Are 'Phase IV' Trials Exploratory or Confirmatory Experiments?

Austin Due
austinjdue@gmail.com
Pre-print. Final version @ *Studies in History and Philosophy of Science* 95 (2022), 126-133.
https://doi.org/10.1016/j.shpsa.2022.07.006
Please cite final/published version.

Abstract
Exploratory experiments are characterized as experiments that do not test hypotheses. Experiments that *do* test hypotheses are often characterized as confirmatory experiments. Philosophers of science have pointed out that research programmes can be both confirmatory and exploratory. However, the way confirmatory and exploratory experimentation are each defined precludes *single* cases of experimentation being jointly confirmatory and exploratory; how can an experiment both test and not test a hypothesis? Here I argue that a recharacterization of the relationship between exploratory and confirmatory experimentation is needed, and I appeal to 'phase IV' trials to show what this recharacterization could look like. In short, I offer a recharacterization of the relationship between exploratory and confirmatory experimentation where the former remains a distinct kind of experimentation but is not necessarily non-hypothesis-testing.

Keywords: experimentation; exploration; clinical trials; phase IV.

1. Introduction[1]

We often worry about the safety and effectiveness of new therapeutic drugs. Pre-market trials discover much about new drugs, but long-term side effects and contraindications are not usually known until a drug is on the market and used in the general population for some time. Also, we often do not know how effective a drug is in populations outside those represented in pre-market trials. This gap in our knowledge around novel pharmaceuticals can be due to (i) permissive policies of regulatory bodies and (ii) evidence-based medicine's insistence of randomized control trials (RCTs) as the 'gold-standard' of evidence. Criticisms of (i) tend to focus on private industry influence in the regulatory process (Gaffney & Lexchin 2018; Lexchin 2016; Light et al., 2013). Criticisms of (ii) are varied and include questions about the purpose and ethics of randomization (Worrall 2002; Worrall 2008), the

[1] My thanks to Ross Upshur, Maya Goldenberg, Mike Miller, Brian Baigrie, Robyn Bluhm, Brian Feldman, Mathew Mercuri, and audiences at the 2021 meetings of the Canadian Society for History and Philosophy of Science and the International Philosophy of Medicine Roundtable as well as two anonymous reviews for their comments and feedback on this and previous iterations of this paper.

external validity of RCTs (Cartwright 2007; Cartwright 2011), and the failure of pre-market RCTs to find unforeseen side effects (Vandenbroucke 2008; Osimani 2014; Stegenga 2016). Moreover, issues of (i) and (ii) overlap in privately funded trials (Sismondo 2008; Sismondo 2009; Solomon 2017).

This combination of factors necessitates post-market drug surveillance or 'pharmacovigilance.' Pharmacovigilance refers to a range of practices related to detecting, evaluating, and preventing drug-related problems (May 2014). Techniques in pharmacovigilance range from the spontaneous reporting of suspected adverse drug reactions to regulatory bodies to 'phase IV' trials. Phase IV trials – those following pre-market phase III RCTs – often occur in 'real world' uncontrolled and nonblinded clinical settings. Phase IV trials blur the lines between clinical research and clinical practice and are relatively understudied entities philosophers of science and medicine engage with, especially in comparison to phase III RCTs. Here I argue that phase IV trials can enrich philosophical accounts of experimentation, specifically the relationship between confirmatory and exploratory experimentation.[2]

Exploratory experimentation is characterized, among other things, as experimentation that does not aim to test or evaluate hypotheses. This is the 'fundamental condition' of exploratory experimentation (Elliot 2007, p. 322-323). Experiments that *do* aim to test or evaluate hypotheses are confirmatory experiments. Philosophers have been right to show that research programmes can be constituted by both exploratory and confirmatory experiments (O'Malley 2007; Waters 2007). However, if a condition of exploratory experimentation is that it does not test hypotheses, it seems that no *single* experiment can be both exploratory and confirmatory; how can an experiment concurrently test and not test a hypothesis? If we want to say a single experiment can be both exploratory and confirmatory, that means the characterization of exploratory experimentation and its relationship to confirmatory experimentation is incomplete at best and wrong at worst. I take it as intuitive that some experiments are confirmatory, some are exploratory, and some can be both. Therefore, a recharacterization of exploratory experimentation and its relationship to confirmatory experimentation is warranted. I argue thinking about phase IV trials show us how this could be done.

[2] The 'lesson' I extract from thinking about phase IV trials here is not something specific to phase IV trials. Rather, phase IV trials are one of, assumedly, many kinds of experiment that would serve the purpose of re-thinking the relation between exploratory and confirmatory experimentation in single cases of experimentation. I discuss this in more detail in section 6 below.

I argue that some phase IV trials are clear cases of experiments that are both confirmatory and exploratory insofar as they aim to test a hypothesis and explore for unforeseen phenomena. Given these kinds of experiments, the relationship between confirmatory and exploratory experimentation can be recharacterized as one that can co-exist in single cases. This means giving up the 'fundamental condition' and no longer defining exploratory experimentation as non-hypothesis-testing. Firstly, I argue that exploration should be defined in terms of its positive aims, not as what it does not do. Secondly, I propose that exploratory and confirmatory experimentation should not be characterized as dichotomous or along a continuum. I will show that with these two points taken together, we can better explain how a single experiment can be both confirmatory and exploratory. Even if it is uncontroversial that a single experiment can have multiple aims, the recharacterization of the relationship between exploratory and confirmatory experimentation is still required for these aims to be held together without tension or contradiction.

To offer an alternative way to think about exploratory experimentation and its relationship to confirmatory experimentation given phase IV trials, I first address some contemporary accounts of exploratory experimentation in section 2. There I show that exploratory experimentation is characterized, among other things, as fundamentally non-hypothesis-testing. Section 3 introduces phase IV trials in general before section 4 addresses a case study of an interventional phase IV trial of imatinib in pediatric leukemia patients. I stress 'interventional' phase IV trials here rather than post-hoc statistical analyses that can also be characterized as phase IV trials. Section 5 then addresses the exploratory component of phase IV trials. I discuss how the side effect discovery process in phase IV trials entails exploring for unforeseen phenomena concurrently with hypothesis testing. Moreover, the paucity of external validity from phase III RCTs makes the exploratory and confirmatory functions in phase IV trials hard to separate. Section 6 then offers an alternative characterization of exploratory experimentation and its relationship to confirmatory experimentation. 'Aiming to test hypotheses' should not be a condition that precludes an experiment from being characterized as exploratory. Before concluding in section 7, I show that because of experiments like phase IV trials that blur the lines between confirmation and exploration, the relationship between confirmatory and exploratory experimentation should be explained differently than diametric points along a continuum. These are distinct kinds of experimentation that can be instantiated together, separately, or not at all.

2. Confirmatory & Exploratory Experimentation

2.1. New Experimentalism & Exploration

With 'new experimentalism' arising in philosophy in the 1980s, there was a call for philosophers to be more attentive to the various functions of experimentation outside hypothesis testing.[3] The joint genesis of the literature around the exploratory function of experimentation is often attributed to works by Richard Burian (1997) and Friedrich Steinle (1997). Though the accounts differ insofar as Steinle's account discusses parameter variation in uncertain contexts and Burian's discusses triangulation, there are overlaps (Elliot 2007; Schickore 2016). In both cases, exploration occurs in cases where no well-developed theories are yet present. Exploratory experimentation is described as a 'mapping' activity that probes the unknown and unforeseen relationships between or within phenomena. Exploration can create and stabilize entities for analysis when theory is lacking (Franklin 2005; Elliot 2007). However, this does not mean exploration is necessarily theory-free. Laura Franklin-Hall (2005) argues that exploratory experiments can be guided by a *background theory*. Background theories can tell us what an entity is and which tools are appropriate to investigate it (Franklin 2005). A background theory might let me know that some unknown entity is a liquid, and that liquids take on the shapes of their containers and boil when heated. However, this background theory does not predict the behavior of the entity. For example, the background theory does not make a prediction about at *which* temperature the unknown liquid will begin to boil.

David Colaço (2018) further argues that in addition to background theories, *local theories* or theories about the specific entity can also play a role in exploratory experiments. Say I *did* know at which temperature our mystery liquid boils; this would be a 'local theory.' Colaço (2018) maintains that if local theories inform the experiment while not being explicitly tested or evaluated, those local theories remain auxiliary to the experiment. Theory being involved but only in an auxiliary sense preserves the exploratory nature of the experiment. In short, what characterizes exploratory experimentation is not that it is free from theory or hypotheses. Rather, what is shared in accounts of exploratory experimentation is that it explicitly does not aim to test or evaluate hypotheses and specific predictions from theories.

---

[3] Jutta Schickore (2016) points out that philosophy-adjacent investigations into experimentation and the 'logic of discovery' were occurring prior to the dawn of 'new experimentalism,' sometimes marked with Hacking's (1983) *Representing and Intervening*.

2.2 A Taxonomy for Exploratory Experiments

As above, an underlying connection between accounts of exploratory experimentation is that exploration does not aim to test or evaluate hypotheses. Discussing various accounts of exploratory experimentation, Kevin Elliot (2007) posits that "…the most fundamental characteristic of [exploratory experimentation] seems to be that it, in contrast to other types of experimentation, does *not* serve the aim of testing theories or hypotheses…[exploratory experimentation] does not involve testing specific predictions of a particular theory…" (2007, 322-323). With that in mind, Elliot provides a taxonomy of exploratory experimentation along three dimensions: (1) the positive aims of the experiment, (2) the role theory plays in the experiment, and (3) methods used to vary experimental parameters (Elliot 2007, p. 323). Not included as roles for theory in (2) are 'being something evaluated' or 'tested.' This reifies the 'fundamental' characteristic of exploratory experiments as non-hypothesis-testing. As mentioned above, philosophers of science have pointed out that confirmatory and exploratory experiments can together constitute research programmes (O'Malley 2007; Waters 2007). However, it is not clear if this is intended for *single* experiments, especially given the 'fundamental' characterization of exploratory experiments as non-hypothesis-testing.

If a fundamental condition of exploratory experimentation is that it does not test hypotheses, saying that some experiment E is both exploratory and confirmatory amounts to saying E both tests and does not test some hypothesis. At best this is vague. I take it as intuitive that we want to say some single experiments can be both confirmatory and exploratory. Moreover, I think there are concrete cases of single experiments that are clearly both confirmatory and exploratory, like the phase IV trials discussed in the proceeding sections. Therefore, a re-thinking of the common characterization of exploratory experimentation as something that does not test hypotheses and its relation to confirmatory experimentation is warranted. This relation will be one that needs to have confirmation and evaluation and exploration co-exist in a non-dichotomous way, fully detailed in section 6.

In sum of this section, the above shows that exploratory experimentation aims to uncover novel phenomena and relationships between phenomena. Exploratory experimentation has been described as involving theories or hypotheses in various ways, as long as those theories and hypotheses are not being evaluated or tested. A role *not* included on Elliot's (2007) dimension (1) is 'aim to confirm hypotheses.' A role *not* included in dimension (2) is 'be tested or evaluated.' This taxonomy incorporates the widely held assumption that exploration does not test hypotheses. Notice also that

'aims' is just one dimension of the taxonomy. Methodological features like (3) are also important to characterize exploratory experiments. If we added to dimension (1) 'aim to confirm hypotheses' or add to dimension (2) 'be tested or evaluated', then the relationship between exploratory and confirmatory experimentation would need re-thinking. This is precisely what phase IV trials can require us to do, since I aim to show that they are single experiments that are both confirmatory and exploratory. Before presenting a concrete case, a more general discussion of phase IV trials and their function in contemporary medical practice and research can be useful and now follows.

3. Phase IV Pharmaceutical Trials

3.1. Why 'IV'?

Novel pharmaceuticals go through 'phases' of research trials before regulatory bodies can approve them for the market.[4] After successful lab and animal studies, novel drugs can be tested on humans. Phase I or 'first-in-human' trials are typically trials on twenty to eighty paid, healthy volunteers.[5] Phase I trials test a new drug at various doses to determine proper therapeutic dosage and acute side effects. Once it is determined that a new drug is safe enough for human use, the drug moves on to phase II trials.

Phase II and phase III trials are respectively smaller and larger RCTs. Larger phase III RCTs can have trial populations in the thousands, though this is not overly common. If a pharmaceutical company presents two positive RCTs deemed 'pivotal' by the US Food and Drug Administration (FDA), this is often sufficient for regulatory approval.[6] Once a drug is on the market, trials to monitor ongoing use and safety are considered phase IV trials. Regulatory bodies might require private companies to run these trials, or public researchers might run the trials themselves. Phase IV trials often occur in clinical settings, thus they are not always randomized or blinded. Not all phase IV trials are 'interventional,' but can include statistical analyses and case control studies on existing data (Suvarna 2010). For the current context, I have interventional phase IV trials in mind, where an intervention is

[4] As per: https://www.fda.gov/patients/learn-about-drug-and-deviceapprovals/drug-development-process.

[5] Volunteers who typically come from lower socioeconomic classes and historically oppressed groups who are compensated little in comparison to acute and long-term risks (Fisher 2020).

[6] That is, most of the time. As of 2016 and the '21st Century Cures Act' the FDA has the power to approve drugs to market without necessarily enforcing the 'two RCTs' rule (Schwartz 2017).

given to a population and results are measured. Because phase IV trials are not always randomized or blinded, they have received some criticism as marketing tools (Kessler et al. 1994; Sismondo 2008). However, randomization and blinding are not always needed to control for unforeseen side effects (Vandenbroucke 2008; Osimani 2014). And, because phase IV trials happen in therapeutic clinical settings, not telling a patient what drug they are prescribed can go against physicians' moral obligations to patients (Bernabe 2016; Bernabe et al. 2014). Regardless of the concerns around phase IV trials, they remain an important part of pharmacovigilance.

3.2. From Efficacy to Effectiveness

Phase III RCTs establish what is commonly understood as the *efficacy* of a new drug (Suvarna 2010). 'Efficacy' describes how well the drug works under ideal or tightly-controlled circumstances. Pre-market industry funded RCTs are often highly selective and have homogenous, young, healthy populations. People with comorbidities taking concurrent medications, the elderly, children, and pregnant women have historically been excluded from many of these trials. The more tightly controlled the RCT, the more 'ideal' it becomes and the better a drug's efficacy is determined. Consequently, many RCTs funded by private industry have notoriously low external validity. This means extrapolating from the RCTs to contexts outside the RCT can be unreliable (Cartwright 2007; Bluhm 2007; Clarke et al. 2013). The paucity of external validity in RCTs can be amended with larger more diverse sample sizes, but this is not overly common.

Once a drug is approved for the market, phase IV trials (if they occur) are the first 'real world' test of a new drug (Suvarna 2010). Phase IV trials can test a drug outside of the artificial contexts of the pre-market RCT. By evaluating the efficacy established in phase III trials, phase IV trials aim to establish the *effectiveness* of the drug.[7] Where drug efficacy is how a drug performs in the ideal circumstances like those of a tightly controlled RCT, effectiveness is how it works in a wider, more heterogenous group of patients with the relevant illness (Suvarna 2010). Phase IV trials are in principle more externally valid than phase III trials, since phase IV trials answer the question of whether or not a drug works in varied situations. This is analogous to Nancy Cartwright's (2007; 2011) claims that RCTs establish that

[7] Or, more precisely, play a part in establishing the effectiveness of a new drug. Given questions about long term use of drugs, co-morbid use of drugs, and potential off-label uses, it is difficult to know when we ever really have a comprehensive certainty about the effectiveness of some drug. Determining effectiveness is a piecemeal process.

a drug 'works somewhere,' but the question we want answered is if it 'works for us.' Efficacy is the evidence that a drug does work somewhere, i.e., in the context of the RCT. Effectiveness is how a drug will work in more heterogenous 'real world' contexts; 'effectiveness' is what phase IV trials can do a better job of telling us as the overlap of clinical research and clinical practice. However, instances of phase IV trials can run into similar methodological and biasing issues of phase III RCTs. Either way, phase IV trials remain an important tool in pharmacovigilance.

In what follows, I show how a phase IV trial can blur the lines between exploratory and confirmatory experimentation. Phase IV trials as described are single experiments that aim to test or evaluate a hypothesis while exploring for unforeseen phenomena and relations between phenomena. The hypothesis that can be evaluated in a phase IV trial is related to the efficacy established in the phase III trials. In other words, a phase IV trial establishes a drug's effectiveness by evaluating (or testing) in the 'real world' the efficacy previously established in phase III. Phase IV trials also explicitly monitor for unforeseen phenomena, e.g., unpredicted side effects. What follows is a concrete case to help illuminate these points.

4. Phase IV Pediatric CML Imatinib Trial

4.1 Millot et al. (2011)

Frédéric Millot et al. (2011) presents the results of a multi-site French national phase IV trial from 2004-2010 on forty-four pediatric patients with newly diagnosed chronic myelogenous leukemia (CML). CML is a rare form of leukemia, accounting for 2%-3% of cases in children. 'Newly' diagnosed meant treatment occurred within two months of the initial CML diagnosis. Patients were given imatinib, an inhibitor of the oncoprotein at fault for CML. Imatinib had already been proven successful in treating CML in adults in phase III trials. Millot et al. (2011) prescribed the pediatric CML patients imatinib to test if it worked as well for newly diagnosed pediatric patients. The trial was interventional, nonrandomized, and nonblinded. Adult dosage of imatinib had been established as 400mg/m^2 daily prior to the trial, and Millot et al.'s (2011) baseline pediatric dosage used in the trial was 260mg/m^2 daily. One measured outcome for the pediatric patients compared to adult patients was 'progression-free survival,' or the amount of time passed surviving without CML reemergence. Millot et al. (2011) reported that at 36 months post-intervention, the progression-free survival rate was 98%, close to previous studies on adults. Biological markers like hematological, cytogenetic, and molecular responses were also similar between the pediatric trial population and adults. The study

explicitly claims to confirm that the dosage and treatment were tolerable (Millot et al., 2011, 2831) and that imatinib could successfully treat pediatric CML. I take it this is an unproblematic example of an experiment, since there is an intervention (imatinib), a target system (pediatric CML patients), and measured outcomes (progression-free survival, biomarkers) that allow us to infer information about the relationship between the intervention and the target system (imatinib can treat newly diagnosed pediatric CML).

4.2. Evaluating Hypotheses & Establishing Effectiveness: Phase IV as Confirmatory

The pediatric imatinib case illustrates how phase IV trials can get us from evidence of drug efficacy to evidence of drug effectiveness. Imatinib was known to treat CML in adults in the late 1990s and early 2000s from pre-market trials. But as is common with pre-market trials, there was little knowledge about imatinib's ability to successfully treat pediatric patients due to the lack of inclusion of children in the trials. There was already some idea that imatinib *should* work for children based off its ability to inhibit the oncoproteins responsible for CML.[8] However, this is what philosophers of medicine call 'mechanistic' evidence of imatinib's therapeutic properties. Pharmaceutical treatment decisions based on mechanistic evidence alone without trial evidence is often suspect in contemporary medical practice, especially according to those sympathetic to evidence-based medicine (Howick 2011). Even some critics of evidence-based medicine attest that 'good' evidence for therapeutic effectiveness requires some integration of trial-based evidence alongside mechanistic evidence (Clarke et al. 2013). Millot et al. (2011) gives trial-based, empirical evidence demonstrating that imatinib does indeed treat CML in pediatric patients, adding to the knowledge of the 'real world' effectiveness of imatinib as a CML treatment.

Knowing that a drug is effective means knowing it works outside the context of the phase III RCT. Not knowing if imatinib worked for pediatric patients meant not having established or contributed to the knowledge about the effectiveness (or lack thereof) of the drug. Phase III RCTs had established the efficacy of imatinib or that imatinib 'worked somewhere': adults with CML. Going from efficacy to effectiveness meant evaluating the phase III-established efficacy in 'real world' contexts like in a phase IV trial. I think it is fair to call the efficacy established in phase III a kind of hypothesis or theory

[8] Knowing that imatinib functions by inhibiting oncoproteins is evidence of the drug's causal capacities, which can sometimes increase the external validity around some study. However, the availability of this kind of knowledge is not always the case with novel drugs, and not the kind of evidence provided by the RCT.

about the drug, and this hypothesis is what phase IV trials evaluate to determine effectiveness. In other words, some phase III trial or trials prove that a drug can treat an illness in ideal circumstances, and phase IV trials evaluate that hypothesis in different 'real world' circumstances. Since a phase IV trial like the one described happens under therapeutic care, there are clear conditions by which the hypothesis is confirmed: if the drug successfully treats what it is hypothesized to treat. So, phase IV trials like the one described are confirmatory experiments insofar as they do positively aim to evaluate a hypothesis or a prediction about a drug.

One might maintain that there is a disconnect between the efficacy established in phase III adult imatinib trials and what was evaluated in the Millot et al. (2011) case, since the phase III adult trials might say something specifically about adults. In that case, we could instead use an example of a phase IV trial that tests the same population as was tested in phase III, modifying exclusion or inclusion criteria or changing trial length. We might also ask if what the phase III trials had established, i.e., 'imatinib works somewhere,' was playing the role of a local auxiliary hypothesis as per Colaço (2018) in the pediatric imatinib study. That 'imatinib can treat CML by inhibiting specific oncoproteins' is a local theory of imatinib; is that being evaluated in the Millot et al. (2011) study? Seemingly, yes. If imatinib inhibited the specific oncoproteins at fault for CML in children but for some reasons did not successfully treat the children's CML, we would know something about the local theory – 'imatinib treats CML by inhibiting specific oncoproteins' – would be amiss. Consider though that it is not 'imatinib can treat CML by inhibiting specific oncoproteins' that is being tested, but 'imatinib can treat CML by inhibiting specific oncoproteins in children' that is being tested in the case above. 'Imatinib can treat CML by inhibiting specific oncoproteins' would in this case be auxiliary. Auxiliary hypotheses, per Colaço (2018), can be incidentally confirmed or denied in an experiment without that incidental confirmation being an aim of the experiment. But, even if that is the case, there is still an aim to confirm or evaluate a hypothesis in the Millot et al. (2011) case, i.e., the modified hypothesis including children. Either way it is correct to say that Millot et al. (2011) does explicitly aim to evaluate some hypothesis about imatinib and is a confirmatory experiment.

In sum of this section, the case of the pediatric imatinib trial above can help illustrate how phase IV trials move from the efficacy of a drug to the effectiveness of a drug. Imatinib was shown in phase III to treat CML, and by evaluating this efficacy in contexts outside the ideal phase III context, the phase IV trials added to our evidence about the drug's effectiveness. In this sense, phase IV trials do aim to

evaluate hypotheses or theories – the efficacy of some drug established in phase III – and the conclusions add to our knowledge about a drug's effectiveness. As such, phase IV trials like this can rightly be called a kind of confirmatory experiment. What now follows is how phase IV trials are concurrently exploratory experiments. With that established, we will then be able to re-think the common characterization of exploratory experimentation and its relation to confirmatory experimentation.

5. Phase IV as Exploratory

The previous section used a concrete example to show how phase IV trials can function as confirmatory experiments insofar as they can aim to evaluate a theory or hypothesis like the efficacy of a drug established in phase III trials. However, this is an incomplete picture of what phase IV trials are. Phase IV trials and pharmacovigilance in general are often construed as exploratory (Jones & Kingery 2014). Phase IV trials explicitly aim to uncover or explore for suspected side effects. Moreover, given the paucity of external validity that can surround a drug's efficacy, the exploratory and confirmatory functions of phase IV can be difficult to separate. Taken with section 4 the proceeding section shows how interventional phase IV trials as single experiments can blur the lines between exploratory and confirmatory experimentation.

5.1 Back to the Taxonomy

In addressing how phase IV trials like the imatinib case are exploratory it is worthwhile to return to Elliot's (2007) taxonomy of exploratory experimentation. The first dimension of the taxonomy was (1) the positive aims of the experiment. The second was (2) the role of theory or hypotheses, and the third (3) was how parameters are varied. Section 4 above addresses dimension (2), the role of hypotheses, in phase IV trials. Hypotheses or theories about a drug's efficacy can be explicitly tested in phase IV. Phase IV trials might also test hypotheses about drugs unrelated to previous trials, as in the case of investigating a suspected secondary therapeutic effect. This alone might show that phase IV trials as described just are confirmatory experiments. However, I aim to show that phase IV trials can fit on or address dimensions (1) and (3), showing that phase IV trials are indeed exploratory.

One of the aims of phase IV trials is to determine the safety of a therapeutic drug. Phase IV trials can explicitly aim to discover unknown, unpredictable phenomena that arise from an untested population taking a novel drug or novel combinations of drugs. Some of these phenomena are construed as

'adverse events' or 'adverse effects.' Because of phase IV's typically larger, more heterogenous trial populations and longer time scales, rarer or more diverse effects arise than in phase III trials. Phase IV trials can aim to discover potential side effects – positive and negative – that can arise in 'real world' settings. We often cannot foresee or predict side effects, though we have a good idea that all drugs have them. In the imatinib case, Millot et al. (2011) aimed to test a hypothesis but also aimed to monitor for and discover any possible side effects. They did find a side effect of imatinib in their pediatric patients: growth-stunting (Millot et al. 2009; Millot et al. 2014). So, phase IV trials fit securely on dimension (1) of the taxonomy since exploration and monitoring for unpredictable, unforeseen phenomena can be a positive aim of phase IV trials.

The relationship between phase IV trials and dimension (3) is more complicated. Fitting on Elliot's (2007) dimension (3) requires the variation of parameters or having methods by which parameters are varied. I think something analogous to parameter variation can occur in phase IV trials. Phase IV trials evaluate efficacy to establish effectiveness by testing outside the conditions or parameters of phase III. We might say the efficacy established in phase III relies on *ceteris paribus* conditions, and some philosophers have argued this (Bluhm 2007; Cartwright 2007). While phase III trials' efficacy rests on *ceteris paribus* conditions, phase IV trials vary these conditions or parameters by evaluating the efficacy in different, less 'stable' conditions. Even if some subgroup in phase IV was represented in phase III, things like change in time scale, comorbidities, or dosage change might be analogous to parameter variation.

Elliot discusses how parameter variation works in nanotoxicology, claiming that particles are tested at a variety of doses and in different biological systems in order to better understand the toxicity of the particles (2007, p. 319). One experiment can show how a toxin works in one instance, and then by changing the parameters of that instance (biological system, dosage, surface chemistry, etc.) more can be discovered about the toxin. We can imagine analogously a new drug being tested in a particular population. To better understand the therapeutic properties (like effectiveness) of that drug, we can change the parameters (test in a different population) and more can be discovered about the drug.

However, one might maintain that these cases are 'follow-up' studies rather than cases of parameter variation within a single experiment. This is intuitive, as phase IV studies are by their nature a kind of 'follow-up' study on market-approved drugs. If that is the case, it is less clear how a *single* phase IV

trial like the one described varies parameters. 'Variation' implies some relation between different things, since in varying a parameter that parameter already is somehow established. In that case, we can imagine that 'varying parameters' implicitly points to two distinct epistemic activities: measuring something at an initial time with an initial parameter and again at a different time with a different, 'varied' parameter. These might be two different experiments, where *between* the experiments the parameters are varied, but *within* the experiments it is not quite right to say parameters are varied. So, one might say that the imatinib phase IV study has different parameters than the phase III studies, but that is not saying the phase IV study 'varies' parameters. Thus, phase IV studies like those described might not, strictly speaking, fit on dimension (3).

I do think this is a reasonable point, however, this would seem to also hold for Elliot's (2007) example of parameter variation. I think two things can be said here: firstly, perhaps meeting all the taxonomy's criteria is not needed to classify experiments as exploratory. I find this plausible, since it might only be the case that an experiment's positive aims (e.g., Elliot's first dimension) suffices for something to be considered an exploratory experiment. However, I think there is a second response that can maintain Elliot's (2007) nanotoxicology example as something that intuitively fits dimension (3) while resolving this 'follow-up' concern with phase IV trials.

Since the context I am discussing is *single* experiments, and parameter variation implicitly brings in multiple epistemic activities, parameter variation might be *categorically unrelated* to whether a single experiment can be exploratory. Parameter variation is something about the exploratory nature of research programmes that are constituted by, among other things, *multiple* experiments. And, my point here is that single experiments can be concurrently exploratory and confirmatory. So, if one is to reject that phase IV trials can analogously fit dimension (3), that is not to say that single phase IV trials cannot be exploratory since dimension (3) is a criterion for programmes, not single experiments. Either way, phase IV trials as described can still be classified as exploratory (and concurrently confirmatory) experiments insofar as their positive aims. Moreover, we can imagine a kind of phase IV trial that *does* vary parameters within the trial/experiment, though this will look different than the imatinib case described.

5.2 Are Phase IV Trials Single Experiments?

One might maintain that phase IV trials as I have described them are something like 'composite' experiments. If a phase IV trial is just two experiments put together by convention, that would preclude the argument that phase IV trials can blur the lines between confirmatory and exploratory experimentation: this counterpoint would say phase IV trials are an exploratory experiment and a confirmatory experiment happening simultaneously, or 'put together'. Just like parameter variation implied something about two epistemic activities above, perhaps exploration and confirmation in the phase IV contexts are separate epistemic activities constituted by separate aims. We might try to imagine the two aims of the imatinib trial being split into two experiments. In one case, the drug is given to pediatric patients with the aim of seeing if it treats CML. In the other case, imatinib is given to the pediatric patients without measuring their CML symptoms, but close attention is paid to any suspected side effects.

In response, I think that establishing the effectiveness of a novel drug cannot occur without considering safety signals.[9] This is not something specific to phase IV trials and effectiveness. Drugs must be safe enough to pass phase I, II, and III stages to get to the market. Negative side effects that arise during the evaluation of a drug can preclude further experiment on it, at least for therapeutic purposes. Safety and effectiveness are conceptually related. Something will not 'work for us' if it causes more overall therapeutic harm than good. Establishing the effectiveness of a drug ideally means having discovered both its therapeutic effects and safety profile, i.e., side effects, outside phase III contexts. Therefore, the aims of a phase IV trial cannot be neatly separated in such a way as to identify two discrete experiments. Even if we could separate those aims into separate experiments, that does not mean that the 'fundamental condition' of exploration is correct; that there are experiments that both aim to test hypotheses and explore for unforeseen phenomena together, even if merely conventionally, challenges the 'fundamentality' of the condition.

Moreover, remember that exploration's positive aims are about uncovering unknown, unforeseen phenomena. When we have a hypothesis that we know has weak external validity and we test it in the chaotic 'real world', is there some investigation into unforeseen phenomena going on? Say I have some

[9] Other than the reasons around just what 'effectiveness' entails, there are bioethical reasons as well that these two aims should not be considered possibly separable; therapeutic outcomes should not be ignored when a drug is given under therapeutic care.

drug *D*. I might know *D* has efficacy in a small homogenous population and my evidence has weak external validity. In giving *D* to a larger more heterogeneous population, is the act of determining *D*'s effectiveness by evaluating its efficacy in the 'real world' a kind of exploration among unforeseen phenomena? Say I run two phase IV trials on *D*. In the first phase IV trial, the population is the same or highly similar to the homogenous population from phase III. In the second phase IV trial, the population is completely different and more diverse and co-morbid than the phase III population. In both phase IV trials I test the same hypothesis and they are both confirmatory insofar as that efficacy is evaluated. It seems that I have *less* justification in predicting what will happen in the second phase IV trial than the first. In the second trial, the relationships among the phenomena are not as well-known to me or predictable as in the first trial. There is more going on or more 'known unknowns' because of how different the contexts are between where that efficacy was established and where it is now being tested.

In the second more diverse case, there seems to be more uncertainty. Does that mean it is more 'exploratory' than the first experiment? Perhaps not. As I will discuss in the next section, the amount of uncertainty *alone* does not entail a distinction between exploratory and confirmatory experimentation. However, the second case does lend itself or allows for more robust possible exploration. The aim to confirm and the aim to explore will remain distinct aims. My point is only to highlight that these aims are likely not so easily separable, not only in the case of phase IV trials explicitly exploring for possible side effects, but also in light of how phase IV trials determine effectiveness by evaluating efficacy in the 'real world'.

6. Confirmatory or Exploratory: The Distinction Revisited

I do not think the above entails that all confirmatory experiments are to some degree exploratory. Consider the scope of the argument presented. Defining exploratory experimentation as non-hypothesis-testing is overly exclusive insofar as some experiments like phase IV trials as described are exploratory and do aim to test hypotheses. Some experiments are exploratory, some are confirmatory, and some are both. However, the question can remain about how we distinguish cases of solely confirmatory experiments from cases of solely exploratory experiments.

As above, uncertainty is inherent in all forms of experiment, but I aim to show that alone it is insufficient to distinguish solely confirmatory experiments from solely exploratory experiments. In

virtue of doing an experiment on some phenomenon, I am looking to gain information about it I did not know before. If this is right, there might be no such thing as a 'purely' confirmatory experiment, since we acknowledge we cannot be certain about the relationships among the phenomena we test. In that sense, even in simple prediction we are looking to uncover 'unknown' relationships in or about our target phenomenon. That there are no solely confirmatory experiments would be, admittedly, an odd conclusion. Though I do not think it would be a damning one. The discussion about the inseparability of the confirmatory and exploratory functions of experiment in phase IV in the previous section might even be construed as evidence for this. However, I think that we can keep these different kinds of experimentation distinct by showing the presence of uncertainty is alone insufficient to qualify an experiment as exploratory.

Consider an example that is intuitively, solely confirmatory. Say I have two well-studied chemical compounds, and I know each of their structural formulas. I might predict what the structural formula of their combination would be in some context. I make the prediction, somehow combine the compounds, and then evaluate if the resulting compound's structural formula matches my prediction. Here I am explicitly testing or evaluating a prediction, and I am quite certain in my prediction.

But do I *know* that my prediction will be correct about the chemical formula? There is inherently some uncertainty beforehand. So maybe even in that case there is some 'exploration' occuring? I want to hesitate against this conclusion. We can distinguish a specific kind of address or attitude towards uncertainty that preserves the possibility of solely confirmatory experiments. This would keep the distinction between confirmatory and exploratory experiments tenable. I posit that exploratory experiments engage with uncertainty in a distinct way that is different than how confirmatory experiments do. In the chemical compound case, the uncertainty comes about through the usual suspects: induction, underdetermination, and human error. In the phase IV case, these same causes of uncertainty are also there. We might say that the uncertain things (e.g., if a drug's effectiveness will be close to the assumed efficacy or what side effects might occur) are the explicit targets of the phase IV trial. But isn't the 'target' of the structural formula case also something we do not know about *a priori*? Doesn't an experiment by its very nature try to show something previously unknown about a phenomenon and its relationships?

Uncertainty alone then is not a feature that distinguishes exploratory from confirmatory exploration. It is the explicit aim of *monitoring* for unforeseen and unpredicted phenomena or relations between phenomena given some uncertainty that makes an exploratory experiment different than a solely confirmatory experiment. There will be confirmatory experiments that do not aim to uncover any unforeseen phenomena. And, there will be exploratory experiments that do not aim to test hypotheses. The important thing to take away is that exploration is exploratory *not because* it does not aim to test hypotheses, but because it aims to discover unforeseen phenomena. Only then does it make sense to say an experiment can be both exploratory and confirmatory. Phase IV trials as described show that the fundamental condition of exploratory experimentation, that exploration does not aim to tests hypotheses, is overly restrictive on what kinds of experiments can count as exploratory.

What follows from this way of thinking about the relationship between confirmatory and exploratory experimentation is that they are not diametrically opposed to one another or along a continuum, as is sometimes thought (O'Malley 2007, p. 349). Confirmatory and exploratory experimentation do not exist along a continuum separated by 'degrees' of hypothesis-testing. As above, 'hypothesis-testing' can be instantiated by an exploratory experiment just as much as by a confirmatory experiment. It is just that a solely exploratory experiment might not evaluate a hypothesis, and a solely confirmatory experiment might not monitor for unforeseen phenomena. Therefore, I propose that the conceptual relation between exploratory and confirmatory experimentation is less like points along a continuum, and more like figure 1 below. Exploration and confirmation are discrete kinds of experimentation, not points along a line. They can co-exist, as in phase IV trials. That means that it is not strictly correct to think that no exploratory experiments can aim to evaluate a hypothesis. Thinking about the relationship in this way does not preclude single cases where exploration and confirmation overlap, as was the case with the common characterization of exploratory experimentation as non-hypothesis-testing. Fig. 1 shows that given the relationship between confirmatory and exploratory experimentation as not dichotomous, there are four kinds of experiment we can categorize.

	Exploratory	~Exploratory
Confirmatory	(1)	(3)
~Confirmatory	(2)	(4)

Fig. 1: Posited Relation Between Confirmatory and Exploratory Experimentation

Figure 1 gives a table showing the relationship between 'exploratory' and 'not exploratory' (with the negation represented by '~') and 'confirmatory and 'not confirmatory' experiments. We can think of something being categorized as exploratory when there are positive aims to monitor unforeseen phenomena. We can think of something as confirmatory when there are positive aims to evaluate or test a hypothesis. This gives 4 different categories of experiments: (1) experiments that are both confirmatory and exploratory, (2) experiments that are exploratory and not confirmatory, (3) experiments that are confirmatory but not exploratory, and (4) experiments that are neither exploratory nor confirmatory. In what follows I use Arabic numerals in discussing the categories on fig. 1, and Roman numerals in discussing the phases of clinical trials.

The kinds of experiments in (1) would be cases of single experiments that are both confirmatory and exploratory, i.e., experiments that have the concurrent aims of evaluating a hypothesis and uncovering unforeseen phenomena. I have used interventional phase IV trials to demonstrate proof of these cases. However, assumedly, phase IV trials are not the only kinds of experiments that would fit within (1). Some phase II/III trials may fit in (1). Some oncology phase I trials might also fit in (1). Whether or not some type of experiment fits in any category will change case-by-case and depend on the experiment's aims. It is also likely the case that many other kinds of experiments on complex systems also fit in (1). I do not think the arguments here are idiosyncratic to the phase IV context. Instead, it

is likely that an analysis of any experiment in category (1) could suffice to prove the point of this paper and show the need to recharacterize exploration and its relation to confirmatory experimentation.

Category (2) would be the exploratory experiments that do not evaluate hypotheses, like those discussed by the philosophers who address exploratory experimentation cited above. Background theories or local auxiliary theories can be present in category (2) experiments but are not the targets of evaluation. Category (3) experiments would be cases of solely confirmatory experimentation with no aims to discover unforeseen phenomena, like the case of the chemical formula above. Also in (3) might be the paradigm cases of confirmatory or 'critical' experiments that have contributed to the historical acceptance and rejection of theories. Finally, category (4) contains experiments that are neither confirmatory nor exploratory. These might be sparse in scientific practice but common in scientific education, though not inclusive of all experiments done in an educational setting. Mixing baking soda and vinegar at the science fair or dissecting a frog in a classroom probably does not play a confirmatory or exploratory role in the technical senses often meant, though I do not think this disqualifies them as experiments. This allows us to maintain that experiments in the classroom are still experiments, even if they are only meant to show students something like how an instrument works without necessarily exploring for some unknown phenomena or confirming a prior hypothesis. Other kinds or types of experimentation could be added to the table to more fully categorize different categories of experimentation as well.

Finally, we can ask what the value of the distinction is at all. The natural world reliably escapes our best attempts to classify, categorize, and explain it. It should be no surprise that our methods of investigating the world are also messy. Philosophers have pointed this out about approaches to scientific methods generally (Waters 2019) and specifically in the context of exploratory experimentation (Schickore 2016). Concepts like 'exploratory experimentation' do not pick out something like a natural kind of thing in the world; it is just our attempt to understand the aims, processes, and practice of science. Labels like 'exploratory' might be merely heuristics to understand why some scientific practice happens. Schickore (2016) points out that the concept of exploratory experimentation has enriched our understanding of experimentation but can be vague. Exploratory experimentation can serve as a heuristic to better characterize different types of experimental practice (Schickore 2016, p. 23-24).

I do not argue against this response. Even if confirmatory and exploratory experimentation are mere heuristics, having a clear idea of where their boundaries end and overlap can still be of value. How the various aims of experimentation are interrelated is an issue worth being clear on. I have demonstrated that phase IV trials are a member of a more general class of experiment that can serve as an example of these overlaps where the lines between confirmatory and exploratory experimentation blur and require us to re-think their relationship.

7. Conclusion

Phase IV trials are tools in pharmacovigilance that monitor the ongoing use and safety of drugs. Some phase IV trials belong to a class of experiments that are jointly confirmatory and exploratory. Since exploratory experiments are often characterized as experiments that do not aim to evaluate hypotheses, and phase IV trials are exploratory experiments that *do* aim to evaluate a hypothesis, we need to re-think exploration's common characterization as non-hypothesis-testing and its relationship to confirmatory experimentation. Philosophers of science have pointed out that these two modes of experimentation can co-exist within research programmes. What I have presented above shows how single experiments can be characterized as both confirmatory and exploratory. Contrary to the common characterization of exploratory experiments, some cases of exploratory experimentation like phase IV trials do involve explicitly evaluating some hypothesis. Therefore, the relationship between exploration and confirmation in experimentation is not one of opposition along a continuum of degrees of hypothesis-testing. This preserves the belief that some experiments can be confirmatory, some can be exploratory, and some can be both. Maintaining this belief with the characterization of exploration as non-hypothesis-testing was problematic. The account argued for above makes that belief tenable. The recharacterization of the relationship between confirmatory and exploratory experimentation can also include other types of experimentation for more detailed classification. Finally, even if 'exploratory experimentation' is something like a heuristic, it is still useful to clarify cases of overlap and its relation to other heuristics like confirmatory experimentation.

References

Bernabe, R. D. L. C. (2016). Ethical Principles in Phase IV Studies. In D. Strech & M. Mertz (Eds.) *Ethics and Governance of Biomedical Research* (p. 97-107). Springer International. https://doi.org/10.1007/978-3-319-28731-7_8.

Bernabe, R. D. L. C., van Thiel, G. J. M. W., Raaijmakers, J. A. M., & van Delden, J. J. M. (2014). The fiduciary obligation of the physician-researcher in phase IV trials. BMC Medical Ethics, 15, 11. https://doi.org/10.1186/1472-6939-15-11.

Bluhm, R. (2007). Clinical trials as nomological machines: Implications for evidence- based medicine. In H. Kincaid, & J. McKitrick (Eds.), Establishing medical reality (pp. 149–166). Springer.

Burian, R. (1997). Exploratory experimentation and the role of histochemical techniques in the work of Jean Brachet, 1938-1952. History & Philosophy of the Life Sciences, 19, 27–45.

Cartwright, N. (2011). A philosopher's view of the long road from RCTs to effectiveness. The Lancet, 377, 9775. https://doi.org/10.1016/S0140-6736(11)60563-1 Cartwright, N. (2007). Are RCTs the gold standard? BioSocieties, 2(1), 11–20.

Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2013). Mechanisms and the evidence hierarchy. Topoi, 33, 339–360.

Colaço, D. (2018). Rethinking the role of theory in exploratory experimentation. Biology and Philosophy, 33, 38. https://doi.org/10.1007/s10539-018-9648-9

Elliot, K. C. (2007). Varieties of exploratory experimentation in nanotoxicology. History & Philosophy of the Life Sciences, 29(3), 174–197.

Fisher, J. A. (2020). Adverse events: Race, inequality, and the testing of new pharmaceuticals. New York: NYU Press.

Franklin, L. R. (2005). Exploratory experiments. Philosophy of Science, 72, 888–899.

Gaffney, A., & Lexchin, J. (2018). Healing an ailing pharmaceutical system: Prescription for reform for United States and Canada. The BMJ, 361, k1039. https://doi.org/ 10.1136/bmj.k1039

Hacking, I. (1983). Representing and intervening: Introductory topics in the philosophy of natural science. Cambridge: Cambridge University Press.

Howick, J. (2011). Exposing the vanities – and a qualified defense – of mechanistic reasoning in health care decision making. Philosophy of Science, 5, 926–940.

Jones, J. K., & Kingery, E. (2014). History of pharmacovigilance. In E. B. Andrews, & N. Moore (Eds.), Mann's pharmacovigilance 3rd edition (pp. 11–24). Wiley & Sons, Inc.

Kessler, D. A., Rose, J. L., Temple, R. J., Schapiro, R., & Griffin, J. P. (1994). Therapeutic-class wars–drug promotion in a competitive marketplace. New England Journal of Medicine, 331, 1350–1353. https://doi.org/10.1056/NEJM19941117 3312007

Lexchin, J. (2016). Private profits versus public policy: The pharmaceutical industry and the Canadian state. University of Toronto Press.

Light, D. W., Lexchin, J., & Darrow, J. J. (2013). Institutional corruption of pharmaceutical and the myth of safe and effective drugs. Journal of Law Medicine & Ethics, 41(3), 590–600.

May, F. (2014). Teaching and learning pharmacovigilance. In E. B. Andrews, & N. Moore (Eds.), Mann's pharmacovigilance (3rd ed., pp. 793–804). Wiley & Sons, Inc.

Millot, F., Baruchel, A., Guilhot, J., Petit, A., Leblanc, T., Bertrand, Y., ... Guilhot, F. (2009). Imatinib is efficient but has a negative impact on growth in children with previously untreated chronic myelogenous leukemia (CML) in early chronic phase (CP): Results of the French national phase IV trial. Blood, 114, 22. https://doi.org/ 10.1182/blood.V114.22.863.863

Millot, F., Baruchel, A., Guilhot, J., Petit, A., Leblanc, T., Bertrand, Y., ... Guilhot, G. (2011). Imatinib is effective in children with previously untreated chronic myelogenous leukemia in early chronic phase: Results of the French national phase IV trial. Journal of Clinical Oncology, 29(20), 2827–2832.

Millot, F., Guilhot, J., Baruchel, A., Petit, A., Leblanc, T., Bertrand, Y., ... Rialland, F. (2014). Growth deceleration in children treated with imatinib for chronic myeloid leukemia. European Journal of Cancer, 50, 3206–3211.

O'Malley, M. (2007). Exploratory experimentation and scientific practice: Metagenomics and the proteorhodopsin case. History & Philosophy of the Life Sciences, 29(3), 335–358.

Osimani, B. (2014). Hunting side effects and explaining them: Should we reverse evidence hierarchies upside down? Topoi, 33, 295–312.

Schickore, J. (2016). 'Exploratory experimentation' as a probe into the relation between historiography and philosophy of science. Studies in History and Philosophy of Science, 55, 20–26. https://doi.org/10.1016/j.shpsa.2015.08.007

Schwartz, J. L. (2017). Real-world evidence, public participation, and the FDA. Hastings Center Report, 47(6), 7–8.

Sismondo, S. (2008). How pharmaceutical industry funding affects trial outcomes: Causal structures and responses. Social Science & Medicine, 66, 1909–1914.

Sismondo, S. (2009). Ghosts in the machine: Publication planning in the medical sciences. Social Studies of Science, 39(2), 171–198.

Solomon, M. (2017). After disclosure. In A. LaCaze, & B. Osimani (Eds.), Uncertainty in pharmacology: Epistemology, methods, and decisions (pp. 439–450). Springer Nature.

Stegenga, J. (2016). Hollow hunt for harms. Perspectives on Science, 24(5), 481–504.

Steinle, F. (1997). Entering new fields: Exploratory uses of experimentation. Philosophy of Science, 64, S65–S74.

Suvarna, V. (2010). Phase IV of drug development. Perspectives in Clinical Research, 1.2, 57–60.

Vandenbroucke, J. P. (2008). Observational research, randomized trials, and two views of medical science. PLoS Medicine. https://doi.org/10.1371/journal.pmed.0050067

Waters, C. K. (2019). Presidential address, PSA 2016: An epistemology of scientific practice. Philosophy of Science, 86, 585–611

Waters, C. K. (2007). The nature and context of exploratory experimentation: An introduction to three case studies of exploratory research. History & Philosophy of the Life Sciences, 29(3), 275–284.

Worrall, J. (2008). Evidence and ethics in medicine. Perspectives in Biology and Medicine, 51(3), 418–431.https://doi.org/10.1353/pbm.0.0040.

Worrall, J. (2002). What evidence in evidence-based medicine? Philosophy of Science, 69(S3), 316–330.