

We are all Bayesian, everyone is not a Bayesian

Mattia Andreoletti*

Andrea Oldofredi[†]

March 13, 2018

Abstract

Medical research makes intensive use of statistics in order to support its claims. In this paper we make explicit an epistemological tension between the conduct of clinical trials and their interpretation: statistical evidence is sometimes discarded on the basis of an (often) underlined Bayesian reasoning. We suggest that acknowledging the potentiality of Bayesian statistics might contribute to clarify and improve comprehension of medical research. Nevertheless, despite Bayesianism may provide a better account for scientific inference with respect to the standard frequentist approach, Bayesian statistics is rarely adopted in clinical research. The main reason lies in the supposed subjective elements characterizing this perspective. Hence, we discuss this objection presenting the so-called *Reference analysis*, a formal method which has been developed in the context of objective Bayesian statistics in order to define priors which have a minimal or null impact on posterior probabilities. Furthermore, according to this method only available data are relevant sources of information, so that it resists the most common criticisms against Bayesianism.

Keywords: Evidence; Clinical trials; Scientific Inference; Frequentism; Bayesian Statistics; Reference Analysis

*European Institute of Oncology, Dept. of Experimental Oncology; University of Milan, Dept. of Oncology and Hemato-oncology

[†]Université de Lausanne, Dept. of Philosophy, 1015 Chamberonne, Lausanne, Switzerland. Email: Andrea.Oldofredi@unil.ch (Corresponding Author)

Contents

1	Introduction	2
2	Case Studies	3
2.1	The Effect of Distant Healing	3
2.2	The Effect of SSRI Fluoxetine on Children and Adolescent Depression	4
3	Discussion	4
3.1	We are all Bayesian	5
3.2	Everyone is not a Bayesian	6
4	Reference Analysis: An Objective Approach to Bayesian Statistics	9
5	Conclusion	12

1 Introduction

How to deal with evidence in empirical sciences has raised many issues in philosophy, especially when related to other topics such as probability, explanation, and causality. Contemporary debates on evidence are prominent among philosophers of medicine especially since the birth of the Evidence Based Medicine movement in the 90s. In particular, one of the most discussed issues is whether or not bare statistical evidence may be considered reliable for warranting a scientific hypothesis; a good example may be the causal relation existing between a drug and the remission of some disease or, considering occupational epidemiology, the relation between workers' exposition to a particular substance and a certain pathology. As expected, across the relevant philosophical literature, the answer to this question is negative (the reader may refer to [Cartwright \(2007\)](#) and [Cartwright \(2010\)](#), [Worrall \(2010\)](#), [Stegenga \(2011\)](#), [Clarke et al. \(2014\)](#)). Despite these critiques, statistical studies - such as Randomized Clinical Trials (RCT) and meta-analysis - are, as a matter of fact, the gold standard methodology of clinical research, and several important decisions are actually made on the basis of their results, as, for instance, the approval of a new drug. This is because, as it has been argued time and again (cfr. [Teira \(2011\)](#), [Teira and Reiss \(2013\)](#)), RCTs are so far the best method to deliver what [Porter \(1996\)](#) describes as "procedural objectivity": since we trust the research methodology, we can agree on experimental results, no matter personal beliefs.

Having said that, in the following sections we are going to show that this supposed objectivity is a scientific ideal rather than a characterizing and inherent property of scientific knowledge: it seems indeed that medical consensus does not arise straightforwardly from the results of clinical research, as one would expect. In support of this view in [Section 2](#) we present two case studies: in the former, RCTs have been conducted to prove the efficacy of alternative remedies; although statistically significant results have been obtained *in favor* of the efficacy of distant healing, the scientific community is reluctant to accept the results. Whereas, in the latter, drugs for adolescent depression have been tested in several clinical trials with inconsistent results: in this case, however, the Food and Drug Administration (FDA) approved the drugs, and physicians are largely prescribing them. One may legitimately ask why it is the case that statistical evidence is sometimes *arbitrarily* accepted or rejected. In [Section 3](#) we offer a potential epistemic explanation, suggesting that a Bayesian framework might contribute to clarify and improve comprehension of medical research, since these case studies show how

personal beliefs play a prominent role in accepting or refuting a scientific hypothesis.

Against this background, despite Bayesianism can better account for scientific inference, Bayesian statistics (an approach to inferential statistics in which also pre-existing knowledge is taken into account) is rarely adopted in clinical research. The main reason lies in the supposed subjectivity of Bayesianism, or more precisely, the supposed subjective selection of priors. Hence, in Section 4 we present the so-called *Reference analysis*, a formal method which has been developed in the context of objective Bayesian statistics¹ in order to define *objective priors* the so-called “non-informative priors”, i.e. priors which have a minimal or null impact on posterior probabilities. Finally, in Section 5 we conclude that Bayesian statistics allows also for more transparent inferences.

2 Case Studies

To illustrate the point that statistical evidence alone does not by itself determine the amount of credibility that is granted to a given medical hypothesis, we focus on two therapeutic interventions, looking at the impact of single studies and meta-analyses as reflected in medical guidelines and commentaries. These studies test the efficacy of the following interventions: (i) prayer or other forms of "distant healing", (ii) the use of selective serotonin re-uptake inhibitors (SSRIs) to treat children and adolescents' depression.

2.1 The Effect of Distant Healing

A study contained in Leibovici (2011) published in the British Medical Journal explored the effect of distant, retroactive intercessory prayer in hospitalized patients with bloodstream infections. The study was performed on a large sample of patients (3393 adults) who were hospitalized between 1990 and 1996; the prayer was spoken in July 2000. The patients were randomly assigned either to the intervention, that is a prayer was spoken over the list of their names several years after their hospitalization, or to the control arm, i.e. no intervention. The clinical outcomes compared were the duration of stay in hospital, the duration of fever, and the overall mortality.

Notably, the intervention group showed shorter duration of fever ($p = 0.04$) and shorter stay in the hospital ($p = 0.01$). The results were published in December 2001, partly as a Christmas joke (the article also featured an image of a cervical smear resembling Rudolph the reindeer), and partly as a reaction to the previous published systematic review of studies assessing the efficacy of distant healing (Astin *et al.* (2001)). In that review the authors came to the conclusion that despite some methodological limitations present in the majority of the studies, further investigation the efficacy of distant healing is advisable since the 57% (13 of 23) of the included studies showed statistically significant treatment effects. Both Leibovici's trial and the systematic review of Astin *et al.* were published alongside comments on how to interpret these results. Many commentators addressed the methodological weaknesses in the trials, but there were also commentaries dealing with deeper issues. For instance, Ted Kaptchuk, known for his research on alternative medicine and placebo effect, asked:

Will more trials of distant healing with increased methodological rigor be helpful? If the results of such trials are negative, there is no problem: rational

¹While Reference analysis is not the only available proposal with which one may establish objective priors, it is certainly a good case study being it a major approach in this field.

and empirical knowledge agree. If the results of such trials are positive, would the evidence be persuasive?

He concludes:

It seems that the decision concerning acceptance of evidence (either in medicine or in religion) ultimately reflects the beliefs of the person that exist before all arguments and observation ([Kaptchuk \(2001\)](#)).

2.2 The Effect of SSRI Fluoxetine on Children and Adolescent Depression

In 2002 a placebo-controlled RCT on the effect of the SSRI fluoxetine (marketed as Prozac) in children and adolescents with depression was published in *The Journal of the American Academy of Child and Adolescent Psychiatry* ([Emslie et al. \(1997\)](#)). Fluoxetine was associated with greater mean improvement in the Children's Depression Rating Scale score compared to placebo though only slightly ($p < 0.05$). The conclusion was that "Fluoxetine 20 mg daily appears to be well tolerated and effective for acute treatment of MDD in child and adolescent outpatients". In 2004, after reports of increased suicide risk in children on SSRI antidepressant medication raised a public scandal, a meta-analysis of randomized controlled trials for various marketed SSRI-compounds was conducted, examining the evidence for their efficacy and safety in children and adolescents ([Whittington et al. \(2004\)](#)). The authors included both published and unpublished trials, in order to minimize publication bias. The results were unsatisfactory: the evidence for the efficacy of selective serotonin re-uptake inhibitors was inconsistent. Some trials for some compounds showed a significant effect, often just a small one, while others did not. As a conclusion, a need for more well-designed trials and greater transparency was expressed. Nonetheless, the same authors in the same paper emphasized fluoxetine as an efficacious, safe drug with a positive risk-benefit ratio based on two RCTs showing slightly positive effects and the meta-analysis. Fluoxetine remains the only SSRI approved by the FDA for treating children and adolescents with depression (age 8 and older).

3 Discussion

We are aware that data regarding the efficacy of psychotropic drugs is murky, to say the least: scandals about withheld negative results, fabricated data, and biased ghostwriters abound. On the other hand, the harm done to people choosing for the ineffective "alternative" treatments in life threatening illnesses is real as well. Biases introduced because of personal interests (may they be of financial nature or not) and/or by experimental sloppy design are real problems in clinical trials (and more generally in applied statistics), may they be for alternative or conventional therapies. Nonetheless, there are differences in the evaluation of the diverse therapeutic approaches that cannot be explained by the aforementioned factors.

While the publication of the study regarding retroactive prayer's efficacy in a medical journal may be clearly interpreted as a joke, or at most to spark philosophical discussion, the latter study concerning the efficacy of SSRIs in children and adolescents has had profound practical consequences, since it led to the market approval by the FDA. Moreover, although the present evidence is inconclusive, the majority of medical doctors and regulators probably would agree that conducting more unbiased trials with SSRIs would benefit our knowledge about their efficacy, while conducting more accurate trials on the efficacy of prayer or homeopathic remedies likely would not. In addition,

meta-analyses in the second case study would regard a rather specific intervention (a certain molecule tested for a specific indication in a well defined group of people), while the meta-analyses of trials investigating the impact of distant healing may concern a relatively broader (and not well defined!) range of interventions and indications: “Meta-analysis of molecule-based therapies”, even if it were feasible, does not sound like a good idea, but on the other hand “Meta-analysis of trials of Muslim prayer at noon against ear infection” sounds like a real waste of time and resources.

But why should it be so? There seems to be a deeper question concerning the standard statistical practice: is it possible that careful statistical analyses based solely on data, as the classical approach to inferential statistics requires, provide correct and complete answers to a given problem disregarding the context (and the relative available knowledge) in which a certain study is conducted?

3.1 We are all Bayesian

The most persuading answer, in our opinion, comes from the Bayesian approach to philosophy of science. In this regard, John Worrall ([Worrall \(2010\)](#)) claims that results of a certain class of studies like that of Leibovici are refuted by medical scientific community because actually *we are all naturally Bayesian*: at the moment of inference we cannot do without our beliefs. We cannot rationally accept the efficacy of retroactive prayer, since we are certain, according to our *background knowledge*, that it cannot have any effect, even though there is some statistical result showing precisely the contrary.

In Bayesian jargon, the statistical evidence obtained is not robust enough in order to modify our prior degree of belief in the efficacy of retroactive prayer for bloodstream infections, which is reasonably zero. As [Leibovici \(2011\)](#) points out: “if the pre-trial probability is infinitesimally low, the results of the trial will not really change it, and the trial should not be performed”. On its part, Worrall concludes that “Fisher’s insistence on not bringing any prior information into the assessment of the impact of a stochastic experiment in order to guarantee the objectivity was an understandable but grievous and enormously error”.

Although Worrall is not concerned with the limitations of classical statistical inference, he grasps a fundamental point about it: as a matter of fact, the claim of “objectivity” of frequentist analysis is somehow illusory, because many factors actually affect the interpretation of the results. When we come across borderline cases, the credibility of medical claims is measured by how much they resemble the generally accepted framework. For instance, even if the specific mechanistic explanation is not established, the model of how serotonin re-uptake inhibitors work matches the expectations that clinicians have about a therapy for psychiatric disorder. These expectations are developed through accumulating data from in vitro and animal studies, by breakthroughs in other areas of medicine which all shaped our “medical *Weltanschauung*”.

If it is true that we are all naturally Bayesian, then one may reasonably expect an implementation of formal Bayesian methods to analyze data. Indeed, the idea that Bayesian statistical inference can enhance the interpretation of randomized clinical trials is widespread also in scientific literature. For instance, [Wijeysundera et al. \(2009\)](#) argue that Bayesian statistical analysis overcomes several limitations of frequentist statistics. First of all “it permits inductive inference by reporting the clinically relevant probabilities for specified treatment effects”, secondly “it can determine the probability for varying magnitudes of therapeutic response”, and finally, above all, it can incorporate external information (e.g. previous evidence, biological plausibility, pre-existing beliefs) when interpreting the result of a study.

Generally, three specific components are necessary in order to analyze the results of a clinical trial in a Bayesian way: (i) the prior probability, (ii) the likelihood, and the (ii) posterior probability. The prior is the probability assigned a priori, based on information independent of the study to a hypothesis or to a parameter to be estimated, the likelihood is a function (not a probability function!) of an unknown parameter given a set of observed outcomes, and the posterior is the end-result, i.e. the probability of the hypothesis given the data resulting from the study and prior information, which can be calculated using Bayes' theorem:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

where $P(\theta|D)$ is the conditional probability of the parameter θ knowing the the observed data D . The posterior $P(D|\theta)$ expresses the probability that the observed outcomes D have been caused by (or in general terms causally related to) the θ , while $P(\theta)$ is the prior probability assigned to θ .²

In their study Wijesyundera and colleagues used Bayesian methods to reanalyze RCTs published in high-impact general medical journals and they found that for positive studies Bayesian and frequentist analyses had good agreement for the presence of any benefit. They close: “Bayesian inference reports probabilities that are theoretically consistent with the probabilities that clinicians are interested in, can be calculated for a range of clinically relevant effects and can be adjusted for differing pre-existing beliefs”.

To conclude, Bayesian reasoning seems a reliable way to describe what actually happens in medical research when decision-making, interpretation and report of randomized clinical trials are at stake. It seems to work quite well even as a normative claim: performing a Bayesian analysis is a more profitable way to justify the interpretation of results. As the epidemiologist Vandenbroucke says:

Science is an intrinsically human affair. When new theories are created and new evidence sought, judgment will retain a subjective element. This does not mean that it is impossible to sift out which interpretation is more valuable (Vandenbroucke (2005)).

Then it seems clear that if there exists an important subjective component in scientific processes, then a Bayesian framework may be the best way to take it into account, since currently it is the only perspective incorporating it.

3.2 Everyone is not a Bayesian

Nevertheless, according to a recent review, Bayesian trials are still a small niche in medical literature (Jack Lee and Chu (2012)), since the medical community is very

²It is useful to emphasize that in the context of Bayesian statistics

probability is *always* a function of two arguments, the event E whose uncertainty is being measured, and the conditions C under which the measurement takes place; “absolute” probabilities do not exist. In typical applications, one is interested in the probability of some event E given the available data D , the set of assumptions A which one is prepared to make about the mechanism which has generated the data, and the relevant contextual knowledge K which might be available. Thus, $Pr(E|D, A, K)$ is to be interpreted as a measure of (presumably rational) belief in the occurrence of the event E , given data D , assumptions A and any other available knowledge K , as a measure of how “likely” is the occurrence of E in these conditions (Bernardo (2003), p. 4).

reluctant to adopt Bayesian statistics in clinical research for its subjective component. Thus, it is mandatory to understand whether or not this critique to Bayesian statistics and to Bayesianism in general is justified. One has to acknowledge that priors are subjective and there is not a standard way to establish them; for instance, it is easy to imagine an hypothetical situation in which there is no information at all about a new treatment so that, in this case, the prior probability would be uninformative, basically it would be a flat distribution that permits the posterior to be determined almost entirely by the data. Then an enthusiastic prior and a skeptical prior might be constructed too and the posterior would be different according to which prior has been chosen. Indeed, the fundamental objection to Bayesian statistics addresses its subjectivist strand: the idea that prior and posterior distributions represent subjective states of knowledge. Here the concern is, first, that scientists should be concerned with objective knowledge rather than subjective beliefs, and second, that it is not clear how to assess subjective knowledge in any case (Gelman and Hennig (2017)). Although the apparent difficulty to meet the standard of objectivity within a Bayesian framework is not the only reason to discard Bayesian statistics, it is probably the most important, as pointed out by Efron (Efron (1986)): “the topic of objectivity [is] in my opinion the linchpin of non-Bayesian success with statistical practitioners”. Indeed, objectivity is one of the crucial factors discriminating scientific thinking from wishful thinking. We all have a special, but apparently justified, expectation: scientists have to draw the same conclusion starting from the same set of data, no matter personal beliefs. This is the main reason why *everyone is not a Bayesian*. Let us discuss this point.

In the first place, looking carefully at the classical approach to inferential statistics it is possible to find components typically ascribed to the Bayesian perspective, i.e. choices in which a priori knowledge and subjectivity determine the results of the inference. If it is true, then it seems inappropriate to claim that according to the frequentist account *only* sample data are relevant in order to provide information concerning an observed phenomenon. For instance, in practical situations usually one *presupposes* the form of the probabilistic model to adopt in order to analyze a certain kind of problem; it is trivial, therefore, to conclude that not only sample data are relevant for the inferential process, but also the previous knowledge of several similar studies and cases must be taken into account, meaning that this pre-existing knowledge (or prior information) allows us to automatically select a given model to represent a particular situation of interest. More precisely, the existing knowledge of the mechanisms that caused the available evidence is one of the elements responsible for the selection of probabilistic models.

Let us clarify this point with an easy example. Suppose to observe k times the event “head” in n independent tosses of a fair coin (n Bernoulli experiments). Our bare evidence consists, thus, in k successes in n independent trials. A relevant piece of information, however, is *not* contained in the available evidence: the mechanism generator of these data. On the one hand, if we consider a *fixed* number of trials, the binomial distribution will be the correct model, since it is a perfect representation of the situation at hand having to do with the number of successes in n independent experiments, whereas if the n trials are the result of a process requiring the realization of exactly k “heads”, the correct model will not be the binomial distribution, but rather the *negative* binomial. Now, if p denotes the occurrence of the event “head”, it is easy to see that the very same evidence leads to profoundly different conclusions depending on the mechanism generator of the data, and consequently on the model selection.

In the case we consider a fixed number of tosses, the stochastic variable X presents the following properties:

1. Mass probability function $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$
2. Mean: $E(X) = np$
3. Variance: $Var(X) = np(1-p)$.

Considering the latter case, X is characterized by:

1. Mass probability function: $P(X = x) = \binom{k+x-1}{x} p^k (1-p)^x$
2. Mean: $E(X) = k(1-p)/p$
3. Variance: $Var(X) = k(1-p)/p^2$

Furthermore, in our example the likelihood for these generator processes assumes different values, consequently the estimation of p in these two cases is naturally different. Hence, what this examples shows is the relevance that pre-existing knowledge has when we prefer a certain model to analyze data in particular situations, so that it should be emphasized that also in the context of classical inferential statistics there are crucial aspects which are not contained in bare sample data, showing that the alleged objectivity of the frequentist approach peacefully coexist with prior knowledge and subjective elements.

More generally, it is worth to be said that usually in statistics subjective elements are widespread, “starting from data preprocessing via data exploration and choice of method onto the selection of how to present which results. Realistically, even one-click methods require user choice on data coding and data exclusion, and these inputs can have big influences on end results such as p -values and confidence intervals” (Gelman and Hennig (2017)).

In this regard, following Sprenger (2017), it seems useful to underline that there is not a clear consensus on what constitutes objectivity in science.³ As well known in philosophical literature, there are many different accounts of scientific objectivity and, interestingly for our discussion, at least three of them pose obvious challenges for Bayesian methods: concordant objectivity, procedural objectivity, and value-free objectivity.⁴ Let us discuss them in order to present in the next section some Bayesian solutions to the problem of objectivity. In the first place, concordant objectivity (intersubjectivity) “occurs when a group of people all agree on an outcome” (Kincaid et al. (2007)); with reference to scientific practice this happens when all the members of the scientific community agree on experimental results. Intersubjective agreement has been long considered by philosophers of science as the cornerstone of scientific objectivity (e.g. one may consider Quine and Davidson’s reflections on this topic). Bayesian inference clearly violates this sense of objectivity since it opens up the possibility of taking into account prior knowledge which might vary between individuals: different scientists using different prior probabilities for analyzing the same data may well reach different conclusions. Secondly, values-free objectivity considers values and subjective judgments as a threat to the impartiality of scientific research. Finally, procedural objectivity occurs when a standardized experimental design is set up such that the same outcome is always produced, regardless of who is performing the experiment. This sense of objectivity

³The reader may refer also to Douglas (2009), where eight different senses of scientific objectivity have been identified.

⁴These are discussed in detail in Sprenger (2017).

is particularly influential in the context of clinical trials, as suggested by David Teira (Teira (2011)) to guarantee the impartiality of medical research with respect to external interests. Bayesian inference clearly violates these two latter senses as well. Especially in areas such as medical research where non-scientific interests at stake are high (financial, ethical, ...) consequence of biased or partial decisions might be very severe. The “trojan horse” of Bayesian procedures are prior distributions. There “non-fact-based information [...] will sneak it [...] wreaking havoc on the research’s interpretation” (Moyé (2008)).

4 Reference Analysis: An Objective Approach to Bayesian Statistics

Usually in the context of Bayesian statistics, if prior information concerning a specific parameter ϕ is available, it must be contained in the prior probability. On the contrary, in cases where no prior information is available, the so-called *objective* Bayesian frameworks provide formal methods to define priors with a *minimal* effect on the posterior distributions in order to minimize the effects of the prior’s selection and letting the data “speaking for themselves”. In this regard, it is worth to underline that in such cases the most relevant source of information are the set of available data, resembling the methodology of the frequentist approach to statistics.

It is well-known that several answers to the problem of finding priors in situation of lacunose or absent prior knowledge have been given throughout the history of statistics; in the first place, Bayes and Laplace tried to solve this issue applying the so-called *principle of insufficient reason*: when there is no information available about ϕ , then the prior should be defined as a uniform distribution, considering equiprobable every outcome of ϕ . Unfortunately, this definition is not satisfactory being not invariant under re-parametrization: if we do not have information concerning ϕ , consequently one would expect that there is no information also regarding $1/\phi$, however, the uniform priors of ϕ and $1/\phi$ are not equivalent. Therefore, one should look for better solutions.

Since the supposed lack of objectivity is a particularly cogent issue for Bayesians, the literature concerning Bayesian statistics faces directly the problem of the selection of priors in a formal manner. In fact, in this section we present a specific method developed primarily by J. Bernardo and J. Berger in order to define objective priors avoiding by construction the aforementioned problems and objections, i.e. prior with a minimal or null effect on posterior probabilities, usually called *reference priors*, hence the label *reference analysis* attributed to these authors.⁵ NB: this methodology obviously does not exhaust the spectrum of the solutions developed by Bayesian statisticians (cfr. Kass and Wasserman (1996) for a more systematic discussion of objective approaches in Bayesian statistics⁶), but it is a useful example to show that it is possible to limit the usual objections against the subjective elements presented by the Bayesian perspective.

The essential idea of Bernardo and Berger is to define a function which is able to maximize the expected divergence between prior and posterior probability distri-

⁵The following discussion is heavily influenced by Bernardo and Ramon (1998). Furthermore, it is important to note that this method is invariant under re-parametrization and avoid by construction the issues affecting the principle of insufficient reason.

⁶Reference analysis is just a specific objective method to minimize the subjective component of Bayesian statistics, and other methodologies are also available, e.g. one should consider Jeffreys (1961), Jaynes (1968). It is important to stress that Jeffreys’ priors and reference priors are equivalent in the case of one-dimensional parameters, while they diverge in the multidimensional case.

butions given the observational data, namely a function which maximizes the missing information in the experiment. Consequently, one should note that one of the most important feature of reference analysis is the definition of the notion of “missing information”. Now, if $f(x, \theta)$ is the probability density function of a one-parameter random variable X , and $T(X)$ a sufficient statistics for the parameter θ , then, assuming the availability of $X_1^n = (X_1, \dots, X_n) = x_n^*$ observations, Bernardo and Berger derive the prior distribution $\pi_{ref}(\theta)$ maximizing the expected value of $K_n(\pi(\theta|x_1^n), \pi(\theta)) = \int \pi(\theta|x_1^n) \log(\pi(\theta|x_1^n)/\pi(\theta)) d\theta$, in this case is the Kullback-Leibler distance, which is a measure of how a given probability distribution $\pi(\theta|x_1^n)$ diverges from a second probability distribution $\pi(\theta)$, when the latter is used to approximate the former. If K_n^π is the expected value of the distance with respect to X one gets $K_n^\pi = E_{x_n^*} \{K_n[\pi(\theta, x_n^*), \pi(\theta)]\}$, the reference prior maximizes $K_\infty^\pi = \lim_{n \rightarrow \infty} K_n^\pi$.

To state it alternatively, the reference posterior has the role to specify exactly what could be learnt and gathered from data concerning a given quantity of interest ϕ , assuming that *only* the set of observed data provides the available information regarding it. With Bernardo’s words:

the reference posterior exploits certain well-defined features of a possible prior, namely those describing a situation where relevant knowledge about the quantity of interest [...] may be held to be negligible compared to the information about that quantity which repeated experimentation (from a particular data generating mechanism) might possibly provide. [...] Any statistical analysis contains a fair number of subjective elements; these include (among others) the data selected, the model assumptions, and the choice of the quantities of interest. Reference analysis may be argued to provide an “objective” Bayesian solution to statistical inference problems in just the same sense that conventional statistical methods claim to be “objective”: in that the solutions only depend on model assumptions and observed data (Bernardo (2003)).

Insisting again on the concept of information, it is interesting to note that the reference approach make use of Shannon’s definition of information in order to keep this notion as precise as possible.⁷ In the context of reference analysis, the more prior knowledge is available the less information is expected to be gained by the performance of experiments, so that the amount of information which is expected from an experiment depends strictly on (i.e. is a function of) the available prior knowledge. Thus, the core idea of the reference analysis may be rephrased as follows:

An infinitely large experiment would eventually provide all missing information; thus, it is possible to obtain a measure of the amount of missing information as a limiting form of a functional of the prior distribution. It is natural to define “vague” prior knowledge as that with the largest missing information: *the reference prior* should then be that which *maximizes the missing information*. (Bernardo and Ramon (1998))

Hence, as we have seen a few lines above, the reference prior is defined as a limit of a sequence of prior distributions which maximizes the expected information in the long run, i.e. “from an increasingly number of experiments”.

⁷The reader should refer to Soofi (1994) for a useful introduction to the technical notion of information applied in statistics.

More precisely, taking into account the simplest case of one parameter⁸, considering an experiment consisting in the observation of a specific set of data D generated by a probability model $p(D|\phi)$, which depend only on the quantity of interest ϕ , and letting be $\mathbf{t} = \mathbf{t}(D) \in \mathcal{T}$ a sufficient statistic about the observed data, the expected amount of information $I\{\mathcal{T}, p(\phi)\}$ provided by the gathered data D concerning ϕ , given the prior knowledge $p(\phi)$ is defined as follows:

$$I\{\mathcal{T}, p(\phi)\} = \int_{\mathcal{T}} \int_{\Phi} p(\mathbf{t}, \phi) \log \frac{p(\mathbf{t}, \phi)}{p(\mathbf{t})p(\phi)} d\phi d\mathbf{t},$$

According to this definition, the amount of information provided by the data is defined by the “expected logarithmic divergence of the prior from the posterior” (Bernardo (2003)), what a few lines above we have written in terms of the Kullback-Leibler distance. This definition tells us that the missing information about the quantity ϕ would be provided by k independent repetitions of the experiment; thus, in the long run, as $k \rightarrow \infty$, $I\{\mathcal{T}, p(\phi)\}$ will ideally provide any missing information concerning ϕ .

The last step to the definition of the reference prior is to consider, for large k , an asymptotic approximation of the posterior distribution (which does not depend on the priors) from which it is possible to derive a sequence of positive functions which are able to induce a sequence of posterior distributions via Bayes theorem, containing the desired reference posterior distribution. With Bernardo’s words: “the reference prior function $\pi(\phi)$ of a univariate parameter ϕ is defined to be the limit of the sequence of the proper priors $\pi_k(\phi)$ which maximize $I\{\mathcal{T}, p(\phi)\}$ in the precise sense that, for any value of the sufficient statistic $\mathbf{t} = \mathbf{t}(D)$, the reference posterior, the intrinsic limit $\pi(\phi|\mathbf{t})$ of the corresponding sequence of posteriors $\{\pi_k(\phi|\mathbf{t})\}$, may be obtained from $\pi(\phi)$ by formal use of Bayes theorem, so that $\pi(\phi|\mathbf{t}) \propto p(\mathbf{t}|\phi)\pi(\phi)$ ” (Bernardo (2003), notation adapted).

Now that we have loosely introduced the main ideas of the reference analysis, it is worth to emphasize that reference priors are *not* probability distributions, but rather functions. This fact implies that they should not be considered degree of beliefs with respect to a certain proposition. Moreover, in support of this fact, often reference priors are not proper probability distributions since do not integrate one. However, it is technically possible to obtain proper posteriors from *improper* priors. The interested reader should refer to Bernardo (1997) for details and to Syversveen (1998) and Kass and Wasserman (1996) for a detailed discussion of problematic/artificial aspects of reference analysis and open problems for future research. Another relevant aspect which deserves to be mentioned is the interpretation of such priors. Firstly, they should not be interpreted as a measure or a representation of ignorance, but they should be rather compared to a default option in a certain computer package (remarkably, the term “reference prior” may recall the idea of standardization); Kass and Wasserman (1996) say “[i]n principle, we could construct a systematic catalog of reference priors for a variety of models. The priors in the catalog do not represent ignorance, but are useful in problems where it is too difficult to elicit an appropriate subjective prior. The statistician may feel that the reference prior is [...] a good approximation to any reasonable subjective prior for that problem”. Thus, there is no an objective prior representing ignorance, and one should keep in mind that the definition of reference priors strictly depends on the inference problem at hand.

⁸More complex cases and examples are discussed in great detail in Bernardo (2003) and Bernardo and Ramon (1998).

Having briefly discussed an objective Bayesian approach, it is possible to speculate on how one may reply to the previous critiques concerning objectivity. It is crucial to underline that often the term “objectivity” refers to the independence of personal biases, as [Gelman and Hennig \(2017\)](#) point out, and it seems fair to claim that objective Bayesian approaches try to provide methods to construct priors in this direction, yielding algorithms based on transparent assumptions and reproducible results. To this regard, concordant objectivity is not violated, since given a particular problem objective priors are by construction defined in order to lead to the same posterior distribution. Hence, different statisticians will agree on the final probability distributions. Recalling the passage of Kass and Wasserman quoted a few lines above, in cases in which it is possible to define an objective prior, the problem of intersubjectivity vanishes because such objective priors represent optimal approximations of subjective priors, so that the subjective elements do not influence the process of inference and data analysis.

In second place, one may also say that Bernardo’s proposal (among others, e.g. those of Jeffreys and Jaynes mentioned in footnote 4) being formal procedures satisfy by construction several epistemic virtues: not only methodological transparency, but also clear and rigorous argumentation, and impartiality. This fact allows to reply also to the Values Free objection. Gelman and Hennig writes, in this regard, that

[f]ormal statistical methods contribute to objectivity as far as they contribute to the fulfillment of these desiderata [transparency and impartiality], particularly by making procedures and their implied rationales transparent and unambiguous ([Gelman and Hennig \(2017\)](#)).

Finally, procedural objectivity is essential also in the context of objective Bayesian statistics, since only data gathered from experiments are relevant for the definition of such priors (trivially, reference priors have been created specifically in order to have the minimal effect on posterior distributions). This specific trait of reference analysis should make this proposal appealing also to frequentist, since the main difference between the Bayesian and frequentist statistics lies in the fact that according to the latter only *data* are relevant.

In sum, we are well aware that currently there is not a unique point of view in the Bayesian community on the validity of objective approaches; however, although it is true that objectivity may be seen as an ambition rather than an achievement as [Gelman and Hennig \(2017\)](#) underline, these approaches aim to achieve something more than desirable, i.e. a sort of standardized priors to use in situations where there is an effective absence of information. Be that as it may, these approaches certainly show how it is possible to provide convincing answers to the usual objections against Bayesianism.

5 Conclusion

Let us conclude summarizing what we achieved in this paper. At the very beginning we questioned the idea that scientific consensus on medical claims is obtained throughout accumulation of RCTs. We briefly presented two case studies which starkly clash with that idea. From a pure epistemic standpoint, we can explain the interpretation of experimental results appealing to Bayesian inference. Following John Worrall, we suggested that we are all naturally Bayesian: the credibility of medical studies is measured by how much they resemble already existing knowledge. We cannot neither accept the efficacy of retroactive prayer nor completely regard SSRI as ineffective, despite statistical evidence showing the opposite is available. Hence, a Bayesian framework seems to better capture

the achievement of medical consensus, considering available information in assessing the results of clinical studies. Nonetheless, Bayesian statistics is rarely adopted in clinical research because of its putative subjectivity in selecting prior probabilities. In contrast, not only we have underlined that also in the frequentist account to inferential statistics several decisions are taken considering pre-existing knowledge and subjective beliefs, but also we presented a formal method which allows a rigorous definition of priors in specific situations where prior information available is negligible. This method does not violate the senses of scientific objectivity which traditionally pose challenges to the Bayesian perspective.

Finally, the adoption of an objective Bayesian statistical analysis would have a further advantage, that is allowing for more transparency. As far we are concerned with scientific objectivity, methodological transparency plays a key role. For instance, with respect to the fluoxetine case, a Bayesian analysis would have made the data more transparent and compelling, without the possibility for the authors to subjectively assess the safety and efficacy of the drug.

In clinical research transparency becomes paramount, because all the methodological choices made by the experimenters might have a significant impact on final results, hence they should be motivated on the basis of rational criteria. As recently argued by Gelman and Hening ([Gelman and Hennig \(2017\)](#)) the adoption of prior probabilities makes the researchers' point of view more transparent. As they put it, “[t]he merit of objectivist Bayesianism is that the approach comes with a much stronger drive to justify prior distributions in a transparent way using principles that are as clear and general as possible”.

Acknowledgements A. Oldofredi is grateful to the Swiss National Science Foundation for financial support (Grant No. 105212-175971).

References

- Astin, J. A., Harkness, E., and Ernst, E. (2001). Distant healing. *Annals of Internal Medicine*, 132(6):903–910.
- Bernardo, J. M. (1997). Non-informative priors do not exist. *Journal of Statistical Planning and Inference*, 65:159 – 189 (with discussion).
- Bernardo, J. M. (2003). Bayesian statistics. In Viertl, R., editor, *Probability and Statistics*. Encyclopedia of Life Support Systems (EOLSS). Oxford, UK: UNESCO.
- Bernardo, J. M. and Ramon, J. M. (1998). An introduction to bayesian reference analysis: inference of the ratio of multinomial parameters. *The Statistician*, 45:1 – 35.
- Cartwright, N. (2007). Are rct the gold standard? *BioSocieties*, 1:11–20.
- Cartwright, N. (2010). What are randomised controlled trials good for? *Philosophical Studies*, 147(1):59–70.
- Clarke, B., Gilles, D., Illari, P., Russo, F., and Williamson, J. (2014). Mechanism and the evidence hierarchy. *Topoi*, 33(2):339–360.
- Douglas, H. (2009). *Science, policy, and the value-free ideal*. University of Pittsburgh Press.
- Efron, B. (1986). Why isn’t everyone a bayesian? *The American Statistician*, 40(1):1 – 5.
- Emslie, G., Rush, A., Weinberg, W., Kowatch, R., Hughes, C., Camody, T., and Rintelmann, J. (1997). A double-blind, randomized, placebo-controlled trial of fluoxetine in children and adolescents with depression. *JAMA Psychiatry*, 54(11):1037–1037.
- Gelman, A. and Hennig, C. (2017). Beyond subjective and objective in statistics. *Journal of the Royal Statistical Society, Series A*, 180:1 – 31.
- Jack Lee, J. and Chu, C. T. (2012). Bayesian clinical trials in action. *Statistics in medicine*, 31(25):2955–2972.
- Jaynes, E. (1968). Prior probabilities. *IEEE Trans. Systems, Science and Cybernetics* 4, pages 227 – 291.
- Jeffreys, H. (1961). *Theory of Probability*. Oxford: Oxford University Press, 3rd edition edition.
- Kaptchuk, T. (2001). Distant healing. *Annals of Internal Medicine*, 134(6):532 – 533.
- Kass, R. and Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*, 91(431):1343 –1370.
- Kincaid, H., Dupré, J., and Wylie, A. (2007). *Value-Free Science: Ideals and Illusions?* Oxford University Press.
- Leibovici, L. (2011). Effects of remote, retroactive intercessory prayer on outcomes in patients with bloodstream infection: randomised controlled trial. *British Medical Journal*, 323:1450.

- Moyé, L. A. (2008). Bayesians in clinical trials: asleep at the switch. *Statistics in medicine*, 27(4):469–482.
- Porter, T. M. (1996). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton University Press.
- Soofi, E. (1994). Capturing the intangible concept of information. *Journal of the American Statistical Association*, 89(428):1243 – 1254.
- Sprenger, I. (2017). The objectivity of subjective bayesianism. *philsci-archive.pitt.edu*, pages 1 – 27.
- Stegenga, J. (2011). Is meta-analysis the platinum standard for evidence? *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 42(4):497 – 507.
- Syversveen, A. R. (1998). Non-informative bayesian priors. interpretation and problems with construction and applications. *Preprint*, pages 1 – 11.
- Teira, D. (2011). Frequentist versus bayesian clinical trials. *Handbook of the philosophy of science. Philosophy of Medicine*, 16:255–298.
- Teira, D. and Reiss, J. (2013). Causality, impartiality and evidence-based policy. In *Mechanism and causality in biology and economics*, pages 207–224. Springer.
- Vandenbroucke, J. (2005). Homeopathy and the growth of truth. *The Lancet*, 366(9487):691 – 692.
- Whittington, C., Kendall, T., Fonagy, P., Cottrell, D., Cotgrove, A., and Boddington, E. (2004). Selective serotonin reuptake inhibitors in childhood depression: systematic review of published versus unpublished data. *The Lancet*, 363(9418):1341 – 1345.
- Wijeysundera, D., Austin, P., Hux, J., Beattle, W., and Laupacis, A. (2009). Bayesian statistical inference enhances the interpretation of contemporary randomized controlled trials. *Journal of Clinical Epidemiology*, 62(1):13 – 21.
- Worrall, J. (2010). Evidence: philosophy of science meets medicine. *Journal of Evaluation in Clinical Practice*, 16(2):356 – 362.