regarding the expected
magnitude of treatment effect. Importantly, assessments of mortality in underpowered studies are limited by the instability of the estimates of treatment effect and therefore lack sufficient precision to inform the design of phase 3 mortality trials. For example, assuming a 6% annual mortality rate, a two-arm trial (1:1 randomization) enrolling 500 patients would have 20% chance of observing a HR greater than 1.0 if truth = HR 0.75; if truth = HR 0.90 (a magnitude of effect that is comparable with the observed effect of several approved therapies for cardiovascular disease and non–small lung cancer) the chance of observing a HR greater than 1.00 would increase to nearly 40% (Figure 4).

Figure 4

Probability of type II error in trials underpowered to assess mortality, based on a two-arm trial with 1:1 randomization, 250 patients per group, and a 6% annual event rate in the placebo group. HR = hazard ratio.

Furthermore, the sample size requirement of 2,600 patients in our analysis is based on the assumption of a 25% reduction in all-cause mortality among patients randomized to treatment with the study drug, a magnitude of effect that is substantially larger than the observed effect of drugs that are commonly used to reduce the risk of mortality in other diseases. In a metaanalysis of 76 trials evaluating the effect of statins in 170,255 patients with a risk factor for a cardiovascular event, Mills and coworkers (15) reported a 10% reduction in all-cause mortality (relative risk, 0.90; 95% confidence interval, 0.86–0.94). This finding was consistent with four other metaanalyses evaluating the effect of statins on cardiovascular events; relative risks for all-cause mortality in these studies ranged from 0.84 to 0.89 (16–19). Additionally, in an analysis of randomized, double-blind, placebo-controlled trials evaluating molecular-targeted therapies for stage IIB–IV non–small cell lung cancer, Hotta and coworkers (20) identified 11 trials evaluating three currently approved agents for which overall survival was a primary or secondary endpoint. HRs for all-cause mortality ranged from 0.70 to 1.60, with only 1 of 11 identified trials reporting a HR less than 0.80 and 7 of the 11 trials reporting a HR greater than or equal to 0.90. Detecting a comparable magnitude of treatment effect on all-cause mortality (HR, 0.90) in a trial enrolling patients with IPF with mild to moderate functional impairment would require more than 17,000
patients and cost hundreds of millions of dollars. In the context of an orphan disease, such a trial is clearly impracticable.

Based on these observations, we believe that the adoption of all-cause mortality as a primary endpoint for IPF clinical trials, although scientifically rigorous and of indisputable clinical relevance, is both impractical and cost-prohibitive. For these reasons, we believe that objective measures of lung function and functional status represent the most appropriate measures of clinical efficacy for therapeutic clinical trials in IPF. Our conclusion is supported by several lines of evidence. First, both FVC and 6-minute walk test distance (6MWD) have been shown to be reliable, valid, and responsive measures of disease status (10, 19), and longitudinal changes in FVC and 6MWD are strong independent predictors of mortality in patients with IPF (11, 12, 21–25). Second, recent estimates for the minimal clinically important difference for FVC (10) and 6MWD (26–29) have established clear thresholds for evaluating the clinical relevance of an observed magnitude of treatment effect in IPF clinical trials. Third, analysis of outcomes in two large, randomized, double-blind, placebo-controlled trials in patients with IPF suggests that a novel composite for disease progression defined on the basis of categorical decrements in FVC and 6MWD increases event rates and has the potential to improve clinical trial efficiency (30). Finally, FVC has been established as a standard primary efficacy endpoint in phase 3 clinical trials in IPF (2–5), supporting the regulatory approval of an IPF therapy in several countries in Europe, Asia, North America, and Latin America (8).

The findings of our study should be interpreted in the context of several important limitations. First, the magnitude of treatment effect used in the sample size calculation and modeling exercise was arbitrarily selected, because there were insufficient data to reliably inform estimates of the magnitude of effect that can be reasonably anticipated in therapeutic trials in patients with IPF. Accordingly, the extent to which a 25% reduction in all-cause mortality represents an achievable outcome in IPF clinical trials is uncertain. However, as noted previously, relevant benchmarks in lung cancer and cardiovascular disease suggest that our assumption regarding the magnitude of treatment effect was comparatively large and therefore provided a conservatively small estimate of sample size (15–20).

Second, our assumptions regarding the estimated proportion of deaths in patients with IPF that are indirectly related to IPF were based on clinical judgment and the “relatedness” of death to the underlying fibrotic disease was not formally adjudicated. We note, however, that the chief aim of this particular analysis was limited to illustrating the relative impact of deaths that are not related to the underlying disease biology on the observed magnitude of treatment effect in a trial that is powered to assess the effect of an antifibrotic agent on all-cause mortality. Although the necessary magnitude of treatment effect on deaths that are directly related to IPF varies according to both the proportion of deaths that are not directly related to IPF and the degree to which such deaths are amenable to intervention with the study drug, the model illustrates an important concept that should be borne in mind when assessing the feasibility of all-cause mortality as a primary endpoint in IPF clinical trials.

Third, our analyses were limited to mortality outcomes and the feasibility of all-cause mortality as a standard primary endpoint in therapeutic clinical trials in patients with IPF. We did not evaluate the feasibility of the composite primary endpoint comprised of all-cause mortality and all-cause nonelective hospitalization that was recently proposed by Raghu and coworkers (6). Our decision was based on the limitations of all-cause nonelective hospitalization as an endpoint, including broad regional variability in hospitalization patterns and the possibility that the reason for hospitalization may be unrelated to the underlying disease biology. Additionally, unlike FVC and 6MWD, hospitalization has not been validated as an outcome in a clinical trial in patients with IPF.
Although we have shown that a clinical risk prediction model comprised of age, respiratory hospitalization, percent FVC, and 24-week change in percent FVC reliably predicts 1-year mortality in patients with IPF (12), the prognostic significance of all-cause nonelective hospitalizations has not been evaluated. Moreover, as noted by Raghu and coworkers (6), a correlation between an endpoint and mortality is insufficient evidence to demonstrate that the endpoint is a reliable surrogate for mortality. The observed correlation between respiratory hospitalization and mortality, although informative in the formulation of an individual patient’s prognosis, is insufficient to establish respiratory hospitalizations as a reliable surrogate for mortality and, in the absence of evidence to demonstrate the validity of either respiratory hospitalizations or all-cause nonelective hospitalizations as a clinical endpoint, the feasibility of primary endpoints based on hospitalizations in IPF clinical trials cannot be adequately evaluated.

Fourth, the power calculations and sample size estimates for a mortality trial in patients with mild to moderate physiologic impairment were based in part on the Kaplan-Meier 2-year survival estimates, with a planned trial enrollment period of 3 years and a follow-up period of 2 years. Although the stability of the estimates over 2 years suggests that mortality rates are unlikely to increase substantially among patients who remain in the trial beyond the first 2 years of follow-up, such a possibility cannot be excluded (31). Therefore, increasing the study duration could decrease the sample size requirements. We believe, however, that the challenges of executing trials of longer duration (including increased cost, decreased patient retention, and an increase in the prevalence of comorbidities over time) limit the feasibility of this approach.

Finally, our analysis was limited to patients with mild to moderate impairment in baseline measures of physiologic function; therefore, the degree to which conclusions regarding the feasibility of mortality trials can be extended to patients with severe physiologic impairment is uncertain. As noted previously, however, the design and execution of properly powered mortality trials in this population is confounded by the increased proportion of patients with end-stage fibrosis and comorbid conditions, and the paucity of data to guide estimates of sample size and study duration.

In summary, our results demonstrate that all-cause mortality rates are relatively low in patients with IPF and mild to moderate physiologic impairment. However, patients with IPF suffer real and meaningful consequences from their disease and therapies aimed at attenuating disease progression are urgently needed. Patients’ interests are therefore best served by efficient clinical trials of limited size and duration that compress drug development timelines and attract further investment in the field, thereby optimizing access to novel therapeutic interventions that slow disease progression and preserve patients’ quality of life.


2. Safety and efficacy of BIBF 1120 at high dose in idiopathic pulmonary fibrosis patients.

3. Safety and efficacy of BIBF 1120 at high dose in idiopathic pulmonary fibrosis patients II.

5. Efficacy and safety of pirfenidone in patients with idiopathic pulmonary fibrosis (IPF).


Correspondence and requests for reprints should be addressed to Talmadge E. King, Jr., M.D., Department of Medicine, University of California San Francisco (UCSF), 505 Parnassus Avenue, San Francisco