

METHODOLOGY

Open Access

Accumulating Evidence and Research Organization (AERO) model: a new tool for representing, analyzing, and planning a translational research program

Spencer Phillips Hey^{1*}, Charles M Heilig² and Charles Weijer³

Abstract

Background: Maximizing efficiency in drug development is important for drug developers, policymakers, and human subjects. Limited funds and the ethical imperative of risk minimization demand that researchers maximize the knowledge gained per patient-subject enrolled. Yet, despite a common perception that the current system of drug development is beset by inefficiencies, there remain few approaches for systematically representing, analyzing, and communicating the efficiency and coordination of the research enterprise. In this paper, we present the first steps toward developing such an approach: a graph-theoretic tool for representing the Accumulating Evidence and Research Organization (AERO) across a translational trajectory.

Methods: This initial version of the AERO model focuses on elucidating two dimensions of robustness: (1) the consistency of results among studies with an identical or similar outcome metric; and (2) the concordance of results among studies with qualitatively different outcome metrics. The visual structure of the model is a directed acyclic graph, designed to capture these two dimensions of robustness and their relationship to three basic questions that underlie the planning of a translational research program: What is the accumulating state of total evidence? What has been the translational trajectory? What studies should be done next?

Results: We demonstrate the utility of the AERO model with an application to a case study involving the antibacterial agent, moxifloxacin, for the treatment of drug-susceptible tuberculosis. We then consider some possible elaborations for the AERO model and propose a number of ways in which the tool could be used to enhance the planning, reporting, and analysis of clinical trials.

Conclusion: The AERO model provides an immediate visual representation of the number of studies done at any stage of research, depicting both the robustness of evidence and the relationship of each study to the larger translational trajectory. In so doing, it makes some of the invisible or inchoate properties of the research system explicit – helping to elucidate judgments about the accumulating state of evidence and supporting decision-making for future research.

Keywords: Translational medicine, Research efficiency, Graph-theoretic model, Robustness, Moxifloxacin, Tuberculosis, Research coordination, Research planning, Decision-making

*Correspondence: spencer.hey@mcgill.ca

¹ Studies for Translation, Ethics, and Medicine (STREAM) Group, Biomedical Ethics Unit, McGill University, Montreal, QC H3A 1X1, Canada
Full list of author information is available at the end of the article

Background

Maximizing efficiency across a drug development trajectory is important for drug developers, policymakers, and human subjects. Limited resources and the ethical imperative of risk minimization demand both that researchers design and plan their studies to maximize the knowledge gained per patient-subject enrolled and that funders and ethical review boards hold them to this standard.

Unfortunately, this standard is often not met. The infamous failure of torcetrapib at phase 3 illuminated inefficiencies in cardiovascular drug development [1-3]. In cancer, the cases of sunitinib for treating hepatic cancer and bevacizumab for treating gastric cancer were both examples of drugs that advanced into phase 3 testing without supporting phase 2 evidence [4]. Across the entire spectrum of drug development, nearly one-third of the drugs abandoned at phase 2 are considered failures not for lack of efficacy, but for 'strategic' reasons [4]. Even within a single company, as Pfizer's internal review showed, almost half of their phase 2 proof-of-concept studies (43%) failed to test the target mechanism of action adequately [5].

These problems all contribute to the common perception that the current system of drug development is beset by inefficiencies [4,6,7], and the growing number of calls for greater coordination across the drug development enterprise [8-10]. Yet, there remain no systematic approaches for representing, analyzing, and communicating the efficiency and coordination of the research enterprise. The available techniques of systematic review and statistical meta-analysis can only tell a part of this story – elucidating partial cross sections of the evidence – but they do not offer any synthesis for the accumulating state of evidence across the entire translational trajectory. For example, a meta-analysis is well suited to identifying trends across a series of similar trials, but it is useless for understanding transitions between the different phases of research. How did the research program progress from pre-clinical to clinical studies? How did the results of phase 1 studies compare with phase 2? What does the failure of translation from pre-clinical to clinical tell us about what we ought to do next? These kinds of questions are entirely consistent with the aim of comprehensively understanding the translational research enterprise, but they are not questions for which a statistical meta-analysis is helpful.

In this paper, we present a first step toward developing a broader, more comprehensive approach to representing and understanding a translational research program: A graph-theoretic tool for representing the Accumulating Evidence and Research Organization (AERO) within a translational trajectory. We begin in the next section by describing the basic methodology – the components of the AERO model and how these are used to construct an AERO graph. We then apply this tool to a case

study involving the antibacterial agent, moxifloxacin, for the treatment of drug-susceptible tuberculosis. We close by discussing the potential utility of our approach for improving the organization, coordination, and efficiency of the drug development enterprise.

Methods

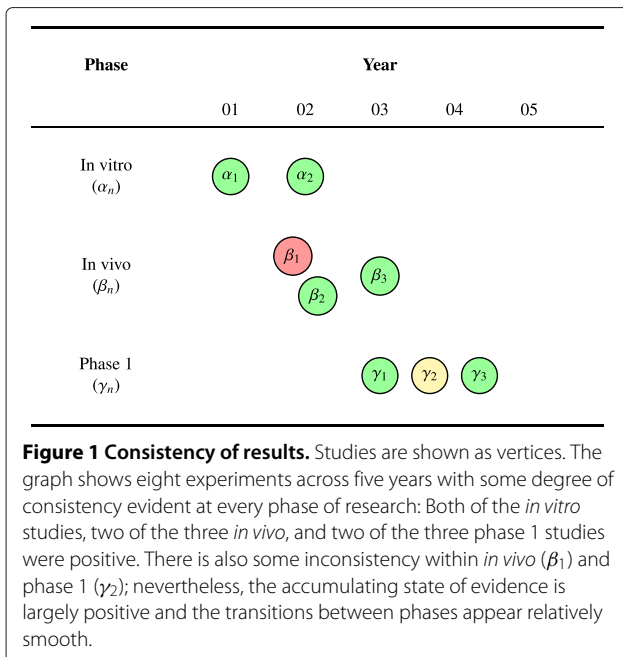
The conceptual foundations of the AERO model emerged out of work in the philosophy of science to develop a general methodology for representing robust scientific evidence [11]. Although there are many different dimensions of robustness that figure in drug development, the version of the AERO model we present here focuses on elucidating just two of these: (1) the consistency of results among studies with an identical or similar outcome metric; and (2) the concordance of results among studies with qualitatively different outcome metrics. For example, a series of animal experiments that all show a similar effect size for a new drug is evidence of consistency. A positive direction of effect on both the surrogate outcome used in a phase 2 trial and the clinical outcome used in a phase 3 trial is evidence of concordance.

The visual structure of the AERO model is a directed acyclic graph (DAG), designed to capture these two dimensions of robustness and their relationship to three basic questions that underlie the planning of a translational research program: What is the accumulating state of total evidence? What has been the translational trajectory? What studies should be done next? We discuss each of these questions in turn, showing how they are represented in the model.

Representing the accumulating state of total evidence

The accumulating state of evidence in a drug development trajectory is constituted by discrete experiments, proceeding from the pre-clinical *in vitro* and *in vivo* experiments to the phase 1, 2 and 3 clinical trials. In the AERO model, we represent each experiment as a node (that is, a vertex) arranged in a two-dimensional space: an x-axis, representing time, and a y-axis, representing the phase of research. The nodes are then color-coded according to the direction of their outcome: studies in support of further research (for example, positive results) are green, studies against further research (for example, negative results) are red, and studies ambivalent toward further research (for example, inconclusive results) are yellow.

Figure 1 is an example AERO graph for a research program with eight experiments across three phases in a five-year span. Although this graph is incomplete, since it does not yet include the arrows (that is, the edges) to illustrate the translational trajectories, it nevertheless captures some features of the research program that are essential to understanding the state of total evidence. For example, we can see that the translation from animals into



humans (that is, *in vivo* into phase 1) is relatively smooth. Two positive animal studies (β_2, β_3) suggest a potential for efficacy in humans and two positive phase 1 studies (γ_1, γ_3) suggest a well-tolerated drug. There is also evidence of consistency at each phase (α_1 and α_2 , β_2 and β_3 , γ_1 and γ_3). Given that consistency between experiments serves to verify findings and control for biases or random errors, which may distort the results of any single experiment, this is a desirable feature for the system to have.

Figure 1 also shows some evidence of concordance. The overall trend of the experimental findings is largely positive across each phase. With the exception of one negative *in vivo* study and one inconclusive phase 1 study, the results have all been favorable to further research. This degree of concordance is another desirable feature of the system. Depending upon the predictive power of the animal model, this pattern of robust results provides good reason for thinking that the experimental agent may be efficacious and safe.

Finally, assuming that β_2 is later than α_2 and γ_1 is later than β_3 , we can see that there was a *de facto* threshold of two positive studies at each phase before proceeding to the next. These thresholds between phases are an important property of the translational trajectory, representing critical (and often expensive) decision points: When is the pre-clinical evidence sufficient to initiate human trials? When is the phase 2 evidence sufficient to justify a pivotal phase 3 trial? Many translational failures (like those mentioned above) can be understood as the result of poor or inappropriate thresholds, wherein the evidence was not

sufficiently mature to warrant advancing a candidate to a later phase of research.

One plausible way to set phase thresholds is to require a certain number of positive studies at each phase before proceeding to the next. Indeed, the US FDA's licensure requirement for two positive phase 3 trials can be thought of as just such a threshold. However, two positives is not the only viable threshold. For example, a certain number of negative studies within a phase could require either that an agent be sent back to an earlier phase or abandoned entirely. It may also be reasonable to require that there is a sufficient number of both positive and negative studies at each phase – the positive studies supporting the efficacy of the therapeutic ensemble for clinical translation and the negative studies supporting a theoretical evidence base that directly informs clinicians about how the experimental agent should not be used [12]. Leaving aside the question of which phase threshold to use, either in general or for some specific research domain, it is sufficient for the purposes here to observe that the AERO graph has the virtue of rendering each of these thresholds visually explicit.

Representing the translational trajectory

In Figure 2, we expand on Figure 1, introducing three phase 2 studies as well as arrows between studies. The arrows represent more precisely the sequence of studies

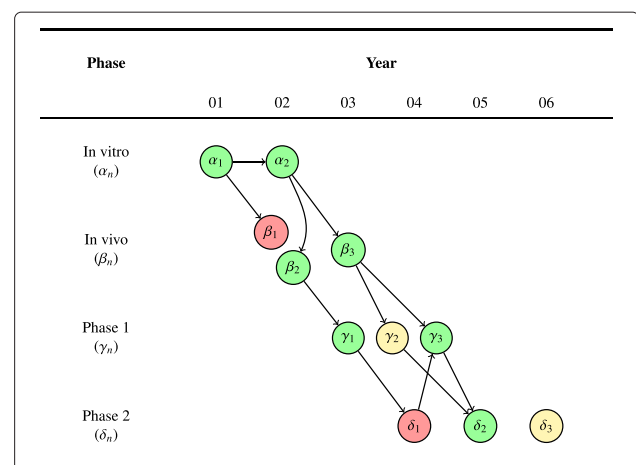


Figure 2 Concordance in a trajectory. Edges show the intellectual lineage. The graph shows eleven experiments across six years. The edges between studies represent the intellectual lineage between them and illustrate the translational research trajectories. Some trajectories are perfectly concordant (for example, $\alpha_1 \rightarrow \alpha_2 \rightarrow \beta_3 \rightarrow \gamma_3 \rightarrow \delta_2$), while others show discordance (for example, $\alpha_1 \rightarrow \beta_1$ or $\alpha_1 \rightarrow \alpha_2 \rightarrow \beta_2 \rightarrow \gamma_1 \rightarrow \delta_1$). The phase 2 studies are also highly inconsistent (that is, one positive, one negative, and one inconclusive study), indicating a relatively rough transition into this phase. Given that phase 2 was initiated after only one positive phase 1 trial, this may indicate that the threshold of evidence used to transition into phase 2 is too low and ought to require at least some degree of consistency.

and capture the intellectual lineage across the translational trajectory. For example, a phase 2 study that uses the same dosage identified in a phase 1 study should be connected by an arrow leading out from the phase 1 node and into the phase 2 node (for example, γ_1 to δ_1). Similarly, a phase 1 human pharmacokinetic and pharmacodynamic study should be connected to the prior *in vivo* study that identified the effective blood concentration (for example, β_3 to γ_2 and γ_3). Borrowing terminology from the language of graph theory, we refer to studies downstream in the intellectual lineage as ‘children,’ and studies upstream as ‘parents’ (we will have more to say about how parentage is established, when we discuss the example in the next section).

This additional structure enriches judgments about consistency and concordance. Where before we could only suggest concordance by the number or ratio of positive results at each phase, now we can track a discrete translational trajectory. For example, the trajectory from α_1 to α_2 to β_2 to γ_1 is a sequence of four positive studies across three phases. Each of these studies built on the evidence in the prior study, and each showed a concordant result.

Notice, however, that the story of this trajectory gets more complicated when we follow it into phase 2. The first phase 2 study, δ_1 , was negative, despite following up the evidence from γ_1 . But negative results are not necessarily uninformative results. Thus, as the figure shows, a later positive phase 1, γ_3 , built upon δ_1 , and eventually led (in conjunction with β_3), to the positive phase 2 result in δ_2 . The complete story of this trajectory, while not one of *perfect* concordance or consistency, can nevertheless be described as favorably robust overall.

Figure 2 also reveals that there was only positive phase 1 study (γ_1) before phase 2 research began (that is, γ_3 is subsequent to δ_1). Again, this represents a *de facto* (if not explicit) judgment about the minimum threshold of evidence necessary to justify proceeding to the next phase. Since consistency requires at least two experiments, advancing into the next phase of research on the basis of only one experiment means that this minimum threshold included no evidence of consistency at the immediately preceding phase. For exactly the same reasons that consistency is desirable, failing to have consistency before proceeding to a subsequent phase is, in general, an undesirable feature of a research system: No verification of findings, no control for bias or random error. While a single, well-designed study may be sufficient, under particular conditions, to advance an intervention, the recent failed attempts by Bayer [13] and Amgen [14] to reproduce earlier findings demonstrates the dangers with such an approach.

Finally, we should point out the orphaned study, δ_3 . This node has no arrows leading into it, reflecting an

experiment that is not directly justified on the basis of prior evidence within the research program. Although the general rule is probably to avoid conducting such studies, there may at times be reasons to draw on evidence or study designs that are external to the current research program. For example, the experiment may be largely based on analogical reasoning, drawing on evidence with a different drug or a different indication (the phase 3 trials of sunitinib and bevacizumab alluded to above could be described in just this way). The AERO graph makes this design choice explicit – the prudence of which will have to be judged by the fruitfulness of the subsequent research.

Planning the next step

The final, and perhaps most important, question we want to address with the AERO graph is: ‘What study should be done next?’ Given the state of accumulating evidence and the current trajectory, what should be the next investigation(s)? In Figure 3, we have added a single, negative phase 3 study, ϵ_1 , following the trajectory from δ_2 . We have also added two blue-lined nodes to represent the contemplated next steps.

Herein lies the real power of the AERO graph: In debating what study should come next, the visualization can sharpen the judgments of researchers, who can now pinpoint the precise translational trajectory or subset of the trajectory that they think ought to inform future research.

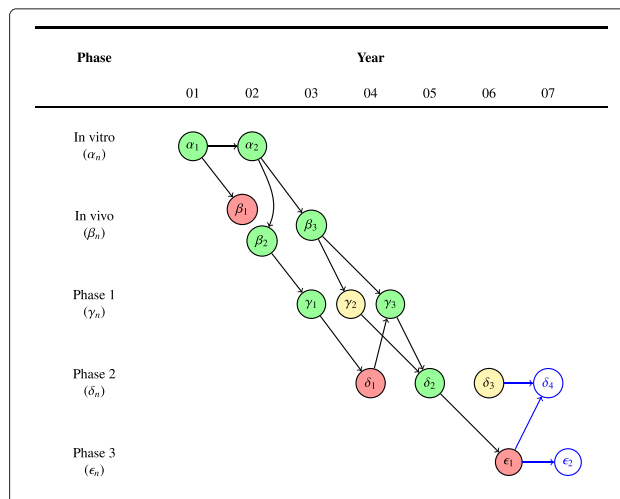


Figure 3 Planning future studies. The graph shows twelve completed experiments across six years along with two contemplated future studies, a fourth phase 2 trial (δ_4) and a second phase 3 trial (ϵ_2). A phase 3 trial (ϵ_1) was initiated following the sole positive phase 2 study (δ_2). The result of this phase 3 trial was negative, discordant with the earlier phase 2 result. Now researchers must decide which study (or studies) to do next: Trust that the accumulated evidence is still sufficient to motivate another phase 3 or return to phase 2 in search of greater consistency and a potential explanation for the discordance between δ_2 and ϵ_1 .

For example, let us suppose that ϵ_1 was negative for lack of clinically significant effectiveness. A researcher could argue that despite the negative result, there is still a largely positive trend (seven out of twelve studies) across the entire trajectory, the drug is very well tolerated (no negatives at phase 1), and the inconsistency across phase 2 studies has been instructive in suggesting a novel variation on the dosage and schedule that ought to be evaluated in a subsequent phase 3 trial. Yet, a different researcher may look at the same graph, identify a mechanism of action common to the inconclusive and negative studies that could explain these results, and recommend δ_4 to test this hypothesis in a less costly, phase 2 trial.

The purpose of the AERO graph is thus not to replace this kind of critical thinking or eliminate disagreement about the state of total evidence. Indeed, researchers may even disagree about whether a particular study ought to be represented as positive or negative. Rather, the purpose of the AERO graph is to sharpen this disagreement by clarifying the state of evidence and translational trajectory. In other words, the prudent course of action within a clinical research program cannot be derived from the state of evidence. The decision of what to do next requires a negotiation between the practical, pragmatic, ethical, and epistemic issues at play. The representational features of the AERO model help to make the epistemic aspects of this judgment explicit.

Results

In its *Global Plan to Stop TB 2011–2015*, the Stop TB Partnership emphasizes the need for improving coordination across the entire development and testing trajectory [10]. Moxifloxacin, an antibacterial agent in the fluoroquinolone family, is one of the new candidate drugs in their pipeline. However, as of 2010, the tuberculosis research community was confronted with a series of inconsistent results across five phase 2 studies with moxifloxacin, and disagreed about whether the state of total evidence supported moving on to conduct phase 3 trials. The results we present in this section are based on our presentation at the Tuberculosis Trials Consortium’s semi-annual meeting in October 2011, where we presented an AERO graph of moxifloxacin’s translational trajectory to help broker this dispute.

Representing the evidence

Inclusion

Table 1 summarizes the results of 19 studies evaluating moxifloxacin for the treatment of drug-susceptible tuberculosis between 1998 and 2009. These studies were included based on a hand search through the citations of the published phase 2 trial reports. This list was then cross-referenced with a PubMed search using the terms ‘moxifloxacin’ and ‘tuberculosis,’ filtered by clinical trials.

Table 1 Summary of moxifloxacin-TB studies

<i>In vitro</i>	Year	Key	Outcome	Children
Ji et al. [15]	1998	u_1	Positive	v_1
Gillespie et al. [16]	1999	u_2	Positive	w_1, w_2
Shandil et al. [17]	2007	u_3	Positive	w_5, w_6
<i>In vivo</i>				
Ji et al. [15]	1998	v_1	Positive	v_2
Miyazaki et al. [18]	1999	v_2	Positive	v_3, w_1
Lounis et al. [19]	2001	v_3	Positive	v_4
Yoshimatsu et al. [20]	2002	v_4	Inconclusive	v_5
Nuermberger et al. [21]	2004	v_5	Inconclusive	u_3, w_3, w_4
Phase 1				
Gosling et al. [22]	2003	w_1	Positive	v_5
Pletz et al. [23]	2004	w_2	Positive	v_5
Gillespie et al. [24]	2005	w_3	Negative	x_1
Johnson et al. [25]	2006	w_4	Positive	x_1, x_2
Nijland et al. [26]	2007	w_5	Negative	–
Peloquin et al. [27]	2008	w_6	Inconclusive	x_4
Phase 2				
Burman et al. [28]	2006	x_1	Negative	w_5, w_6, x_3, x_4
Rustomjee et al. [29]	2007	x_2	Positive	x_3
Conde et al. [30]	2009	x_3	Positive	x_5
Dorman et al. [31]	2009	x_4	Negative	–
Wang et al. [32]	2009	x_5	Positive	–

While this method is less rigorous than a complete systematic review, it is sufficient for the illustrative purposes of this paper. Further development on harmonizing the AERO model with the methods of systematic review is underway.

Extraction

The ‘year’ in Table 1 refers to the year of publication. The ‘key,’ generated sequentially by phase, refers to the corresponding node in Figure 4. The outcome was extracted from the published manuscript, based on the direction of the observed effect, the authors’ recommendation for further research, and any expressed qualifications or reservations. The studies listed as ‘children’ are based on a transitive reduction of the total network of citations, reconstructed to represent the historical translation of evidence as accurately as possible.

Non-positive outcomes

We acknowledge that the inconclusive status of the two *in vivo* studies is debatable. Despite evidence of efficacy, we nevertheless classified Yoshimatsu et al. as inconclusive due to the authors’ concerns about toxicity at the recommended dosage [20]. Nuermberger et al. did not

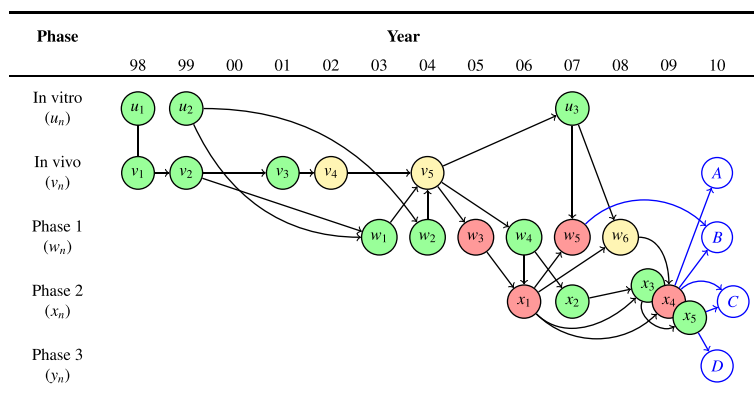


Figure 4 Complete AERO graph for moxifloxacin in an anti-tuberculosis regimen. The graph shows 19 completed experiments across 12 years along with four contemplated future studies. The overall trend of study results was positive until the transition into phase 2 (x_1), when significant discordance (for example, $w_4 \rightarrow x_1$ and $w_6 \rightarrow x_4$) and inconsistency (that is, negative results in x_1, x_4 vs. positive results in x_2, x_3, x_5) began to emerge. Researchers must now decide how to proceed in the face of an equivocal state of total evidence: Investigate mechanisms of discordance between animal models and human trials (A); investigate drug interactions (B); further investigate efficacy and evaluate predictivity of specific phase 2 trial designs (C); or proceed to a decisive phase 3 effectiveness trial (D).

show an improvement on their primary outcome (time-to-culture-negative), but did show a dramatic increase in early potency when moxifloxacin was substituted for isoniazid, one of the drugs in the standard regimen [21]. Gillespie et al. showed no difference in the early bactericidal activity between moxifloxacin and isoniazid [24]. Nijland et al. showed that moxifloxacin plasma concentrations are reduced when it is administered with isoniazid and rifampicin [26]. Peloquin et al. showed favorable population pharmacokinetics with moxifloxacin, but levofloxacin had the most favorable profile in their study [27]. Both Burman et al. and Dorman et al. showed no improvement in the time-to-culture-negative when moxifloxacin was substituted into the standard regimen for ethambutol and isoniazid, respectively [28,31].

Visual representation

Figure 4 is the complete AERO graph based on the information in Table 1 and the possible future studies, $A \dots D$. It was rendered using the *tikz* vector illustration package for LaTeX. All of the non-straight edges were shaped manually in order to aid visual comprehension. The edge between u_1 and v_1 is not an arrow because these two studies are published in the same report.

Analysis

The first thing to notice about Figure 4 is how much messier is the picture of an actual translational research program compared to the toy example we discussed above. One immediate advantage of the AERO graph is that it illuminates the point about translational research: It is not a linear process, but a complex network of overlapping investigation types. Phases can even be repeated,

as when evidence from a downstream phase is used to inform a subsequent upstream investigation (for example, $v_5 \rightarrow u_3$).

But however descriptively accurate is this complexity, there is a prescriptive question about how organized and systematic a research program in translational medicine ought to be. Is this overlapping web of studies an intrinsic part of translational medicine? Or can it be made more orderly, with one study and one phase proceeding after another without the need to backtrack? While we will not take up this intriguing question here, it is worth pointing out that the AERO model facilitates such an inquiry.

So what can we now say about moxifloxacin's translational trajectory? Across the entire trajectory, the trend is largely positive, with a 3:1 positive to negative ratio. There is also evidence of consistency at every phase and no negative results in the pre-clinical studies. Although to our knowledge there was no explicit evidence threshold established for this trajectory, we can nevertheless observe that for both the pre-clinical to clinical and phase 1 to phase 2 transition, the *de facto* threshold is three positive studies. Thus far, these would all seem to be encouraging properties. In fact, prior to the first negative phase 2 study in 2006, the evidence for moxifloxacin was overwhelmingly favorable.

After 2006, significant inconsistency and discordance emerges. Two of the five phase 2 studies, x_1 and x_4 , are negative and show discordance with the earlier phase 1 and pre-clinical outcomes. Indeed, there is a subset of the total trajectory, proceeding through w_3 to x_1 and then w_5 and x_4 , that is both consistently and concordantly negative. This underscores the fact that robustness is not only

a property of positive results. Results can also be robustly negative.

Yet, one negative sub-trajectory is not necessarily fatal for the whole translational trajectory. In this case, we can see an overall trend toward positive outcomes, but it is not obvious that this trend is sufficient to justify a phase 3 trial. Therefore, we now consider how the AERO graph can sharpen arguments for or against each of the four proposed future studies.

A: The argument in favor of *A*, another *in vivo* study, is supported by the striking discordance between the pre-clinical and clinical results along the following sub-trajectory:

$$\dots v_5 \rightarrow w_3 \rightarrow x_1 \rightarrow x_4 \Rightarrow A$$

Given that the *in vivo* models and phase 2 trials are both evaluating an efficacy endpoint, rather than the safety, pharmacokinetic, or early activity endpoints evaluated in phase 1, the discordance between *in vivo* and phase 2, combined with phase 2 inconsistency, casts a reasonable doubt on the predictive power of the animal models. For example, Dorman et al.'s (x_4) hypothesis, comparing the replacement of moxifloxacin for isoniazid in the standard regimen, was informed by Nuermberger et al.'s (v_5) finding in a murine model. The failure to translate this result supports revisiting the *in vivo* design and re-evaluating the import of *in vivo* experiments.

But we should note that pursuing another *in vivo* study does not necessarily invalidate the prior phase 1 or phase 2 results. Those studies can still be internally valid and their evidence can remain reliable and informative in future analyses. The rationale in favor of another animal study is that it may be more productive or efficient in the long term to try and explain the discordance between some of the extant animal and human results, rather than simply continuing to test moxifloxacin-containing regimens in humans.

B: The argument in favor of *B*, another phase 1 study, is supported by the inconsistencies at phases 1 and 2. In addition to the above sub-trajectory, which tracks the negative results at phase 2, this line of reasoning adds a second trajectory of emphasis:

$$\begin{aligned} \dots v_5 \rightarrow w_3 \rightarrow x_1 \rightarrow x_4 \Rightarrow B \\ \dots x_1 \rightarrow w_5 \Rightarrow B \end{aligned}$$

These sub-trajectories suggest a need to better understand the pharmacokinetic properties of moxifloxacin when used with the other drugs in the standard anti-tuberculosis regimen. Nijland et al.'s (w_5) finding that some combinations with moxifloxacin produce lower plasma concentrations relate directly to this point.

C: The argument in favor of *C*, another phase 2 trial, emphasizes differences in experimental design across the phase 2 studies, supported by the following sub-trajectories:

$$\begin{aligned} \dots v_5 \rightarrow w_4 \rightarrow x_1 \rightarrow x_3 \rightarrow x_5 \Rightarrow C \\ \dots x_1 \rightarrow x_4 \Rightarrow C \\ \dots x_2 \rightarrow x_3 \rightarrow x_5 \Rightarrow C \end{aligned}$$

Just as we discussed a robustly negative sub-trajectory, a researcher could emphasize the robustly positive sub-trajectory and claim that if any doubts about moxifloxacin's potential efficacy remain, that these ought to be eliminated by another, rigorous phase 2.

Importantly, the justification for *C* turns, in no small part, on assumptions about the purpose of the phase 2 trial and the threshold of evidence for proceeding to phase 3. If phase 2 trials are supposed to limit the number of candidate interventions for phase 3 strictly, then it may be reasonable to require at least a 2:1 positive-to-negative trial ratio, for example, before advancing. In which case, *C* seems a reasonable option. If, however, phase 2 trials are only supposed to rule out dangerous candidates, then three positive trials is arguably sufficient, and therefore, *C* would seem to be unnecessary.

D: This leads to the argument for *D*, a phase 3 trial, which could be justified on the grounds that moxifloxacin has been evaluated in three positive phase 2 trials, has not been ruled out as a novel candidate for inclusion in an effective anti-tuberculosis regimen, and is supported by the sub-trajectory:

$$\dots v_5 \rightarrow w_4 \rightarrow x_2 \rightarrow x_3 \rightarrow x_5 \Rightarrow D$$

As we have already mentioned, the overall picture of moxifloxacin's trajectory is largely positive – not perfect, but arguably robust. Therefore, it may be reasonable to proceed to phase 3, despite the inconsistency across phase 2. Indeed, we should acknowledge here the REMox-TB trial (NCT00864383), an in-progress phase 3 trial evaluating a moxifloxacin-containing regimen for the treatment of drug-susceptible tuberculosis. The results of this trial, although not yet available, can be understood, in part, as an empirical test of this line of reasoning. A positive and reproducible finding in the REMox-TB trial would be evidence that the AERO graph of Figure 4 reflects a promising trajectory. Inversely, a negative finding should cast doubt on the idea that such a trajectory is robust enough to justify a phase 3 trial.

Summary

Whichever direction is ultimately selected for future research, the AERO model can be dynamically updated in light of those new study results as they become available.

For example, suppose that moxifloxacin researchers were to pursue both options *A* and *C* – another animal study and another phase 2 trial, respectively. And assume further that both of these turn out to be positive. This could potentially show that the researchers now better understand the predictive relationship between the animal and human models and perhaps finally tip the balance of phase 2 evidence clearly in favor of a moxifloxacin regimen.

Yet, even before these studies are executed, researchers could use the AERO graph to strategize about future states of evidence. In contemplating option *B*, another phase 1 trial, one might reasonably question if that result, whatever it turns out to be, would be informative and useful enough to justify the opportunity cost of not pursuing *A*, *C* or *D*.

We should also acknowledge that these alternative research strategies are neither exhaustive nor mutually exclusive. A well-funded research program may be able to pursue all of these and other options simultaneously. The aim in this section is simply to show how the AERO graph can help to clarify and sharpen the rationale for the various possibilities and in so doing, aid decision-making about the directions for further research.

That said, the more fundamental questions elucidated by the moxifloxacin trajectory – calling into question the predictive value of the animal model, reconsidering the purpose for phase 2, and the appropriate thresholds of robustness at each phase transition – should not be overlooked, even if the decision is made to proceed to a phase 3 trial. Part of what Figure 4 illustrates is that there are many unanswered questions about moxifloxacin's trajectory and the underlying causal relationships, and until at least some of these are addressed, a phase 3 trial, whatever its outcome, will be less informative than it could be, and hence, reflect an inefficient research strategy over the long term.

Discussion

Stepping back from this specific case study, there are a number of possible elaborations and applications of the AERO modeling approach that are worth discussing. To begin with, the color-coding we presented here, which classifies studies as positive, negative, or inconclusive, has significant virtues of simplicity and ease of interpretation, but this is far from the only option. For example, a continuous shading scale could be used to represent the effect size or precision of each study; or a five-point ordinal scale could be used, corresponding to the region of posterior interval. The fundamental structure of the AERO model is the directed acyclic graph with study-type strata; the other graphical properties can be extensible in any way that supports decision-making.

The AERO graph can also be thought of as reflecting the maturity of causal knowledge within a given

research domain. A robust field of mostly green nodes would suggest that investigators have full command of the mechanisms in the causal system, whereas a field of predominantly red or yellow nodes would suggest utter lack of contact with the causal factors. A pattern of pre-clinical green that consistently turns to red in clinical translation should cast doubt on the validity of pre-clinical models. A thin thread of green would perhaps suggest just a lucky find, while an evenly balanced network of green and red nodes would suggest a new, emerging domain with which investigators may have only a limited understanding.

This relates to another extension of the approach – comparing and contrasting multiple trajectories. We could see such a comparative analysis being useful in a number of ways: For example, the AERO graph of a successfully translated agent could be used as the model for how new agents in a treatment domain ought to be vetted; or a population of AERO graphs could elucidate systematic differences in efficiency or risk and benefit across different research domains.

Looking ahead to envision the potential utility of the AERO model, particularly as a means to improve the efficiency and coordination of the drug development enterprise, we believe that publications and protocols could both benefit from including an AERO graph. Just like other kinds of visual tools, such as Thorpe et al.'s PRECIS graph [33] or Langan et al.'s graphical augmentation to the meta-analytic funnel plot [34], serve to aid judgments about the quality and direction of evidence, so too can the AERO model help to make invisible or inchoate properties of the research system explicit.

Indeed, AERO graphs could provide a convenient check for investigators, institutional review boards, journal editors, and physicians. The comprehensibility of the representation allows for anyone even modestly familiar with a particular domain to be able to compare and contrast alternative representations. As a consequence, a report that ignores a substantial portion of the available evidence can be more easily detected, since its representation will differ dramatically. This is not to contradict what we claimed earlier about disagreement, as different researchers may have equally legitimate interpretations of their field. However, the apparent contrast between interpretations of the evidence will demand an explanation for why the authors represented the field in one way rather than another.

Conclusion

The AERO model provides an immediate visual representation of the number of studies done at any mode, depicting both the direction of evidence and the relationship of each study to the larger translational trajectory. In so doing, it helps to address the widespread concerns about efficiency, coordination, and organization in

translational medicine. To be sure, the visual representation does not capture all of the available information about a research program. For example, the details of differing experimental designs or conduct may be relevant to understanding failures of robustness. Just as we showed in the previous section, the argument for one course of action over another relies upon knowledge of these additional details about each experiment. Nevertheless, the AERO model provides a systematic representation that is capable of sharpening these judgments and revealing some of the existing patterns across a translational trajectory.

We recognize that the approach we have presented here is but a preliminary sketch. While we have applied the AERO model to a single case study, much more work is needed to develop the approach to its full potential. But particularly given the stakes, thinking about ways to better analyze and judge the structure of clinical research programs as a whole seems a vital line of inquiry. The AERO model is one piece of this inquiry.

Abbreviations

AERO: Accumulating Evidence and Research Organization; DAG: Directed acyclic graph.

Competing interest

The authors declare that they have no competing interests.

Authors' contributions

All authors contributed to the conception and design of the manuscript. SPH wrote all drafts. CMH and CW commented and suggested changes on all drafts. All authors read and approved the final manuscript.

Acknowledgements

We thank Jonathan Kimmelman, William R MacKenzie, and Andrew Vernon for helpful comments and suggestions on earlier drafts. The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

Author details

¹Studies for Translation, Ethics, and Medicine (STREAM) Group, Biomedical Ethics Unit, McGill University, Montreal, QC H3A 1X1, Canada. ²Centers for Disease Control and Prevention, Division of Tuberculosis Elimination, 1600 Clifton Rd, NE, MS E10, Atlanta, GA, 30333, USA. ³Rotman Institute of Philosophy, Department of Philosophy, Western University, 1151 Richmond Street, London, N6A 5B8, ON, Canada.

Received: 5 January 2013 Accepted: 8 May 2013

Published: 30 May 2013

References

1. Tall AR, Yvan-Charvet L, Wang N: **The Failure of Torcetrapib: Was it the molecule or the mechanism?** *Arteriosclerosis, Thrombosis, Vasc Biol* 2007, **27**:257–260.
2. Joy TR, Hegele RA: **The failure of torcetrapib: what have we learned?** *British J Pharmacol* 2008, **154**:1379–1381.
3. Zhao L, Jin W, Rader D, Packard C, Feuerstein G: **A Translational Medicine perspective of the development of torcetrapib: Does the failure of torcetrapib development cast a shadow on future development of lipid modifying agents, HDL elevation strategies or CETP as a viable molecular target for atherosclerosis? A case study of the use of biomarkers and Translational Medicine in atherosclerosis drug discovery and development.** *Biochem Pharmacol* 2009, **78**:315–325.
4. Arrowsmith J: **Phase II failures: 2008–2010.** *Nat Rev Drug Discov* 2011, **10**:1.
5. Morgan P, Graaf PHVD, Arrowsmith J, Feltner DE, Drummond KS, Wegner CD, Street SDA: **Can the flow of medicines be improved? Fundamental pharmacokinetic and pharmacological principles toward improving Phase II survival.** *Drug Discov Today* 2012, **17**:419–424.
6. Kola I, Landis J: **Can the pharmaceutical industry reduce attrition rates?** *Nat Rev Drug Discov* 2004, **3**:711–715.
7. Arrowsmith J: **Phase III and submission failures: 2007–2010.** *Nat Rev Drug Discov* 2011, **10**:1.
8. Boucher HW, Talbot GH, Bradley JS Jr, JEE Gilbert, D, Rice LB, Scheld M, Spellberg B, Bartlett J: **Bad Bugs, No Drugs: No ESKAPE! An Update from the Infectious Diseases Society of America.** *Clin Infect Dis* 2009, **48**:1–12.
9. Seymour L, Ivy SP, Sargent D, Spriggs D, Baker L, Rubinstein L, Ratain MJ, Blanc ML, Stewart D, Crowley J, Groshen S, Humphrey JS, West P, Berry D: **The design of phase II clinical trials testing cancer therapeutics: consensus recommendations from the clinical trial design task force of the National Cancer Institute Investigational Drug Steering Committee.** *Clin Cancer Res* 2010, **16**:1764–1769.
10. StopTBP Partnership: **The global plan to stop TB.** *World Health Organization* 2010. [http://www.stoptb.org/assets/documents/globalplan/TB_GlobalPlanToStopTB2011-2015.pdf].
11. Hey SP: *Meta-heuristic strategies in scientific judgment.* PhD thesis. University of Western Ontario; 2011.
12. Kimmelman J: **A theoretical framework for early human studies: Uncertainty, intervention ensembles, and boundaries.** *Trials* 2012, **13**:173.
13. Prinz F, Schlange T, Asadullah K: **Believe it or not: how much can we rely on published data on potential drug targets?** *Nat Rev Drug Discov* 2011, **10**:712–713.
14. Begley CG, Ellis LM: **Raise standards for preclinical cancer research.** *Nature* 2012, **483**:531–533.
15. Ji B, Lounis N, Maslo C, Truffot-Pernot C, Bonnafoux P, Grosset J: **In Vitro and In Vivo activities of Moxifloxacin and Clinafloxacin against mycobacterium tuberculosis.** *Antimicrob Agents Chemother* 1998, **42**:2066–2069.
16. Gillespie S, Billington O: **Activity of Moxifloxacin against Mycobacteria.** *J Antimicrob Chemother* 1999, **44**:393–395.
17. Shandil RK, Jayaram R, Kaur P, Gaonkar S, Suresh BL, Mahesh BN, Jayashree R, Nandi V, Bharath S, Balasubramanian V: **Moxifloxacin, Ofloxacin, Sparfloxacin, and Ciprofloxacin against Mycobacterium tuberculosis: evaluation of In Vitro and Pharmacodynamic indices that best predict In Vivo efficacy.** *Antimicrob Agents Chemother* 2007, **51**:576–582.
18. Miyazaki E, Miyazaki M, Chen J, Chaisson R, Bishai W: **Moxifloxacin (BAY12-8039), a new 8-methoxyquinolone, is active in a mouse model of tuberculosis.** *Antimicrob Agents Chemother* 1999, **43**:85–89.
19. Lounis N, Bentoucha A, Truffot-Pernot C, Ji B, O'Brien R, Vernon A, Roscigno G, Grosset J: **Effectiveness of once-weekly Rifapentine and Moxifloxacin regimens against Mycobacterium tuberculosis in mice.** *Antimicrob Agents Chemother* 2001, **45**:3482–3486.
20. Yoshimatsu T, Nuernberger E, Tyagi S, Chaisson R, Bishai W, Grosset J: **Bactericidal activity of increasing daily and weekly doses of Moxifloxacin in Murine tuberculosis.** *Antimicrob Agents Chemother* 2002, **46**:1875–1879.
21. Nuernberger E, Yoshimatsu T, Tyagi S, O'Brien R, Vernon A, Chaisson R, Bishai W, Grosset J: **Moxifloxacin-containing regimen greatly reduces time to culture conversion in Murine tuberculosis.** *Am J Respir Crit Care Med* 2004, **169**:421–426.
22. Gosling R, Uiso L, Sam N, Bongard E, Kanduma E, Nyindo M, Morris R, Gillespie S: **The bactericidal activity of Moxifloxacin in patients with pulmonary tuberculosis.** *Am J Respir Crit Care Med* 2003, **168**:1342–1345.
23. Pletz M, Roux AD, Roth A, Neumann K, Mauch H, Lode H: **Early bactericidal activity of Moxifloxacin in treatment of pulmonary tuberculosis: a prospective, randomized study.** *Antimicrob Agents Chemother* 2004, **48**:780–782.
24. Gillespie S, Gosling R, Uiso L, Sam N, Kanduma E, McHugh T: **Early bactericidal activity of a Moxifloxacin and Isoniazid combination in smear-positive pulmonary tuberculosis.** *J Antimicrob Chemother* 2005, **56**:1169–1171.

25. Johnson J, Hadad D, Boom W, Daley C, Peloquin C, Eisenach K, Jankus D, Debanne S, Charlebois E: **Maciel E, et al.: Early and extended early bactericidal activity of Levofloxacin, Gatifloxacin and Moxifloxacin in pulmonary tuberculosis.** *Int J Tuberc Lung Dis* 2006, **10**:605–612.
26. Nijland HMJ, Ruslami R, Suroto AJ, Burger DM, Alisjahbana B, van Crevel, R, Aarnoutse RE: **Rifampicin reduces plasma concentrations of Moxifloxacin in patients with tuberculosis.** *Clin Infect Dis* 2007, **45**:1001–1007.
27. Peloquin CA, Hadad DJ, Molino LPD, Palaci M, Boom WH, Dietze R, Johnson JL: **Population Pharmacokinetics of Levofloxacin, Gatifloxacin, and Moxifloxacin in adults with pulmonary tuberculosis.** *Antimicrob Agents Chemother* 2008, **52**:852–857.
28. Burman W, Goldberg S, Muzanye JJG, Engle M, Mosher A, Choudhri S, Daley C, Munsiff S: **Zhao Z, et al.: Moxifloxacin versus Ethambutol in the first 2 months of treatment for pulmonary tuberculosis.** *Am J Respir Crit Care Med* 2006, **174**:331–338.
29. Rustomjee R, Lienhardt C, Kanyok T, for TB (OFLOTUB) study team G: **A phase II study of the sterilising activities of Ofloxacin, Gatifloxacin and Moxifloxacin in pulmonary tuberculosis.** *Int J Tuberc Lung Dis* 2008, **12**:128–138.
30. Conde M, Efron A, Loreda C, Souza GD, Gracxa N, Cezar M, Ram M, Chaudhary M, Bishai W, Kritski A, Chaisson R: **Moxifloxacin versus Ethambutol in the initial treatment of tuberculosis: a double-blind, randomised, controlled phase II trial.** *Lancet* 2009, **53**:1314–1319.
31. Dorman S, Johnson J: **Goldberg S, et al.: Substitution of Moxifloxacin for Isoniazid during intensive phase treatment of pulmonary tuberculosis.** *Am J Respir Crit Care Med* 2009, **180**:273–280.
32. Wang JY, Wang JT, Tsai T, Hsu C, Yu C, Hsueh P, Lee L, Yang P: **Adding Moxifloxacin is associated with a shorter time to culture conversion in pulmonary tuberculosis.** *Int J Tuberc Lung Dis* 2009, **14**:65–71.
33. Thorpe KE, Zwarenstein M, Oxman AD, Treweek S, Furberg CD, Altman DG, Tunis S, Bergel E, Harvey I, Magid DJ, Chalkidou K: **A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers.** *J Clin Epidemiol* 2009, **62**:464–475.
34. Langan D, Higgins JPT, Gregory W, Sutton AJ: **Graphical augmentations to the funnel plot assess the impact of additional evidence on a meta-analysis.** *J Clin Epidemiol* 2012, **65**:511–519.

doi:10.1186/1745-6215-14-159

Cite this article as: Hey et al.: Accumulating Evidence and Research Organization (AERO) model: a new tool for representing, analyzing, and planning a translational research program. *Trials* 2013 **14**:159.

Submit your next manuscript to BioMed Central
and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

