Re-Print: Developing and then Confirming a Hypothesis Based on a Chronology of Several Clinical Trials: A Bayesian Application to Pirfenidone Mortality Results

AUCTORES
Globalize your Research

Biomedical Research and Clinical Reviews

Authored by

Zhengning Lin[1*] and Donald A Berry[2]

[1]Ascendis Pharma, Palo Alto, CA, USA.

[2]Berry Consultants, Austin, TX, USA.

Published Date

September 15, 2021

Published in the Journal of

Biomedical Research and Clinical Reviews

Auctores Publishing, LLC

16192 Coastal Highway

Lewes, DE 19958,

United States of America

AUCTORES
Globalize your Research

Biomedical Research and Clinical Reviews

Zhengning Lin *

Open Access

Research Article

Re-Print: Developing and then Confirming a Hypothesis Based on a Chronology of Several Clinical Trials: A Bayesian Application to Pirfenidone Mortality Results

Zhengning Lin[1]* and Donald A Berry[2]

[1]Ascendis Pharma, Palo Alto, CA, USA.

[2]Berry Consultants, Austin, TX, USA.

***Corresponding Author:** Zhengning Lin, Ascendis Pharma Inc, 500 Emerson Street, Palo Alto, CA 94301, USA.

Received Date: 11 May 2021 | **Accepted Date:** 06 September 2021 | **Published Date:** 15 September 2021

Citation: Z Lin and Donald A Berry. (2021) Re-Print: Developing and then Confirming a Hypothesis Based on a Chronology of Several Clinical Trials: A Bayesian Application to Pirfenidone Mortality Results. *Biomedical Research and Clinical Reviews*. 4(5); DOI: 10.31579/2692-9406/066

Abstract

Designing a study for independent confirmation of a treatment effect is sometimes not practical due to required large sample size. Post hoc pooling of studies including those for learning purposes is subject to selection bias and therefore generally not suitable for confirmation of a treatment effect. We propose a Bayesian approach which calibrates the role of prior information from historical studies for learning and confirming purposes. The amount of prior information to be combined with current study data for the purpose of hypothesis confirmation depends on the overall strength of prior information for hypothesis generation. The method is illustrated in the analysis of mortality data for the pirfenidone NDA.

The Bayesian analysis provides a formal method to calibrate the role of information from historical evidence in the overall interpretation of results from both historical and concurrent clinical studies. The increased efficiency of using all available data is especially important in drug development for rare diseases with serious consequences, where limited patient source prohibits large trials, and unmet medical needs demand rapid access to treatment options.

Keywords: hypothesis generation; hypothesis confirmation; historical data; rare disease; learn and confirm; idiopathic pulmonary fibrosis

Background

In clinical drug development, early phase studies are designed for learning, for generating and testing hypotheses. Later phase studies are designed for confirmation of treatment effects for regulatory approval. The process of developing and confirming hypotheses applies to a collection of several studies as well as individual studies. Earlier confirmatory studies may generate refined or new hypotheses to be confirmed by later confirmatory studies, and the cycle can go on and on. The setting to confirm a hypothesis based on data exclusively from an individual study can be inefficient and sometimes not feasible in practice due to required large sample size, especially in the case of low event rate for a rare disease. Although data pooling from multiple studies can provide reasonable sample size for hypothesis confirmation, post hoc data pooling including those for hypothesis generation purposes is not scientifically solid, and pre-specification of data pooling without early learning is often unrealistic.

Bayesian statistics has a natural framework to incorporate prior information from earlier studies, for the purpose of evaluating treatment effect from new study data. We propose a Bayesian approach which calibrates the role of prior information from earlier studies for learning and confirming purposes. It formally discount historical information for the purpose of confirming a treatment effect in a prospectively designed study. This approach recognizes the hypothesis generation aspect of prior information while using the residual information for confirmation purposes with increased statistical efficiency. Learning is viewed as continuum rather than regarding "study" to be the learning unit. We illustrate the method in the analysis of mortality data for the pirfenidone NDA.

To help readers with different professions to link Bayesian posterior probabilities to the widely used p-values, we use the term "analogous" to describe comparable levels of statistical significance between the two approaches of statistics. For example, a posterior probability of 0.975 for treatment benefit is analogous to a one-sided p-value of 0.025 (or a two sided p-value of 0.05) in terms of statistical significance, which is a conventional cut point for statistical significance in the current regulatory environment. This linkage is important to compare the two approaches of statistics with comparable level of statistical significance, although the

meaning of posterior probabilities and p-values are quite different within each of the two approaches of statistics: "P = 0.025" is not interpreted as "the probability of alternative hypothesis is 0.975", while the same data can produce a posterior probability of 0.975 for treatment benefit with a "non-informative" prior.

The Pirfenidone NDA

The pirfenidone NDA includes a total of three placebo-controlled studies to demonstrate efficacy for idiopathic pulmonary fibrosis (IPF), a rare and ultimately fatal lung disease with no treatment in the US at the time of NDA. Studies PIPF-004 and PIPF-006 were conducted with a minimum of 72 weeks of double-blind placebo control, while Study PIPF-016 was a 52 week double-blind placebo controlled study started after completion of the early two studies. The primary endpoint is percent predicted FVC, although mortality is considered as the ultimate endpoint with the limitation of low statistical power to be the primary endpoint.

The results of clinical studies PIPF-004 and PIPF-006 suggested that the evident slowing of disease progression caused by pirfenidone might translate into lower mortality. Therefore, the prospective plan of the subsequent confirmative study PIPF-016 included 52-week all-cause mortality and treatment-emergent IPF-related mortality as secondary endpoints. However, PIPF-016 was not powered to detect clinically important effects on either type of mortality. Assuming a total of 31 deaths from any cause (as actually observed in the study overall—refer to Table 1) and an eventual log-rank test, a large treatment effect with 0.5 hazard ratio has only 49% power to detect a treatment difference. Assuming a total of 10 treatment-emergent IPF-related deaths, the study has only 19% power with the same hazard ratio assumption.

	PIPF-016			PIPF-004			PIPF-006		
Mortality	PIR (N=278) n (%)	PBO (N=277) n (%)	RR	PIR (N=174) n (%)	PBO (N=174) n (%)	RR	PIR (N=171) n (%)	PBO (N=173) n (%)	RR
All-cause	11 (4.0)	20 (7.2)	0.55	5 (2.9)	13 (7.5)	0.38	6 (3.5)	9 (5.2)	0.67
TE IPF-related	3 (1.1)	7 (2.5)	0.43	2 (1.1)	8 (4.6)	0.25	2 (1.2)	7 (4.0)	0.29

Note: Table reports the number of 52-week all-cause and TE IPF-related mortality events for PIR and PBO.

IPF = idiopathic pulmonary fibrosis; PBO = placebo; PIR = pirfenidone; RR = relative risk, of PIR to PBO;

TE = treatment-emergent.

Table 1: *Mortality data from Trials PIPF-016, PIPF-004 and PIPF-006 (All Randomized Patients)*

To achieve greater power, the protocol and statistical analysis plan of PIPF-016 indicate that the events in PIPF-016 will be pooled with those censored at one year in PIPF-004 and PIPF-006. Results from the pooled analyses provide reasonably convincing evidence for a positive conclusion, as shown in Table 2.

	All-Cause Mortality		TE IPF-Related Mortality	
	Pirfenidone 2403 mg/d (N = 623)	**Placebo (N = 624)**	**Pirfenidone 2403 mg/d (N = 623)**	**Placebo (N = 624)**
Patient death, n (%)	22 (3.5)	42 (6.7)	7 (1.1)	22 (3.5)
Hazard ratio[a] (95% CI)	0.52 (0.31,0.87)		0.32 (0.14,0.76)	
p-value[b]	0.0107		0.0061	

a Hazard ratio was based on the Cox proportional hazard model.

b p-value was based on the log-rank test.

Note: Table reports 52-week all-cause and TE IPF-related mortality data for PIR and PBO.

CI = confidence interval; IPF = idiopathic pulmonary fibrosis; PBO = placebo; PIR = pirfenidone; RR = relative risk, of PIR to PBO; TE = treatment-emergent.

Table 2: *Mortality Data from PIPF-016, PIPF-004, and PIPF-006 Pooled (All Randomized Patients)*

The consistency of the mortality results across the three trials as shown in Table 1 and the efficacy of pirfenidone in slowing the progression of IPF support a pooling strategy. However, there is a recognized limitation of the pooled mortality analysis because it was specified after results of the earlier trials were available, although before the start of PIPF-016. As a result patients in those two trials cannot be considered exchangeable with patients in trial PIPF-016 for the purpose of confirmation of treatment effect, as the earlier trials are partly hypothesis generating.

A standard analysis for discounting prior information is via a Bayesian statistical approach [1]. The results of these earlier trials are relevant for addressing the final question, but at less than their face value. Hence in the context of trial PIPF-016 they should not count fully [2, 3].

Pooling of Mortality Data

In view of the limited power for addressing mortality in Study PIPF-016, the statistical analysis plan (SAP) for PIPF-016 prospectively defines a pooling analysis with the mortality information from Studies PIPF-006 and PIPF-004 as a secondary analysis:

Mortality data from Study PIPF-016 also will be pooled with data from the pirfenidone 2403 mg/d and placebo groups from Studies PIPF-004 and PIPF-006. For the pooled analysis, the PIPF-004 and PIPF-006 results will be censored at Study Day 365 if an event has not occurred earlier in order to allow the three studies to contribute comparable follow-up times to the pooled analysis.

The mortality analyses using the log-rank test described in the SAP for PIPF-016 provide p-values consistently less than 0.05 when using full pooling as shown in Table 2, while results based on PIPF-016 alone have p = 0.1045 for all-cause mortality and p = 0.2258 for treatment emergent IPF-related mortality.

We carried out a Bayesian analysis that discounts previous Studies PIPF-004 and PIPF-006 but borrows some inferential strength from these studies in estimating the effect of pirfenidone on reduction of mortality as compared with placebo for the PIPF-016 study [4, 5].

Statistical Modeling

We use a Bayesian analysis to synthesize mortality results from Study PIPF-016 and the combination of Studies PIPF-004 and PIPF-006. The prospectively defined analysis for mortality endpoints in the statistical analysis plan for PIPF-016 was a time-to-event log-rank test of the hazard ratio. However, since the duration of follow-up is predetermined to be one year for all patients, we analyze the dichotomous outcomes of deaths within the first year. An advantage of using dichotomous outcomes (instead of time-to-event outcomes) is its simplicity of modeling with complete data transparency at each step of calculation, which is important for ease of communication of a complex concept to different professions. The approach to synthesize mortality data can be applied similarly to survival data with appropriate modeling.

Let the labels for Studies PIPF-004, PIPF-006, and PIPF-016 be s = 4, 6, and 16, respectively. In study s the number of subjects on placebo (PBO) is m_s and on pirfenidone (PIR) is n_s. In study s there are x_s deaths in the PBO group and y_s in the PIR group. We assume that the numbers of deaths within the PIR and PBO groups in study s are distributed as binomial:

$$x_s \sim Binomial(m_s, p_s);$$

$$y_s \sim Binomial(n_s, q_s). \qquad \text{for } s = 4, 6, \text{ and } 16.$$

In the Bayesian framework we can use the data from these historical studies to form a prior distribution on the mortality event rates for Study PIPF-016. The data from PIPF-016 can then be combined with the prior distribution formed from the historical study data to calculate the posterior distribution of the mortality event rates.

Historical Prior

We assume beta prior distributions on the mortality event rates in Study PIPF-016:

$$p_{16} \sim Beta(a_{PBO}, b_{PBO});$$
$$q_{16} \sim Beta(a_{PIR}, b_{PIR});$$

In particular, for both groups PBO and PIR we specify the borrowing of the historical data as a fraction borrowed parameter (θ) as:

$$a_{PBO} = q(x_4 + x_6) + a_0$$
$$b_{PBO} = q[(m_4 - x_4) + (m_6 - x_6)] + b_0$$

and

$$a_{PIR} = q(y_4 + y_6) + a_0$$
$$b_{PIR} = q[(n_4 - y_4) + (n_6 - y_6)] + b_0,$$

where θ is a number between 0 and 1 to reflect the amount of borrowing of information between Study PIPF-016 and historical Studies PIPF-004 and PIPF-006. If θ = 1, then the historical studies are pooled with Study PIPF-016, whereas if θ = 0, then the historical data are completely discounted. The original prior, before any of the three studies, for both PIR and PBO is assumed to be a uniform distribution, with $\alpha_0 = \beta_0 = 1$.

We use the symbol q for the death rate for PIR in Study PIPF-016 and p for the death rate on PBO, dropping the subscript 16 in both cases.

The prior distributions of q and p before Study PIPF-016 but after Studies PIPF-004 and PIPF-006 depend on θ. Figure 1 shows two special cases for all-cause mortality, one with θ = 1 and the other with θ = 0.50. In the case θ = 0, complete discounting of the earlier studies, both prior densities are uniform: equal to a constant for the whole interval from 0 to 1.

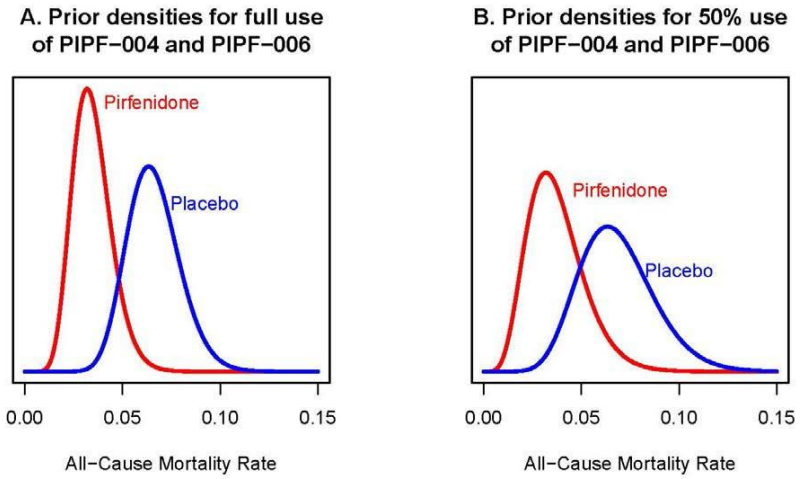

Figure 1: *The Prior Densities for Pirfenidone (PIF) and Placebo (PBO) Using Full Borrowing and 50% Borrowing*

Posterior Distribution: Updating Historical Prior with Study PIPF-016 Results

The posterior distribution of p and q given the results of PIPF-016 also has a beta distribution:

$$q|n_{16},y_{16} \sim Beta(y_{16}+a_{PIR},n_{16}-y_{16}+b_{PIR});$$
$$p|m_{16},x_{16} \sim Beta(x_{16}+a_{PBO},m_{16}-x_{16}+b_{PBO}).$$

Results

We provide results depending on θ, the amount of borrowing from PIPF-004 and PIPF-006. For each θ we draw 1 million samples from the posterior distributions of *p* and *q* and we report:

1) Posterior probability of superiority of PIR vs PBO (this is the proportion of samples where $q < p$)

2) Posterior mean of the relative risk (q/p)

3) 95% credible interval of the relative risk

Table 3 shows results for both all-cause mortality and treatment-emergent IPF-related mortality.

	No Borrowing		Full Borrowing		Tipping Point
	Bayesian Prob. of Superiority (analogous two-sided p-value)	Log-rank Reported p-value	Bayesian Prob. of Superiority (analogous two-sided p-value)	Log-rank Reported p-value	Borrowing needed to achieve 0.975 probability of superiority for pirfenidone
All-cause	0.951 (0.098)	0.1045	0.9947 (0.0106)	0.0107	29%
TE IPF-related	0.890 (0.220)	0.2258	0.9975 (0.0050)	0.0061	38%

IPF = idiopathic pulmonary fibrosis; TE = treatment-emergent.

Table 3: *Mortality Results from Bayesian Analysis*

All-cause Mortality

Under our Bayesian analysis and with no borrowing of information from PIPF-004 and PIPF-006, the posterior probability that pirfenidone is superior to placebo in terms of the all-cause mortality event rates is 0.951. This is analogous (in the sense of comparable statistical significance) to a one-sided p-value of 0.049 and a two-sided p-value of 0.098. In the other extreme, under full borrowing of information from PIPF-004 and PIPF-006, the posterior probability that PIR is superior to PBO in terms of all-cause mortality event rates is 0.9947. This is analogous to a one-sided p-value of 0.0053 and a two-sided p-value of 0.0106. This is very similar to the p-value under full pooling and the log-rank test on a time-to-event endpoint of 0.0107.

Figure 2 shows the probability of superiority for varying θ, reflecting a varying amount of borrowing from PIPF-004 and PIPF-006 as well as the estimated relative risks and 95% credible intervals for each. The figure shows that the "tipping point" where the probability of superiority is 0.975 (analogous to a one-sided p-value of 0.025) is θ = 0.29. So borrowing 29% or more of the mortality information from Studies PIPF-004 and PIPF-006 (which means discounting these two studies by 71% or less) gives statistical significance for all-cause mortality.

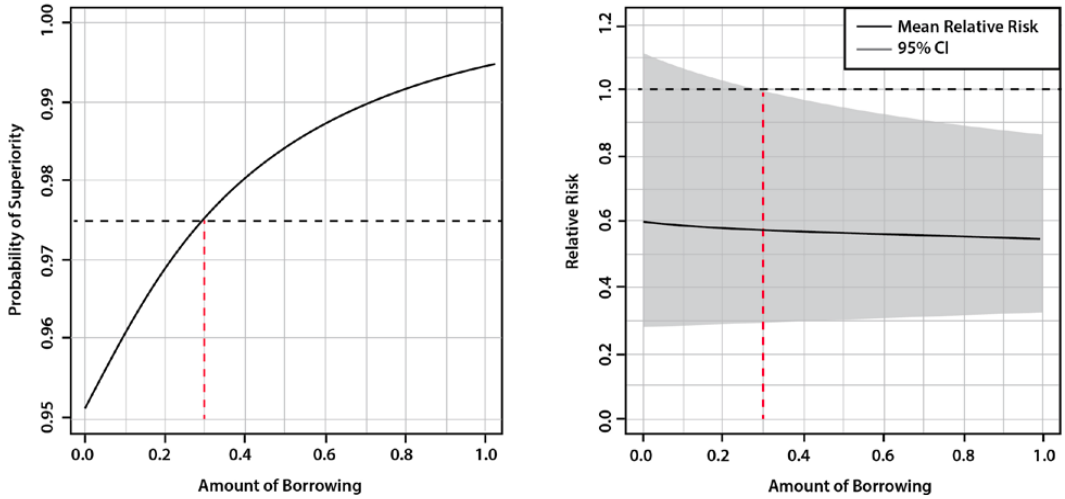

Figure 2: *All-cause Mortality*

Treatment-emergent IPF-related Mortality

Figure 3 shows similar results for event rates of treatment-emergent IPF-related mortality. In particular, under no borrowing of information from previous studies the posterior probability that PIR is superior to PBO is 0.89. At the other end of the scale, under full borrowing from Studies PIPF-004 and PIPF-006, the posterior probability that PIR is superior to PBO is 0.9975. The "tipping point" where the probability of superiority is 0.975 (analogous to a one-sided p-value of 0.025) is θ = 0.38. So borrowing 38% or more from Studies PIPF-004 and PIPF-006 (or discounting these two studies by 62% or less) gives statistical significance for treatment emergent IPF-related mortality.

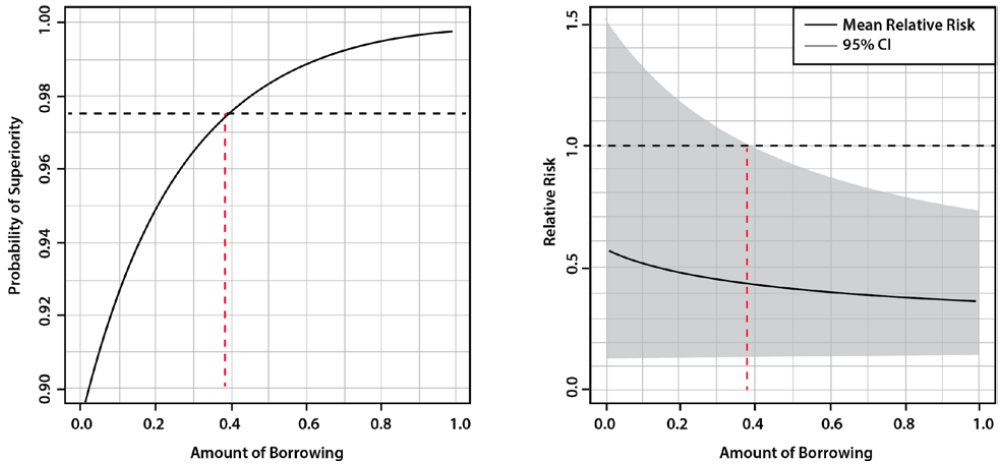

Figure 3: *Treatment-emergent IPF-related Mortality*

Calibrating the role of prior information

In view of limited power in assessing a possible reduction in mortality due to pirfenidone in comparison with placebo, the statistical analysis plan for Study PIPF-016 prospectively specified pooling the mortality results of PIPF-016 with those from two previous studies, PIPF-004 and PIPF-006. The mortality-related events in these previous studies was partially hypothesis generating. Our Bayesian analysis recognizes the hypothesis generating aspect of these earlier studies while using the residual information as a prior distribution for PIPF-016 by partially discounting the earlier studies.

Figure 4 shows this division into hypothesis generating and confirmation. The former is shown in Panel A, showing 50% of the information in Studies PIPF-004 and PIPF-006. In Figure 4A the probability that pirfenidone is superior to placebo is 91%, which provides substantial motivation to establish the hypothesis that pirfenidone reduces all-cause mortality. For the prior distribution in Panel B for assessing all-cause mortality in Study PIPF-016 the posterior probability of superiority calculated in Figure 2 is 98.4%. The corresponding calculation for treatment-emergent IPF-related mortality in Figure 3 again assuming 50% use of results from Studies PIPF-004 and PIPF-006 is also 98.4%. The analogous two-sided p-value is 0.032. In both cases the results provide ample evidence of confirmation.

Figure 5 is in the same format as Figure 4. It shows the analogous parts of the information on all-cause mortality from Studies PIPF-004 and PIPF-006 at the tipping point of 71% of information for hypothesis generation and confirmation.

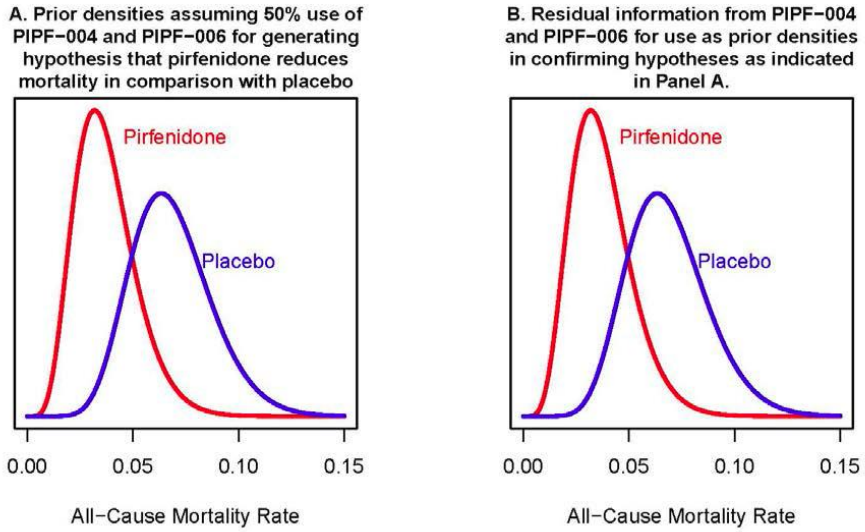

Figure 4: *These Two Panels Show the Posterior Densities of the Results from Studies PIPF-004 and PIPF-006 (refer to Figure 1A) Divided in Two, Half for Hypothesis Generating (Panel A) and the Other Half to Serve as the Prior Information for Study PIPF-016 in Confirming the Hypothesis (Panel B)*

The two graphs are identical to accentuate the equality of the information content in this division. In both panels the "numbers of deaths" are 5.5 out of 172.5 "patients" on PIR and 11 out of 173.5 "patients" on PBO. In Panel A, assuming a uniform distribution prior to studies PIPF-004 and PIPF-006, the probability that PIR is superior to PBO is 91%, which provides substantial motivation to establish the hypothesis that PIR reduces all-cause mortality. For the prior distribution in Panel B for assessing all-cause mortality in Study PIPF-016 the posterior probability of superiority calculated in Figure 2 is 98.4%. The corresponding calculation in Figure 3 again assuming 50% use of results from PIPF-004 and PIPF-006 is also 98.4%. PBO = placebo; PIR = pirfenidone.

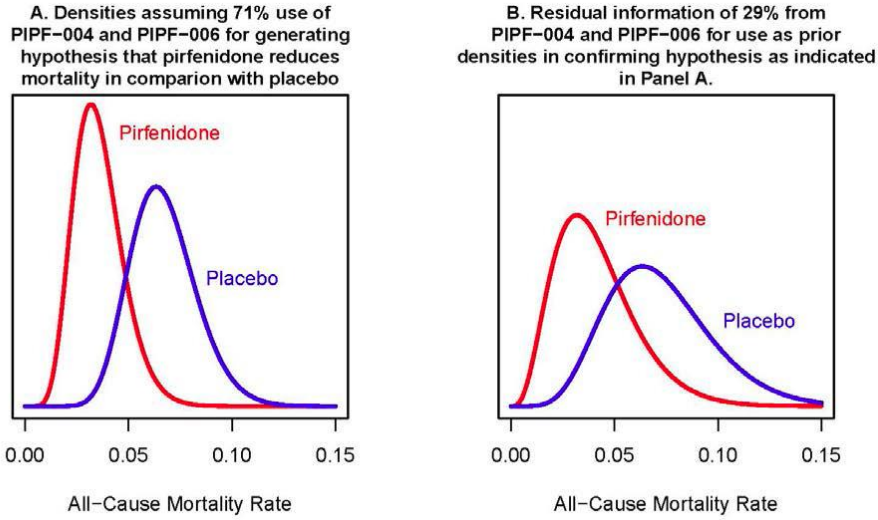

Figure 5: *This Figure Shows the Information Division between Hypothesis Generation and Hypothesis Confirmation at the "Tipping Point" Described in the Text*

The two panels show the posterior densities of the results from Studies PIPF-004 and PIPF-006 divided in two, 71% for hypothesis generating (Panel A) and the 29% to serve as the prior information for Study PIPF-016 in confirming the hypothesis (Panel B). In Panel A, assuming a uniform distribution prior to Studies PIPF-004 and PIPF-006, the probability that PIR is superior to placebo is 93.8%. Panel A contains more than twice as much information for hypothesis generation as Panel B does for confirmation. In Panel A the "numbers of deaths" are 7.81 out of 244.95 "patients" on PIR and 15.62 out of 246.37 "patients" on PBO. In Panel B the "numbers of deaths" are 3.19 out of 100.05 "patients" on PIR and 6.38 out of 100.63 "patients" on PBO.
PBO = placebo; PIR = pirfenidone.

An example of a Bayesian analysis using 50% borrowing from a previous study in a registration setting is Boston Scientific's WATCHMAN® Left Atrial Appendage Closure Therapy (FDA, 2013)[6].

The prior distributions in this Bayesian analysis are empirically based. Berry et al. [3], Berry [7] describe how to use other available information subjectively to improve the accuracy of Bayesian conclusions. For example, the effectiveness of pirfenidone in shifting the stage of IPF may be reasonably regarded to result in an end-stage shift, that is, a mortality reduction. Evidence for this possibility and other information can be incorporated into the prior distributions of this report using methods described in these references.

In summary, a helpful feature of the Bayesian analysis described above is that it provides a way to calibrate the role of the information from the earlier studies in the overall interpretation of the results from all studies. The range of this calibration includes no use of the information from the previous studies at one end and full use of the previous studies in a pooled analysis at the opposite end. The middle ground with respect to the calibration provides a reasonably convincing basis for a positive conclusion with respect to the totality of information from all three studies. Discussions to determine an appropriate amount to borrow from previous studies are included in the following section.

Discussion

Borrowing information

Borrowing information from prior studies to confirm treatment effect becomes increasingly important in drug development, especially in the field of rare disease, with opportunities of increased efficiency of delivering effective treatments to patients. In many cases, combining information from multiple studies is the only way practical to confirm treatment effect, like the case of mortality data for pirfenidone[8]. The Bayesian mortality analysis for pirfenidone illustrated how information from prior studies can be formally incorporated to confirm efficacy for a prospectively designed study not independently capable for such confirmation. It discounted prior study data to account for its hypothesis generating aspect without ignoring the information for the purpose of hypothesis confirmation.

The appropriate amount to borrow (θ) depends on if the discounted amount (1- θ) reasonably establishes the treatment benefit as a hypothesis to be confirmed. The Bayesian calculation translates this concept into the probability of treatment benefit based on the discounted fraction of previous study data. If the probability is large enough to establish the hypothesis, such as 90%, then the residual fraction from previous studies can be borrowed and integrated with new study data for independent hypothesis confirmation.

Determining an appropriate amount to borrow requires subjective judgement. There is no established convention to determine if a particular probability, say, 60%, is considered large enough to establish a hypothesis. Without additional information (such as data of reliable biomarkers), a default probability value of 90% should be sufficient for the purpose of generating hypotheses. The actual discount may be adjusted with a different corresponding probability than 90% based on subjective judgement using extra knowledge such as mechanism of action, similarity of study design, data consistency, and etc. If the prior data are compelling from virtually identical study design, borrowing a moderate amount is reasonable, as was illustrated using 50% borrowing for the pirfenidone mortality data. The probability of superiority for pirfenidone based on 50% of previous all-cause mortality data is over 90%, which is sufficient for hypothesis generation purposes. Borrowing the remaining 50% to form a prior of the new study for confirmation

purposes is therefore reasonable. The subjective nature of this determination should not discourage borrowing of valuable information from previous studies, as the alternative of ignoring compelling data from previous studies is much more problematic. In practice the sponsor and the regulatory agency should discuss an agreement before un-blinding of the prospective study to avoid ambiguity of study outcomes. In case of no pre-specified agreement, like the case of pirfenidone NDA, the analysis provides valuable information for understanding the overall data strength of treatment effect for regulatory decisions. Although justifying a particular fraction of borrowing can be difficult especially on a post-hoc basis, the tipping point calculation provides an intuitive and objective tool to evaluate the evidence of treatment effect based on a wide range of borrowing fraction had it been pre-specified, so that a positive conclusion is possible in a relatively conservative manner when data evidence is strong. Using the pirfenidone data as an example, the regulatory review team may determine if borrowing at least 29% from previous studies is justifiable for a positive conclusion of treatment effect on all-cause mortality.

Relation to the power prior model

The statistical model of borrowing historical data in Section 3 is a special case of the power prior model discussed by Ibrahim and Chen [9], and Ibrahim et al [10]. Ibrahim, Chen and SinHA[11] provided a formal justification of the power prior for Bayesian inference. The model for pirfenidone has a fixed borrowing fraction from pooled historical data for the advantage of simplicity in method communication, which is very important in the regulatory environment of drug development where the majority of professions are not statisticians. The identical study design, similarity of study population of the two historical studies supports data pooling (of the two historical studies) with a single discount fraction. In many other cases a more general power prior model may be appropriate to allow for a data driven dynamic borrowing through a hierarchical model with differences across historical studies and treatment arms [9, 12]. While such models are worth to be further studied, they are beyond the scope of this paper.

The cycle of learning and confirmation

Clinical drug development includes cycles of learning and confirmation [13]. Bayesian statistics has a natural framework for constant learning, and therefore the potential of improved efficiency for learning and confirmation. The Bayesian mortality analysis demonstrates that learning and confirming of hypotheses can be achieved without necessarily using "study" as the learning unit. It makes confirmation of treatment effect on mortality achievable without planning an impossibly large IPF study. In practice, the proposed approach should avoid or address the issues of selection bias and multiplicity, commonly reported as misuses of p-values [14, 15].

A focus of statistical application

This paper focuses on the application of the proposed method rather than the treatment effect of pirfenidone. We discuss pirfenidone's treatment effect for readers' appreciation of the importance of this approach. For interested readers, we adopted the study analysis plan's method of using one-year mortality data from the previous two studies to be consistent with the new study design, instead of using all mortality data from the previous studies which had various follow up duration from one and a half year and beyond. Contrary to many statisticians' opinion, we believe that using data with the same follow up duration is more suitable statistically, with the limitation that conclusions of the treatment effect are applicable to one year of treatment. Using the same duration of follow up data requires minimum statistical assumption compared to the alternative of using data with different duration of follow up that requires some assumption of no time difference. Although appropriate modeling can handle duration differences with additional assumptions, it is beyond the scope of this paper. We are aware of the potential selection bias of choosing one year mortality data instead of all mortality data. Therefore the one year mortality data from the previous two studies should be discounted for the purpose of hypothesis confirmation. A review of mortality data with different cuts of duration should help to understand the robustness of findings with one year duration. The pirfenidone treatment effect with a much longer duration is not assumed to be the same as with one year duration, and is beyond the scope of this discussion.

Conclusions

The Bayesian analysis provides a formal method to calibrate the role of information from historical evidence in the overall interpretation of results from both historical and concurrent clinical studies. The increased efficiency of using all available data is especially important in drug development for rare diseases with serious consequences, where limited patient source prohibits large trials, and unmet medical needs demand rapid access to treatment options. This Bayesian application illustrates that when results from historical studies are compelling, independent confirmation of treatment effect can be achieved more efficiently using a statistical integration of current and historical studies.

References

1. Berry D. A, and Stangl D. K. (1996). "Bayesian Methods in Health-Related Research," In Bayesian Biostatistics. New York: Marcel Dekker, 1-66.
2. Berry D. A. (2009). "Statistical Innovations in Cancer Research," In: Holland J, Frei T et al, editors. Cancer Medicine. 8th ed. London: BC Decker, 446-463.
3. Berry D. A. (2006). "A guide to drug discovery: Bayesian clinical trials," Nature Reviews Drug Discovery, 5; 27-36.
4. Viele K, Berry S, Neuenschwander B, Amzal B, et al. (2013). Use of historical control data for assessing treatment effects in clinical trials, Pharmaceutical Statistics. 13(1); 41-54.
5. Hobbs B. P, Carlin B. P, Mandrekar S J, Sargent D J. (2011). "Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials," Biometrics. 67(3); 1047-1056.
6. FDA Food and Drug Administration (11December2013), FDA Executive Summary P130013: Boston Scientific WATCHMAN® Left Atrial Appendage Closure Therapy. Meeting of the Circulatory System Devices Panel. Page 19: "In an attempt to incorporate PROTECT AF data yet avoid overly influencing the trial results, FDA and the sponsor agreed to use prior data in the analyses with a discounting weight of 50%. This means that the number of events and exposure time observed in PROTECT AF were down-weighted by 50% when using them as prior information in the final analysis of event rates".
7. Berry D. A. (1996). "Statistics: A Bayesian Perspective", Belmont, California: Duxbury Press.
8. Bradford W. Z, Cohen A. H, and Leff J A. (2013). Selection of clinical meaningful primary end-points in phase 3 clinical trials in idiopathic pulmonary fibrosis (letter). Am J Respir Crit Care Med. 187(11): 1269-1270.
9. Ibrahim J. G, Chen M. (2000). Power Prior Distributions for Regression Models, Statistical Science. 15(1): 46-60.
10. Ibrahim J. G, Chen M, Gwon Y, and Chen F. (2015). The power prior: theory and applications, Statistics in Medicine. 34(28): 3724-3749.
11. Ibrahim J. G, Chen M, and Sinha D. (2003). On Optimality Properties of the Power Prior, Journal of the American Statistical Association, 98(461): 204-213.
12. Hobbs B P, Sargent D J, Carlin B P. (2012). Commensurate Priors for Incorporating Historical Information in Clinical Trials Using

General and Generalized Linear Models. Bayesian Analysis, 7(3): 639-674.
13. Sheiner L. B. (1997). Learning versus confirming in clinical drug development, Clinical Pharmacology & Therapeutics. 61(3): 275-291.
14. Wasserstein R L. and Lazar N A. (2016). The ASA's statement on p-values: context, process, and purpose. The American Statistician. 70(1): 1-127.
15. Berry D A. (2016). P-Values Are Not What They are Cracked Up to be, The American Statistician. 1-2.

DOI: 10.31579/2692-9406/066