PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Mapping the preclinical to clinical evidence and development trajectory of the oncolytic virus talimogene laherparepvec (T-VEC): a systematic review
AUTHORS	Lalu, Manoj; Leung, Garvin; Dong, Yuan Yi; Montroy, Joshua; Butler, Claire; Auer, Rebecca; Fergusson, Dean

VERSION 1 – REVIEW

REVIEWER	Sumimasa Nagai Deputy Director, Translational Research Center, The University of Tokyo Hospital, Tokyo, Japan Consulting or an advisory role: Takara Bio Inc
REVIEW RETURNED	07-Feb-2019

GENERAL COMMENTS	Objective of this study is clinically relevant. However, authors should address the following issues: 1. In Table 1 and 2, authors should describe the definition of "Risk of biases" in detail. Authors explained "Risk of biases" too briefly without references in the METHODS section. 2. In Table 3,4, and 5, authors should describe the definition of "Yes", "No", "High risk", "Low risk", and "Unclear" in detail. 3. Authors mentioned the FDA and EMA reports too briefly without references in the DISCUSSION section. Authors should thoroughly described how the FDA and EMA evaluated the Amgen's filing dossier.

REVIEWER	Frances Collichio The University of North Carolina, Chapel Hill Dr Collichio does research for the University of North Carolina. She derives salary support for her research from Amgen, Novartis, and Merck pharmaceutical companies.
REVIEW RETURNED	05-Apr-2019

GENERAL COMMENTS	I appreciate the opportunity to review the manuscript BMJ open - 2019–029475 “Diminishing returns along the road to translation: a systematic review T-VEC’s preclinical to clinical development trajectory”. This is a well written paper that represents a great deal of work. There was very interesting information on construct validity and bias. I am concerned about your conclusion that the drug development of TVEC was not a successful model. You tried to show that response rate in preclinical models was higher than in the

	early clinical trials and from there, the phase III study. Oncology drug development in phase I studies is not meant to look at efficacy. These studies are meant to look at safety. Furthermore, the phase III study looked at durable response rate as its primary endpoint, not response rate. I think that the paper should state that there was successful development of TVEC despite the difficulties in apply preclinical models. The purpose of the paper was to use the drug development of talimogene laherparepvec (TVEC) as a successful model of preclinical translation to clinical approval of a product. To achieve this goal, the authors performed extensive literature review including all of the clinical and pre-clinical studies of TVEC for treatment of any malignancy. *Ex vivo, in vitro* and duplicate studies were excluded. The data were extracted independently using a standardized dating extraction form. The clinical studies were collected by patient characteristics. Risk of bias was assessed. Construct validity, the concept of how much a preclinical experience corresponds to a clinical entity was assessed. A total of 8,852 references were identified and 7,890 references were excluded. Clinical studies were published between 2006 and 2016. Sample sizes ranged from 17 to 295. Pre-clinical studies reported a complete response rate up to 100% for injected tumors and 80% for contralateral tumors. A phase I study reported a complete response rate of 0%. Phase I/II studies in melanoma showed a complete response of 20 to 22% (1,2). The phase III OPTiM trial showed a complete response rate of 10.8% (3). The authors attempted to assess safety however; they were unable to obtain patient level data from the studies. They found that all pre-clinical studies had high risk of construct validity. The preclinical models had a high risk of bias. For clinical trials, the early phase trials had high or unclear risk of bias. The phase III OPTiM study had the lowest risk of bias. The authors concluded that they were unable to answer their main question, which was to show a successful road map of pre-clinical to clinical development using TVEC as a model. Showing a

	road map of drug development is an important topic and I appreciate the authors tackling the project but I think they have made a major flaw. I am not an expert in the area of drug development but I do have a lot of expertise in oncology, melanoma, and the product that was assessed in this report. The major flaw in the study was their conclusion that translation from pre-clinical to clinical showed a detriment in the advocacy of the product. That is not true. The major endpoint of the OPTiM trial was durable response for greater than or equal to six months. Subjects with such a response had their data independently reviewed (3). OPTiM met this endpoint and the treatment was approved by the FDA. Complete response was not the end point. Complete response, while laudable, is not necessarily clinically meaningful in patients with advanced disease. Complete response may have no correlation with durable response or survival in cancer trials. The authors have a flawed conclusion as they are looking at complete response in the preclinical trials compared to complete response in the clinical studies. The authors also state that as trials moved from preclinical to phase I there was a disconnect. Phase I studies are done on patients who are generally unable to obtain standard of care treatments and the response rates are not the goal of a phase 1 study. The goal of these studies is safety. Response is not a goal of these studies. It is erroneous to expect to see a good result from them. Another flaw in this report is the paragraph under safety of treatment. I don't think it would be possible to get patient level data for a report like this but on this point, I am not certain as I do not do studies like this one. Side effects are included in the literature, including in the OPTiM study (3). I have several questions about the study. Under results, they stated 8,852 records (I think they meant references) were screened and 7890 references were excluded. I couldn't find in the text or in the figure why the references were excluded. I suppose this was because of duplicate studies, or studies that reported *ex vivo* or *in vitro* experiments but it would have been helpful to see why they

	were excluded. The section on validity assessment is outside of my area of expertise so I would appreciate it if the editors have another reviewer address this section. The statement that the phase III OPTiM trial had the lowest efficacy of published melanoma clinical trials is not true. Again the OPTiM study looked at durable response for greater than or equal to six months as the primary endpoint, not response rate. The section on statistical analysis is short. I don't not know if this is an adequate assessment for this type of review and I would appreciate it the other reviewer could lend expertise here. I am confused by the legend to Figure 2. The top part of figure 2 shows the pre-clinical studies and the bottom part shows the clinical studies. The non-melanoma/mixed studies are in a blue bar and the melanoma studies are in an orange bar. There are no blue bars in the "clinical "section in the lower part of this figure. Table 1A needs a legend. Table 1 does not need to be labeled 1 "A". Table 2 does not need the "B". In Supplemental Table 3, the Andtbacka, OPTiM study, dose 2 is 3 weeks after the first dose. Cycle 2 starts 5 weeks after the first dose and all subsequent cycle are every 4 weeks. I did not crosscheck the other studies on timing of doses. What is the purpose of the PRISMA checklists? Is the first one older? I found the Table 4 and 5 interesting and they added to the information presented by the authors. Figure 1, which shows the study selection flow, includes the excluded records. It is in this section that the authors could state why the records were excluded. There is a typo in the abstract. It states that "the study aimed to conducted" and it should be "the study aimed to conduct" In conclusion this was a well written study that was a lot of work for

	the authors. The goal of the study was to show the development of a product from pre-clinical to clinical work. The authors concluded that they were unable to paint a clear picture of how the evidence was used in preclinical studies to clinical findings. But, there is a major flaw in their report. They used response rate from preclinical to clinical to justify drug development. Phase I studies are not designed to look at response rate. The OPTiM phase III study that was used to approve TVEC had durable response, not response rate as its primary endpoint (3). If the authors were to re-write the paper showing that there was successful development of TVEC despite the issues that they uncovered in the pre-clinical models, I think that the report would be valid. They also need to include the purpose and methods of phase I, II, and III trials in their paper and show that this was a valid construct for TVEC. They could conclude that they have raised and important issue that research should continue to seek models that have a smooth transition from preclinical to clinical development but this work is likely to be very difficult given the vast difference of medications on humans versus animal models. 1. Hu JC, Coffin RS, Davis CJ, Graham NJ, Groves N, Guest PJ, et al. A phase I study of OncoVEXGMCSF, a second-generation oncolytic herpes simplex virus expressing granulocyte macrophage colony stimulating factor. Clinical cancer research : an official journal of the American Association for Cancer Research. 2006;12(22):6737-47. 2. Senzer NN, Kaufman HL, Amatruda T, Nemunaitis M, Reid T, Daniels G, et al. Phase II clinical trial of a granulocyte-macrophage colony-stimulating factor-encoding, second-generation oncolytic herpes virus in patients with unresectable metastatic melanoma. J Clin Oncol. 2009;27(34):5763-71. 3. Andtbacka RHI, Kaufman HL, Collichio F, Amatruda T, Senzer N, Chesney J, et al. Talimogene Laherparepvec Improves Durable Response Rate in Patients With Advanced Melanoma. Journal of Clinical Oncology. 2015;33(25):2780-8.

REVIEWER	Alex Sverdlov Novartis, USA
REVIEW RETURNED	17-Jun-2019

GENERAL COMMENTS	The manuscript presents a systematic review of the literature describing the development program (pre clinical and clinical

	studies) of an FDA-approved oncolytic virus therapy to treat advanced melanoma. The authors performed a systematic search of the published studies and found lack of a clear translational research roadmap in that early pre-clinical positive findings are frequently not reproducible in larger clinical trials. I think this is an important case study demonstrating the need for better translational research and communication of the results. The methodology is scientifically sound and the conclusions are supported by the data. I found the manuscript well written and up to the point. A few minor comments for the authors' consideration: • Safety analysis is missing – you may want to elaborate on why this was the case and how to improve reporting of safety in translational research. • Figure 2 – please explain which method was used to produce 95% confidence intervals for the response rates. • In the authors' opinion, what should a successful translational roadmap look like? What key data should be documented and reported and at what points?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Sumimasa Nagai
Institution and Country: Deputy Director, Translational Research Center, The University of Tokyo Hospital, Tokyo, Japan Please state any competing interests or state 'None declared': Consulting or an advisory role: Takara Bio Inc

Objective of this study is clinically relevant. However, authors should address the following issues:

1. In Table 1 and 2, authors should describe the definition of "Risk of biases" in detail. Authors explained "Risk of biases" too briefly without references in the METHODS section.

We have now described the specific domains included in the risk of bias assessment as a footnote to Table 1.

2. In Table 3,4, and 5, authors should describe the definition of "Yes", "No", "High risk", "Low risk", and "Unclear" in detail.

This is the standard method of assessing risk of bias of clinical and preclinical studies in systematic reviews. It is difficult to write a simple definition for these as a whole chapter of the Cochrane Handbook is dedicated to defining risk of bias (Chapter 8) and the preclinical adaptation of this is elaborated on in a separate paper (SYRCLE risk of bias tool). We have referred to the Cochrane Handbook and the SYRCLE risk of bias tool for readers interested in learning more about systematic review methodology.

3. Authors mentioned the FDA and EMA reports too briefly without references in the DISCUSSION section. Authors should thoroughly described how the FDA and EMA evaluated the Amgen's filing dossier.

We have now provided references in the discussion for the two briefing documents. In relation to further discussion revolving around the handling of AMGEN's filing dossier, there is no preclinical evidence reported in the dossier regarding melanoma models, which was the focus of the discussion point presented. We feel describing the entire regulatory decision processes thoroughly within the manuscript to be unnecessary. If reviewers deem that

addressing this point in more detail is required, we would kindly request further clarity on exactly what is being asked.

Reviewer: 2
Reviewer Name: Frances Collichio
Institution and Country: The University of North Carolina, Chapel Hill Please state any competing interests or state 'None declared': Dr Collichio does research for the University of North Carolina. She derives salary support for her research from Amgen, Novartis, and Merck pharmaceutical companies.

1. I appreciate the opportunity to review the manuscript BMJ open -2019–029475 "Diminishing returns along the road to translation: a systematic review T-VEC's preclinical to clinical development trajectory". This is a well written paper that represents a great deal of work. There was very interesting information on construct validity and bias. I am concerned about your conclusion that the drug development of TVEC was not a successful model. You tried to show that response rate in preclinical models was higher than in the early clinical trials and from there, the phase III study. Oncology drug development in phase I studies is not meant to look at efficacy. These studies are meant to look at safety. Furthermore, the phase III study looked at durable response rate as its primary endpoint, not response rate. I think that the paper should state that there was successful development of TVEC despite the difficulties in apply preclinical models.

We thank the reviewer for their positive comments. Based on your comments as well as the Editor's suggestion, we have now detailed the phases of drug development in the Introduction.

"Although the process of clinical translation is complicated, the transition from bench-to-bedside often starts with preclinical research. These investigations (usually on animals or cells), are aimed at studying efficacy, pharmacokinetics and dynamics, as well as detailing safety.(4) Next, a drug is tested in a phase I clinical trial, which usually contains a small number of participants and is aimed at studying the safety of the drug. If a drug is safe, it may proceed to phase II which are larger than phase I studies and are designed to test safety, pharmacokinetics, pharmacodynamics, and optimal dosing regimens. They may also offer preliminary evidence of drug efficacy. Finally, a methodologically rigorous phase III study is performed. These studies are designed and powered to test efficacy in the patient population of interest (usually against a comparator such as placebo), as well as identify rarer adverse events which may have gone unnoticed in a smaller phase I or II study.(**5)"**

In addition, we have added the following to the discussion to clearly state the T-VEC.

"Although TVEC was successful in terms of gaining regulatory approval, its translational path is complicated, and the pieces of the evidence puzzle do not easily fit together. While we appreciate that translation is not a predictable linear process, it is difficult to learn from the example of TVEC given the available and reported pre-clinical and clinical evidence. We feel that the incongruence in terms of clinical conditions, and efficacy estimates speak to the difficulties of assessing and determining generalizable strategies for successful translation."

Reviewer: 3
Reviewer Name: Alex Sverdlov
Institution and Country: Novartis, USA
Please state any competing interests or state 'None declared': None declared

The manuscript presents a systematic review of the literature describing the development program (preclinical and clinical studies) of an FDA-approved oncolytic virus therapy to treat advanced melanoma. The authors performed a systematic search of the published studies and found lack of a clear translational research roadmap in that early pre-clinical positive findings are frequently not reproducible in larger clinical trials. I think this is an important case study demonstrating the need for better translational research and communication of the results. The methodology is

scientifically sound and the conclusions are supported by the data. I found the manuscript well written and up to the point. A few minor comments for the authors' consideration:

1. Safety analysis is missing – you may want to elaborate on why this was the case and how to improve reporting of safety in translational research.

We agree. Unfortunately, safety was very poorly and underreported between trials. Different categorization of AEs was used between trials, which rendered pooling of results impossible. Without the patient level data to ensure independence of events, a safety analysis may have provided false or misleading conclusions. For this reason, we had specifically asked Amgen for patient level data in order to conduct a safety analysis. Unfortunately, despite two meetings with their group responsible for T-VEC, they were unwilling to share the data.

As such, we have added the following additions have been made to the results section, and the discussion section.

"Studies did not specify what percent of adverse events were repeated adverse events from the same patient(s), used different criteria for recording and reporting adverse events, categorized them differently, etc. Therefore, we were unable to pool adverse events or interpret findings reliably."

"The reporting of harms in clinical trials remains an issue in the scientific community, and represents a roadblock to translational success. Some basic steps required to improve the reporting of safety in translational research include the development of standardized scales and instruments, instituting active rather than passive surveillance for toxicity, including detailed information on participant withdrawals due to toxicity, reporting the timing, frequency, and duration of clinically relevant events, and the the publication of raw data."

2. Figure 2 – please explain which method was used to produce 95% confidence intervals for the response rates.

A short statement has been added to the methods section.

3. In the authors' opinion, what should a successful translational roadmap look like? What key data should be documented and reported and at what points?

We have now added the following to address this important point:

"While we recognize that translation is not a linear process, we should observe consistent and coherent patterns. Moving forward, we suggest that preclinical and clinical studies for emerging therapies should be fully reported and attention should be given to validities in order to develop more precise estimates of effect early in development. Investigators should carefully match their preclinical model to the intended clinical population; when possible, both disease states and outcomes measured should have high construct validity. Following successful exploratory preclinical studies, investigators should consider preclinical systematic reviews and designing methodologically rigorous confirmatory and/or multicenter preclinical studies. These steps may allow preclinical testing to more accurately forecast downstream clinical results in human patients. Within the trajectory of clinical development (i.e. once clinical trials have been initiated), careful consideration of methods to reduce bias should also be considered (although, this may not be possible for the earliest phase trials). We believe these steps will provide unbiased and valuable information that will ultimately provide patients with cancer therapies that match their preclinical and early clinical promise."

VERSION 2 – REVIEW

REVIEWER	Sumimasa Nagai Translational Research Center, The University of Tokyo Hospital, Tokyo, Japan Consulting or an advisory role: Takara Bio Inc
REVIEW RETURNED	24-Sep-2019

GENERAL COMMENTS	Authors appropriately responded to the reviewers' comments.

REVIEWER	Frances Collichio The University of North Carolina, Chapel Hill Dr Collichio receives salary support from Novartis, Amgen, GSK for clinical trial research. She is also a clinician who treats patients with Talimogene Laherparepvec.
REVIEW RETURNED	09-Oct-2019

GENERAL COMMENTS	Thank you for the opportunity to look at the revised version of your manuscript, now called “Mapping the pre-clinical to clinical evidence and development trajectory of oncolytic virus tell imaging Talimogene Laherparepvec (TVEC): a systemic review. This revision is much improved. My comments will address the clinical aspects of this manuscript since that is my expertise. The overall emphasis of the manuscript, explaining that you are using this as a model of drug development was clear. I think it is important to state in the beginning, on page 4, that you chose TVEC as your model. Please explain why other models were not chosen. You have a sentence that says, “examine a few agents”, but you do not state why other agents were not included. The background section includes a new paragraph on the process of clinical translation explaining the differences between phase 1, phase 2, and phase 3 trials. I found this very helpful. Your primary and point for advocacy is complete response. You do state that secondary endpoints were survival, response rate, time to treatment failure, and disease stability. The FDA approved TVEC based on the registry trial, OPTIM. The endpoint was six-month durability of response. You should state this. You could also add References from 2018 in 2019 in which real world application of this medication continues to show clinical benefit. There is one paper that shows a complete local response of 39% (Louie et al. J Am Coll Surg. 2019 Apr;228(4):644-649. doi: 10.1016/j.jamcollsurg.2018.12.027. Epub 2019 Jan 25.) You could add that TVEC is the recommended treatment by the National Comprehensive Cancer Center for patients with in transit melanoma. You explain how studies were chosen and you have a nice figure to show this. I was surprised by how many trials were excluded. Since this is an area that I’m not an expert in, please add another sentence or two about why so many studies were excluded in your final analysis.

	On page 11 line 48, you say the TVEC was administered “adjuvant” to chemotherapy. Can you clarify this? Was TVEC given after the chemotherapy? In oncology, the word adjuvant more commonly means, “given after surgery”. You also use the word “chemotherapy” but in several trials the TVEC is combined with immunotherapy such pembrolizumab and ipilimumab. You could get around this difficulty by using “systemic therapy” rather than chemotherapy. For the supplemental table 4, please add that the squamous cell carcinoma (Harrington reference) is from “Squamous Cell Carcinoma of the Head and Neck”. SCC of the head and neck is very different from SCC of the skin so adding these words will clarify this point. In your discussion, please restate that you would hope to produce a roadmap of translation from pre-clinical to clinical and you used TVEC as an example. Please restate why other biologic agents were not chosen.

REVIEWER	Alex Sverdlov Novartis, USA
REVIEW RETURNED	06-Oct-2019

GENERAL COMMENTS	None.

VERSION 2 – AUTHOR RESPONSE

1. The overall emphasis of the manuscript, explaining that you are using this as a model of drug development was clear. I think it is important to state in the beginning, on page 4, that you chose TVEC as your model. Please explain why other models were not chosen. You have a sentence that says, “examine a few agents”, but you do not state why other agents were not included.

   Many resources have been dedicated to developing oncolytic virus (OV) therapy. However, only one agent has been FDA approved. We therefore felt as though this was a perfect avenue to explore. We have recently published a study summarizing the large number of OVs that have been tested preclinically (Fergusson et al, Mol Ther Oncolytics, 2019, PMID: 31276026). A statement has now been added to the introduction “TVEC was chosen as a model due to the fact that it is the only approved oncolytic virus therapy to date, despite the multitude of agents under investigation”.

2. Your primary and point for advocacy is complete response. You do state that secondary endpoints were survival, response rate, time to treatment failure, and disease stability. The FDA approved TVEC based on the registry trial, OPTIM. The endpoint was six-month durability of response. You should state this.

   A short statement has been added to the methods section recognizing this limitation “We acknowledge the limitation of this approach, given the FDA approved TVEC based on the OPTIM trial,(13) the primary endpoint of which was durable response rate.” We recognize that this is an important outcome. However, not many of our included studies reported on this outcome. Therefore, in order to be able aggregate data across studies we used the outcome of complete response, which was more consistently reported across included studies.

3. You could also add References from 2018 in 2019 in which real world application of this medication continues to show clinical benefit. There is one paper that shows a complete local response of 39% (Louie et al. J Am Coll Surg. 2019 Apr;228(4):644-649. doi: 10.1016/j.jamcollsurg.2018.12.027. Epub 2019 Jan 25.) You could add that TVEC is the recommended treatment by the National Comprehensive Cancer Center for patients with in transit melanoma.

   This reference has been inserted into the discussion to reinforce an existing point. "Nonetheless, T-VEC has shown some efficacy in treating refractory melanoma(25) and numerous.." The information regarding the NCCC has also been added to the discussion as well. "It is also the recommended treatment by the National Comprehensive Cancer Center for patients with in-transit melanoma."

4. You explain how studies were chosen and you have a nice figure to show this. I was surprised by how many trials were excluded. Since this is an area that I'm not an expert in, please add another sentence or two about why so many studies were excluded in your final analysis.

   An extra sentence has been added to the results section to provide further clarity. "…another 938 articles were excluded for reasons such as wrong study design (i.e. review article), or wrong study intervention (i.e. a different cancer therapeutic)." These systematic searches are designed to be sensitive but non-specific, thus the number of excluded articles is very typical of preclinical/clinical systematic reviews.

5. On page 11 line 48, you say the TVEC was administered "adjuvant" to chemotherapy. Can you clarify this? Was TVEC given after the chemotherapy? In oncology, the word adjuvant more commonly means, "given after surgery". You also use the word "chemotherapy" but in several trials the TVEC is combined with immunotherapy such pembrolizumab and ipilimumab. You could get around this difficulty by using "systemic therapy" rather than chemotherapy.

   Wording suggestions have been applied to the manuscript.

6. For the supplemental table 4, please add that the squamous cell carcinoma (Harrington reference) is from "Squamous Cell Carcinoma of the Head and Neck". SCC of the head and neck is very different from SCC of the skin so adding these words will clarify this point.

   This has now been changed within the table.

7. In your discussion, please restate that you would hope to produce a roadmap of translation from pre-clinical to clinical and you used TVEC as an example. Please restate why other biologic agents were not chosen.

   This has been reinforced within the discussion. "We hoped to synthesize the evidence to produce a clear road map of T-VEC's translation in the published literature to follow the journey of a successful biotherapeutic, to be used as a blueprint for similar efforts in the future."