Molecular Therapy
Oncolytics
Original Article

AMERICAN SOCIETY of GENE & CELL THERAPY

Assessing the Completeness of Reporting in Preclinical Oncolytic Virus Therapy Studies

Dean A. Fergusson,[1,2] Neil L. Wesch,[1] Garvin J. Leung,[1,2] Jenna L. MacNeil,[1] Isidora Conic,[1] Justin Presseau,[1,3] Kelly D. Cobey,[3,4,5] Jean-Simon Diallo,[6,7] Rebecca Auer,[2,6,7] Jonathan Kimmelman,[8] Natasha Kekre,[2] Nader El-Sayes,[2,6] Ramya Krishnan,[2,6] Brian A. Keller,[2,6] Carolina Ilkow,[6,7] and Manoj M. Lalu[1,9,10,11]

[1]Clinical Epidemiology Program, BLUEPRINT Translational Research Group, Ottawa Hospital Research Institute, Ottawa, ON K1H 8L6, Canada; [2]Faculty of Medicine, University of Ottawa, Ottawa, ON K1H 8M5, Canada; [3]School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON K1H 8M5, Canada; [4]Centre for Journalology, Ottawa Hospital Research Institute, Ottawa, ON K1H 8L6, Canada; [5]Department of Psychology, University of Stirling, Stirling FK9 4LA, UK; [6]Cancer Therapeutics Program, Ottawa Hospital Research Institute, Ottawa, ON K1H 8L6, Canada; [7]Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, ON K1H 8M5, Canada; [8]Biomedical Ethics Unit, McGill University, Montreal, QC H3A 0G4, Canada; [9]Regenerative Medicine Program, Ottawa Hospital Research Institute, Ottawa, ON K1H 8L6, Canada; [10]Department of Anesthesiology and Pain Medicine, University of Ottawa, The Ottawa Hospital, Ottawa, ON K1H 8L6, Canada; [11]Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, ON K1H 8M5, Canada

Irreproducibility of preclinical findings could be a significant barrier to the "bench-to-bedside" development of oncolytic viruses (OVs). A contributing factor is the incomplete and non-transparent reporting of study methodology and design. Using the NIH Principles and Guidelines for Reporting Preclinical Research, a core set of seven recommendations, we evaluated the completeness of reporting of preclinical OV studies. We also developed an evidence map identifying the current trends in OV research. A systematic search of MEDLINE and Embase identified all relevant articles published over an 18 month period. We screened 1,554 articles, and 236 met our *a priori*-defined inclusion criteria. Adenovirus (43%) was the most commonly used viral platform. Frequently investigated cancers included colorectal (14%), skin (12%), and breast (11%). Xenograft implantation (61%) in mice (96%) was the most common animal model. The use of preclinical reporting guidelines was listed in 0.4% of articles. Biological and technical replicates were completely reported in 1% of studies, statistics in 49%, randomization in 1%, blinding in 2%, sample size estimation in 0%, and inclusion/exclusion criteria in 0%. Overall, completeness of reporting in the preclinical OV therapy literature is poor. This may hinder efforts to interpret, replicate, and ultimately translate promising preclinical OV findings.

INTRODUCTION

Incomplete and non-transparent reporting of animal experiments[1] contributes to the irreproducibility of basic science studies.[2,3] Such irreproducibility may help explain high attrition rates in drug development.[4] To improve the level of reporting rigor and improve reproducibility, the NIH has identified essential items to be reported for all preclinical experimental animal research.[5] This core set of preclinical reporting guidelines (NIH-PRG) lists items that should, at a minimum, be included in a preclinical publication.[6] To properly evaluate, interpret, and reproduce an experiment's findings, readers need a clear understanding and appreciation for how the experiment was conceived, executed, and analyzed. This information is necessary to judge the validity of findings. Incomplete and poor reporting does not allow for an accurate appraisal of the experiment's value, and important findings (positive or negative) could be missed. As demonstrated in clinical research, the quality and rigor of research improves when the reporting of key methodological items improves.[7,8]

Cancer immunotherapy is a rapidly growing field.[9] Oncolytic, or "cancer-killing," viruses (OVs) comprise a promising therapeutic platform that can elicit anti-cancer immune responses.[10–13] Despite the potential benefits of OV therapy and a multitude of preclinical candidates, it should be noted that there is currently only one OV approved for human use by the U.S. Food and Drug Administration and European Medicines Agency (Imlygic for advanced melanoma, approved in 2015).[14] The current state of OV preclinical study design and reporting is unknown, as is whether this could explain in part the relatively small number of approved candidates, despite the first reported studies of the modern OV era being published nearly 30 years ago.[15]

Our primary objectives were to review and assess the rigor in the design and completeness of reporting of preclinical *in vivo* OV therapy studies by applying the recommendations of the NIH-PRG. Furthermore, our study sought to produce an evidence map highlighting the recent global state of OV research and identify commonly studied viruses, cancers, and animal models. To address this, we used rigorous methods commonly used for systematic reviews, including

Received 25 January 2019; accepted 9 May 2019;
https://doi.org/10.1016/j.omto.2019.05.004.

Correspondence: Manoj M. Lalu, MD, PhD, Clinical Epidemiology Program, BLUEPRINT Translational Research Group, Ottawa Hospital Research Institute, 501 Smyth Road, P.O. Box 201B, Ottawa, ON K1H 8L6, Canada.
E-mail: mlalu@toh.ca

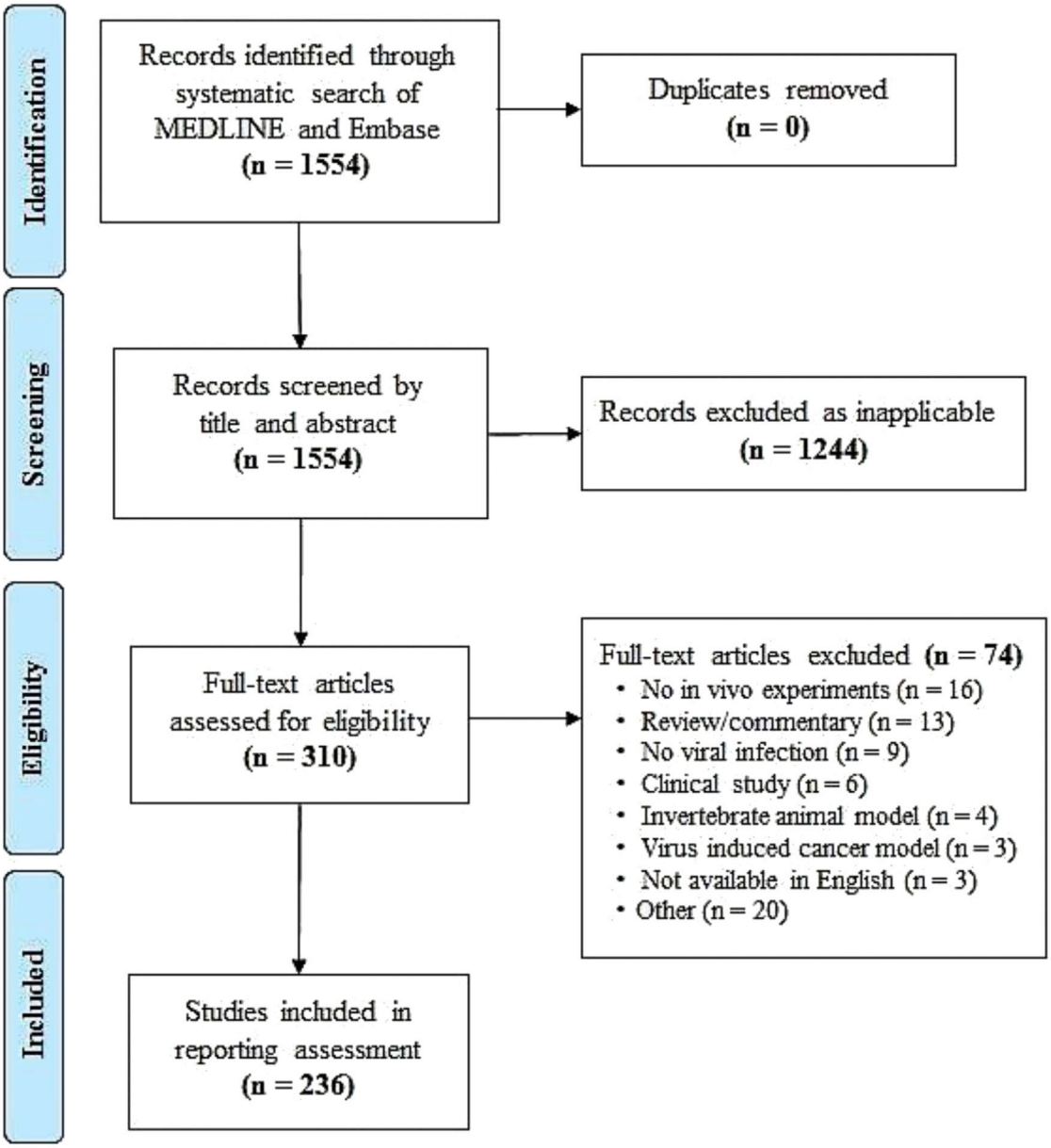

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Flow Diagram for Study Selection

developing an *a priori* protocol, a systematic search strategy to identify eligible articles, and screening of articles and extraction of data performed independently and in duplicate. Our appraisal of basic experimental design concepts and key reporting elements aims to increase transparency in preclinical OV research and to improve "bench-to-bedside" translation for OV immunotherapies.

RESULTS

Study Selection

Our search identified 1,554 records. Title and abstract screening excluded 1,244 records, with a subsequent full-text screening excluding an additional 74 records. Study screening and selection are summarized in Figure 1. Two-hundred thirty-six articles were included in our review; the full list and individual data of the included studies can be found on Open Science Framework (OSF; https://osf.io/j2dwm/).

Epidemiology of Studies

Articles came from 20 different countries, based on the corresponding author's residency at the time of publication (Figure 2). Most common were the United States (n = 72, 31%) and China (n = 63, 27%). The articles were published in 2016 (n = 169,

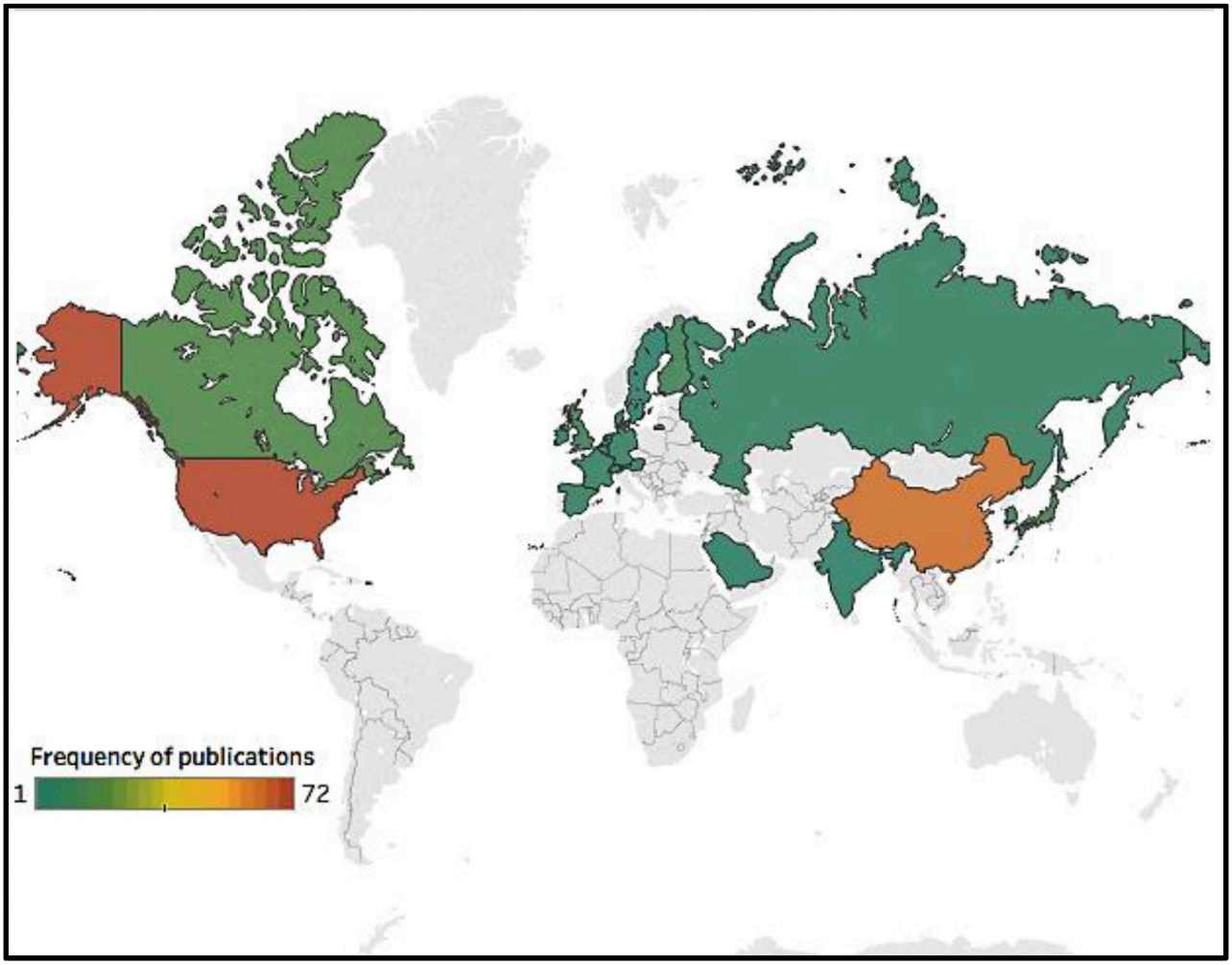

Figure 2. Country of Publication

Information based on the corresponding author's residency at the time of the included article publication (image created using Tableau Software, Seattle, WA, USA).

72%) and the first 6 months of 2017 (n = 67, 28%) in 85 journals. The three most commonly acknowledged sources of research funding were government bodies (n = 196, 43%), academic institutions (n = 107, 24%), and foundations/charities (n = 94, 21%; total sources of funding N = 455). The most commonly reported outcome was tumor size/burden (n = 204, 32%), followed by animal survival (n = 116, 18%), viral infectivity (n = 109, 17%), and host antitumor response (n = 94, 15%; total number of reported outcomes N = 630). Seventy-six (32%) articles had titles that clearly indicated that they contained preclinical experiments. The full results for study design and publication characteristics can be found in Table S1.

In total, 26 different viral platforms were used. The most common was adenovirus (n = 104, 43%) followed by vaccinia virus (n = 34, 14%) and herpes simplex virus (n = 27, 11%; N = 241). The viral treatment most commonly used was intratumoral injection (n = 176, 68%) of an actively replicating virus (n = 182, 77%) in a monotherapy regimen (n = 152, 59%).

Cancerous cell lines from 23 distinct types of tissue were used (N = 293). The most common were colorectal (n = 42, 14%), skin (n = 35, 12%), breast (n = 33, 11%), liver (n = 28, 10%), brain (n = 25, 9%), and lung (n = 25, 9%). This corresponds with the incidence of these cancers in humans.[16] The majority of cancer models used heterotopic (n = 170, 66%; N = 259), xenograft (n = 153, 61%; N = 252), and implant (n = 235, 99%; N = 238) methods.

The animal models were based on four species, with the majority of experiments using mice (n = 231, 96%; N = 241). In total, 26,044 animals were used; 25,516 of these were mice. Explicit ethics approval for animal use was referenced in the majority of articles (n = 172, 73%). The housing and handling of animals was described in 104

articles (44%). The full results for viral, cancer, and animal model data can be found in Table S2.

Assessment of Reporting

Completeness of reporting against each of the seven NIH-PRG domains was assessed. Each section is labeled by the domains it addresses, and a brief summary of the items recommended by the NIH-PRG is provided to guide and orient the reader. The full results for our reporting assessment can be found in Figure 3 and Table S3.

The Listing of Reporting Guidelines

The use of community-based nomenclature and reporting standards are encouraged in the NIH-PRG.[5] One article (0.4%) listed the use of reporting guidelines (Animal Research: Reporting *In Vivo* Experiments [ARRIVE])[17] during study design and manuscript preparation.

Measurement Techniques

Authors are required to report on whether their findings were substantiated under a range of conditions (e.g., different viral dosages and different models).[5] This was reported in 225 articles (95%). Furthermore, the number of subjects used and measurements performed should be sufficiently described, such that a distinction between biological and technical replicates can be made.[5] One-hundred five articles (44%) reported the number of times a measurement was repeated for at least one outcome (e.g., tumor size was measured with calipers three times at each time point). Two articles (1%) reported the number of times a measurement was repeated for *all* experimental outcomes. Similarly, the number of animals used to measure each outcome was reported in 25 articles (17%).

Statistics and Sample Size Estimation

Statistical tests used, the exact value of n, and defined measures of central tendency and dispersion must be included.[5] Statistical tests used were listed in 225 (95%) articles, the exact value of n in 148 (63%) articles, and measures of central tendency and dispersion in 203 (86%) and 206 (87%) articles, respectively. All four items were reported on in 116 articles (49%).

To determine an appropriate number of animals, an *a priori* power calculation is required.[5] The primary outcome, sample size used, and the rationale for the sample size must be reported.[5] A primary outcome was explicitly stated in 12 articles (5%). A sample size calculation was listed in 10 articles (4%); of these, 4 articles (2%) included the method of calculation (e.g., software package). No articles reported all items related to sample size estimation.

Randomization and Blinding

To reduce selection bias, the randomization of animals and the method of randomization must be reported.[5] This is only applicable when a study includes multiple experimental arms, which occurred in 234 articles (99%). Random group assignment was stated in 103 articles (44%); of these, the method of randomization (e.g., use of a software package) was described in 2 articles (1%).

In addition, to reduce performance and selection bias, the NIH-PRG recommends that authors report whether experimenters were blinded to group assignment and outcome assessment, respectively.[5] Blinding of personnel conducting experiments was listed in 7 articles (3%) and blinding of personnel assessing outcomes was listed in 15 articles (6%). Complete blinding (i.e., reporting of all items) was listed in 5 articles (2%).

Inclusion and Exclusion Criteria

Criteria for the exclusion of any data (e.g., animals, results) must be reported to minimize selection bias.[5] Furthermore, authors are expected to include information regarding the total number of animals originally procured for the study.[5] The total number of animals used was included in 7 articles (3%). The exclusion of any data, or lack thereof, was listed in 11 articles (5%). To address selective outcome reporting, the experimental design (as described in the Materials and Methods section) was compared to the reported results; 228 articles (97%) listed all the findings. Furthermore, the NIH-PRG requires any pilot or preliminary experiments, as well as any null or negative results to be reported.[5] Pilot studies were reported in 27 articles (11%). Null or negative results were explicitly described in 96 articles (41%). No articles reported on all of the items recommended in the inclusion/exclusion criteria domain.

DISCUSSION

This review was a comprehensive appraisal of the recent preclinical *in vivo* OV immunotherapy literature. Included studies were published in nearly 100 journals and originated from 20 countries. Twenty-six different viral platforms (and hundreds of genetically modified strains) were used. Twenty-three cancer models (and hundreds of cell lines) were investigated.

Our reporting assessment indicated that basic components of experimental design, such as randomization and blinding, as well as key elements of study methodology, such as sample size estimation and *a priori*-defined inclusion/exclusion criteria, are poorly reported. Of the 21 items identified from the NIH-PRG recommendations, the vast majority were absent across included studies. The statistics domain was partially reported, with 49% of studies reporting all items. Across the other six NIH-PRG domains complete reporting was 2%, at most. Collectively, these results demonstrate that the quality of reporting in the preclinical *in vivo* OV therapy literature is poor and the publication of the NIH-PRG has not yet had a substantive influence.[18]

Numerous guidelines for thorough *in vivo* preclinical study design and reporting exist.[19] This includes the extensive ARRIVE preclinical reporting guidelines that were published in 2010 (and included all elements of the NIH-PRG).[17] Those guidelines were endorsed by hundreds of journals (including many in our sample), learned societies, and funders; however, adherence has been poor.[20–22] In response, the simplified NIH-PRG were created in order to focus on seven key items to report; our study demonstrates that adherence to even these core reporting elements continues to be poor. Several studies have demonstrated common trends of poor reporting. Similar to

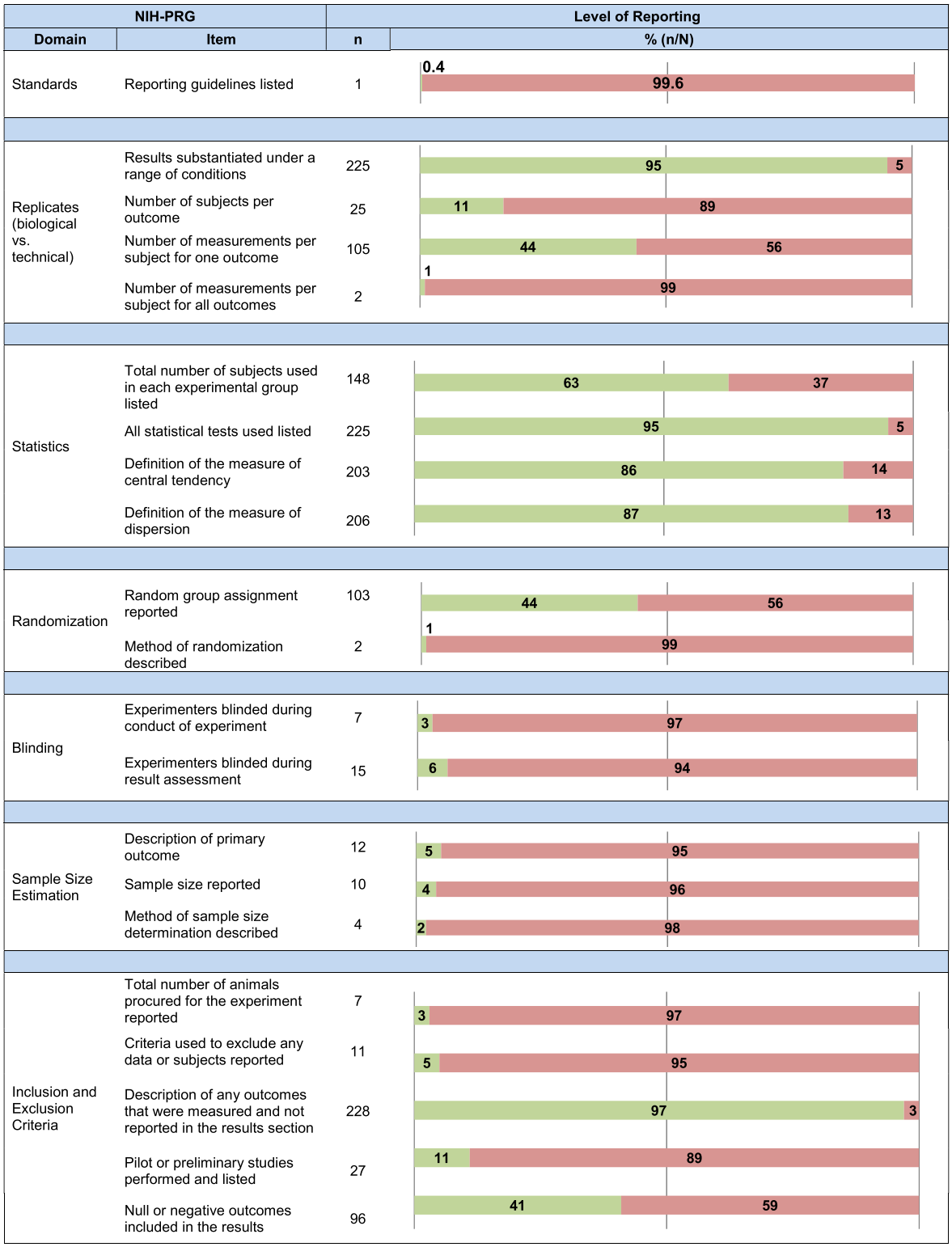

Figure 3. Reporting Assessment Results

Completeness of reporting across all included studies (N = 236) against the deconstructed NIH preclinical reporting guidelines (NIH-PRG) where n is the number of times each item was reported. Green and red correspond to an item being reported or not reported, respectively.

our study, Macleod et al.[23] showed that reporting both on randomization and the method of randomization was quite low (1.8%) in *in vivo* studies published by the *Nature* publication group; interestingly, after mandatory completion of a reporting checklist was instituted, the rate increased to 11.2% in these journals. In other studies, randomization was reported between 0% and 16% of included articles, depending on the journal and topic.[20,22] Comparable results have also been found for blinding, with Macleod et al.[23] finding 4% of articles in their sample reporting on blinding of either experimenter or outcome assessors (which increased to 22.8% after institution of a mandatory checklist). Baker et al.[20] found approximately 20% of journals reported on blinding in their sample of studies. It is important to note that, even if a study did not employ randomization, blinding, sample size calculations, and other factors, the final publication could still easily adhere to the guidelines by transparently reporting that these methods were not used.

The Importance of Methodological Rigor and Transparent Reporting

The importance of rigorous study design and transparent reporting needs to be encouraged. Omission of key concepts in the preclinical setting has been associated with the exaggeration of effect sizes, type I (i.e., false positive) and type II (i.e., false negative) errors, and overestimation of the potential for successful bench-to-bedside translation.[24–26] Incorporating methodological rigor into study design and reporting has been shown to reduce bias in the clinical setting.[27–30] Accordingly, many commentators argue that fundamental concepts of clinical trial methodology, such as randomization and blinding, should be incorporated into preclinical experiments to increase study validity.[31] These practices allow other researchers to evaluate the risk of bias in study reports: randomization addressing selection bias and blinding addressing performance bias. Within the cancer field, the importance of these design elements and reporting is evidenced by recent preclinical systematic reviews and meta-analyses of sunitinib[32] and sorafenib.[33] These studies suggested that poor design and incomplete reporting of preclinical studies likely contributed to overestimated effect sizes in animal experiments relative to subsequent clinical studies in human patients. Within this context, our assessment of the OV literature is an important one that should be considered as efforts to clinically translate OV therapies continue. Within the present study, important methods to reduce bias, such as randomization and blinding were reported in only a minority of studies. It is unclear whether these items are not considered and are therefore omitted from study design or they are performed and simply not reported—a push toward complete reporting would shed light on this.

Strengths and Limitations

The choice of the NIH-PRG for our reporting tool and the method through which the recommendations were deconstructed and items operationalized into a checklist are supported by the expertise of the multidisciplinary research team, including basic scientists, who performed this study. Portions of this checklist were previously used by members of our group.[34] Furthermore, to ensure that the relevant study characteristics were correctly collected, our team included several OV experts (C.I., J.S.D., R.A., N.E.-S., R.K., and B.A.K.).

Regarding potential limitations, the validity of included studies and their risk of bias was not evaluated; this limitation reflects reporting being the primary focus of our study. If risk of bias was evaluated, we would anticipate that most studies would be deemed "unclear risk of bias" in a majority of domains, due to a lack of reporting.[35] Second, it is possible that researchers are aware of reporting guidelines and rigorous study design, but interpreted items differently or simply did not report measures that were performed. Third, authors from countries outside of the United States and those who do not receive funding from the NIH may be unaware of these guidelines. Finally, a number of articles (n = 29, 12%) were published in *Oncotarget*, a journal that was recently delisted in MEDLINE.[36] However, when comparing the level of reporting in *Oncotarget* to all other articles, we found no meaningful differences (Table S4).

Future Steps

We believe the impetus for change will need to come through enforcement by external sources, such as funding agencies and journals (editorial boards, in particular).[21,23] Examples include mandatory checklists at the time of submission[37] or requirement that the study be preregistered[38] on an open-source database prior to study commencement (similar to clinical trials).

Our study demonstrates that researchers as a community are either not yet aware of reporting guidelines or have not yet appreciated their significance. That is, they may be aware of some of these concepts (e.g., randomization and blinding), but may not appreciate the consequence that less rigorous design decisions have on results (e.g., exaggerating effect size). Here, we posit a few explanations. First, a contributing factor may be the lack of available formal training opportunities for basic scientists concerning these design and reporting elements. Clear guidelines on *how* these methods should be conducted and reported need to be established. Nonetheless, even in the absence of implementation of these methods, authors should be able to report that they were not used. Second, the impracticality of implementing these measures in some situations should be considered. For example, a single experimenter cannot be blinded to group allocation or it may be impossible to fully blind experiments (e.g., blinding may not be possible in animals treated with OVs, as they develop clearly visible signs associated with the virus). Again, in these situations, a lack of blinding could still be reported, for transparency. Third, studies adhering to a more rigorous design may require more resources (e.g., extra personnel to blind appropriately), access to which may vary widely among research groups and implementation of which may not be rewarded or compensated for by funding agencies at present. This, however, still does not interfere with transparent reporting regarding the application of these design elements. Last, it is possible that the lack of uptake of reporting guidelines, by other researchers and journal editors discourages any change of practices common in the laboratory. Without

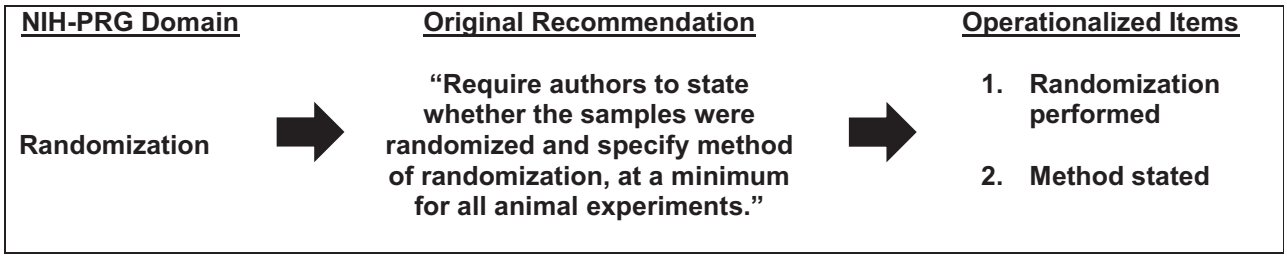

Figure 4. Constructing the Reporting Checklist

The NIH Preclinical Reporting Guidelines (NIH-PRG) domain of randomization was deconstructed into two unidimensional items, and each was operationalized into a "yes" or "no" question.

tangible incentives, the time required for researchers to implement more rigorous study methodologies and fully report studies may be outweighed by pressures to publish rapidly.

Conclusions

This review is an extensive reporting assessment against an important set of published guidelines in a growing field of cancer research. Given the critical need for new cancer treatments, efficient allocation of resources, and reproducible experiments, it is crucial that preclinical *in vivo* OV research be conducted and reported in a transparent, methodical, and rigorous way. The items investigated here are fundamental to clinical research and have demonstrated benefits in preclinical research. Future work will need to determine how these guidelines and their underlying principles should be applied to preclinical research.

MATERIALS AND METHODS

Protocol

Using the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocol (PRISMA-P) statement[39] as a guide, a study protocol was constructed and deposited on OSF (OSF; https://osf.io/j2dwm/).[38] During completion of our review, there were no deviations from the protocol. The protocol was publicly posted concurrent with the commencement of data extraction. Although not a systematic review, the PRISMA statement[40] was used in the drafting of this manuscript.

Eligibility Criteria

We included all preclinical *in vivo* studies that used an OV therapy (prophylactic or interventional) to treat cancer, published between January 2016 and June 2017 (18 months). There were no limitations on viral platform, cancer model, animal model, comparisons, outcomes, or experimental design. Only original research, in the form of full-text English publications, was included. Abstracts, letters, reviews, and commentaries were excluded. Any *in vitro*, *ex vivo*, or clinical experiments were excluded.

Search Strategy and Article Screening

In consultation with an information specialist, a systematic search strategy was developed (OSF; https://osf.io/j2dwm/), and searches of MEDLINE and Embase were performed. Titles that met our eligibility criteria were identified and uploaded to an audit-ready, cloud-based software (Distiller SR; Evidence Partners, Ottawa, ON, Canada). Two independent reviewers performed the process of study selection. Title and abstract screenings were performed independently and in duplicate using an accelerated screening method (one reviewer required to include, two reviewers required to exclude).[41] A subsequent full-text screening was performed in duplicate by two independent reviewers (consensus required to exclude). Any conflicts were resolved through discussion with a third team member.

Data Extraction

Full-text articles which met our inclusion criteria were retrieved and uploaded to Distiller SR. All extractors (N.L.W., G.J.L., J.L.M., and I.C.) were affiliated with the Ottawa Hospital Research Institute. To ensure an adequate level of reviewer agreement, all extractors participated in a pilot training exercise. Each extractor was provided with a training document and independently extracted five articles. Feedback from a core group member (M.M.L.) was given, and a further five articles were independently extracted. At this point, the training exercise was completed, as inter-rater agreement met our pre-defined level of 80%. Two trained members of the extractor team independently extracted data from each article in duplicate. Data were collected concerning common study characteristics (e.g., country of corresponding author, journal, funding source, virus administered, cancer studied, and animal model used) and the level of reporting against the NIH-PRG domains (e.g., randomization and blinding). Conflicts were resolved through consensus discussions between extractors or with a senior investigator when needed. Online methods and supplemental information were retrieved whenever referenced.

Reporting Quality Assessment of Included Studies

The NIH-PRG recommendations are grouped into seven domains: (1) use of community-based reporting standards, (2) distinction between biological and technical replicates, (3) statistics, (4) sample size estimation, (5) randomization, (6) blinding, and (7) inclusion/exclusion criteria.[5] Each of these domains contains a multifaceted recommendation that addresses multiple points. To assess the level of reporting, each domain was deconstructed into unidimensional items. These items were operationalized into 21 "yes" or "no" questions, each addressing a single point. (The full list of questions is contained in our protocol.[42]) For example, the domain of randomization

was deconstructed into questions regarding whether random group assignment was reported and whether the method of randomization was described (Figure 4). All items mentioned in the NIH-PRG were considered relevant to our reporting assessment. This 21-item checklist served as our reporting assessment tool.

Data Synthesis

For each item addressed by the NIH-PRG recommendations, the level of reporting was expressed as the total number of times each item was reported (n) across all studies (N). This descriptive statistic, frequency count, is expressed nominally (n out of N) and as a percentage (n/N). No formal statistical analysis was performed.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.omto.2019.05.004.

AUTHOR CONTRIBUTIONS

Conceptualization and methodology: M.M.L., J.P., K.D.C., J.-S.D., R.A., J.K., C.I., and D.A.F.; validation: N.E.-S., R.K., and B.A.K.; formal analysis: M.M.L., N.L.W., and D.A.F.; investigation: M.M.L., N.L.W., G.J.L., J.L.M., I.C., J.P., K.D.C., J.-S.D., R.A., J.K., N.E.S., R.K., B.A.K., C.I., and D.A.F.; resources: M.M.L. and D.A.F.; data curation: N.L.W., G.J.L., J.L.M., and I.C.; writing–original draft preparation: M.M.L., N.L.W., and D.A.F.; writing–review and editing: all authors; and funding acquisition: M.M.L. and D.A.F.

CONFLICTS OF INTEREST

The authors declare no competing interests.

ACKNOWLEDGMENTS

The authors thank Risa Shorr, Information Specialist and Librarian at The Ottawa Hospital for performing the systematic literature search. This work was supported by Biotherapeutics for Cancer Treatment (BioCanRx, Government of Canada funded Networks of Centres of Excellence) under Clinical, Social, and Economic Impact Grant CSEI3 awarded to M.M.L. and D.A.F.

REFERENCES

1. Kilkenny, C., Parsons, N., Kadyszewski, E., Festing, M.F., Cuthill, I.C., Fry, D., Hutton, J., and Altman, D.G. (2009). Survey of the quality of experimental design, statistical analysis and reporting of research using animals. PLoS ONE *4*, e7824.
2. Jarvis, M.F., and Williams, M. (2016). Irreproducibility in preclinical biomedical research: Perceptions, uncertainties, and knowledge gaps. Trends Pharmacol. Sci. *37*, 290–302.
3. Ramirez, F.D., Motazedian, P., Jung, R.G., Di Santo, P., MacDonald, Z.D., Moreland, R., Simard, T., Clancy, A.A., Russo, J.J., Welch, V.A., et al. (2017). Methodological Rigor in Preclinical Cardiovascular Studies: Targets to Enhance Reproducibility and Promote Research Translation. Circ. Res. *120*, 1916–1926.
4. Levin, L.A., and Danesh-Meyer, H.V. (2010). Lost in translation: bumps in the road between bench and bedside. JAMA *303*, 1533–1534.
5. NIH. (2014). Principles and guidelines for reporting preclinical research. https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research.
6. Landis, S.C., Amara, S.G., Asadullah, K., Austin, C.P., Blumenstein, R., Bradley, E.W., Crystal, R.G., Darnell, R.B., Ferrante, R.J., Fillit, H., et al. (2012). A call for transparent reporting to optimize the predictive value of preclinical research. Nature *490*, 187–191.
7. Turner, L., Shamseer, L., Altman, D.G., Schulz, K.F., and Moher, D. (2012). Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. Syst. Rev. *1*, 60.
8. Pussegoda, K., Turner, L., Garritty, C., Mayhew, A., Skidmore, B., Stevens, A., Boutron, I., Sarkis-Onofre, R., Bjerre, L.M., Hróbjartsson, A., et al. (2017). Systematic review adherence to methodological or reporting quality. Syst. Rev. *6*, 131.
9. Rosenberg, S.A., Yang, J.C., and Restifo, N.P. (2004). Cancer immunotherapy: moving beyond current vaccines. Nat. Med. *10*, 909–915.
10. Russell, S.J., Peng, K.W., and Bell, J.C. (2012). Oncolytic virotherapy. Nat. Biotechnol. *30*, 658–670.
11. Varghese, S., and Rabkin, S.D. (2002). Oncolytic herpes simplex virus vectors for cancer virotherapy. Cancer Gene Ther. *9*, 967–978.
12. Miest, T.S., and Cattaneo, R. (2013). New viruses for cancer therapy: meeting clinical needs. Nat. Rev. Microbiol. *12*, nrmicro3140.
13. Kirn, D. (2001). Oncolytic virotherapy for cancer with the adenovirus dl1520 (Onyx-015): results of phase I and II trials. Expert Opin. Biol. Ther. *1*, 525–538.
14. Andtbacka, R.H., Kaufman, H.L., Collichio, F., Amatruda, T., Senzer, N., Chesney, J., Delman, K.A., Spitler, L.E., Puzanov, I., Agarwala, S.S., et al. (2015). Talimogene laherparepvec improves durable response rate in patients with advanced melanoma. J. Clin. Oncol. *33*, 2780–2788.
15. Liu, T.-C., Galanis, E., and Kirn, D. (2007). Clinical trial results with oncolytic virotherapy: a century of promise, a decade of progress. Nat. Clin. Pract. Oncol. *4*, 101–117.
16. The United States Cancer Statistics Working Group. (2010). United States cancer statistics: 1999–2006 incidence and mortality web-based report. Atlanta: US Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. https://wonder.cdc.gov/wonder/help/CancerMIR-v2013.html.
17. Kilkenny, C., Browne, W.J., Cuthill, I.C., Emerson, M., and Altman, D.G. (2010). Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. PLoS Biol. *8*, e1000412.
18. Moher, D., Avey, M., Antes, G., and Altman, D.G. (2015). The National Institutes of Health and guidance for reporting preclinical research. BMC Med. *13*, 34.
19. Henderson, V.C., Kimmelman, J., Fergusson, D., Grimshaw, J.M., and Hackam, D.G. (2013). Threats to validity in the design and conduct of preclinical efficacy studies: a systematic review of guidelines for in vivo animal experiments. PLoS Med. *10*, e1001489.
20. Baker, D., Lidster, K., Sottomayor, A., and Amor, S. (2014). Two years later: journals are not yet enforcing the ARRIVE guidelines on reporting standards for pre-clinical animal studies. PLoS Biol. *12*, e1001756.
21. Hair, K., Macleod, M.R., and Sena, E.S. (2018). A randomised controlled trial of an Intervention to Improve Compliance with the ARRIVE guidelines (IICARus). bioRxiv, https://doi.org/10.1101/370874.
22. Gulin, J.E.N., Rocco, D.M., and García-Bournissen, F. (2015). Quality of reporting and adherence to ARRIVE guidelines in animal studies for Chagas disease preclinical drug research: a systematic review. PLoS Negl. Trop. Dis. *9*, e0004194.
23. The NPQIP Collaborative Group (2019). Did a change in Nature journals' editorial policy for life sciences research improve reporting? BMJ Open Science *3*, e000035.
24. Sena, E., van der Worp, H.B., Howells, D., and Macleod, M. (2007). How can we improve the pre-clinical development of drugs for stroke? Trends Neurosci. *30*, 433–439.
25. Macleod, M.R., van der Worp, H.B., Sena, E.S., Howells, D.W., Dirnagl, U., and Donnan, G.A. (2008). Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. Stroke *39*, 2824–2829.
26. Bebarta, V., Luyten, D., and Heard, K. (2003). Emergency medicine animal research: does use of randomization and blinding affect the results? Acad. Emerg. Med. *10*, 684–687.
27. Sacks, H., Chalmers, T.C., and Smith, H., Jr. (1982). Randomized versus historical controls for clinical trials. Am. J. Med. *72*, 233–240.

28. Ioannidis, J.P., Haidich, A.B., Pappa, M., Pantazis, N., Kokori, S.I., Tektonidou, M.G., Contopoulos-Ioannidis, D.G., and Lau, J. (2001). Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA *286*, 821–830.

29. Colditz, G.A., Miller, J.N., and Mosteller, F. (1989). How study design affects outcomes in comparisons of therapy. I: Medical. Stat. Med. *8*, 441–454.

30. Miller, J.N., Colditz, G.A., and Mosteller, F. (1989). How study design affects outcomes in comparisons of therapy. II: Surgical. Stat. Med. *8*, 455–466.

31. Ioannidis, J.P., Greenland, S., Hlatky, M.A., Khoury, M.J., Macleod, M.R., Moher, D., Schulz, K.F., and Tibshirani, R. (2014). Increasing value and reducing waste in research design, conduct, and analysis. Lancet *383*, 166–175.

32. Henderson, V.C., Demko, N., Hakala, A., MacKinnon, N., Federico, C.A., Fergusson, D., and Kimmelman, J. (2015). A meta-analysis of threats to valid clinical inference in preclinical research of sunitinib. eLife *4*, e08351.

33. Mattina, J., MacKinnon, N., Henderson, V.C., Fergusson, D., and Kimmelman, J. (2016). Design and reporting of targeted anticancer preclinical studies: a meta-analysis of animal studies investigating sorafenib antitumor efficacy. Cancer Res. *76*, 4627–4636.

34. Canadian Critical Care Translational Biology Group (2016). The devil is in the details: incomplete reporting in preclinical animal research. PLoS ONE *11*, e0166733.

35. Sena, E.S., Currie, G.L., McCann, S.K., Macleod, M.R., and Howells, D.W. (2014). Systematic reviews and meta-analysis of preclinical studies: why perform them and how to appraise them critically. J. Cereb. Blood Flow Metab. *34*, 737–742.

36. McCook, A. (2017). Widely used US government database delists cancer journal. Retraction Watch. http://retractionwatch.com/2017/10/25/widely-used-u-s-government-database-delists-cancer-journal/.

37. Han, S., Olonisakin, T.F., Pribis, J.P., Zupetic, J., Yoon, J.H., Holleran, K.M., Jeong, K., Shaikh, N., Rubio, D.M., and Lee, J.S. (2017). A checklist is associated with increased quality of reporting preclinical biomedical research: A systematic review. PLoS ONE *12*, e0183591.

38. Nosek, B.A., Ebersole, C.R., DeHaven, A.C., and Mellor, D.T. (2018). The preregistration revolution. Proc. Natl. Acad. Sci. USA *115*, 2600–2606.

39. PRISMA-P Group (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ *350*, g7647.

40. Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gøtzsche, P.C., Ioannidis, J.P., Clarke, M., Devereaux, P.J., Kleijnen, J., and Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med. *6*, e1000100.

41. Khangura, S., Konnyu, K., Cushman, R., Grimshaw, J., and Moher, D. (2012). Evidence summaries: the evolution of a rapid review approach. Syst. Rev. *1*, 10.

42. Wesch, N.L., Lalu, M.M., Fergusson, D.A., Presseau, J., Cobey, K.D., Diallo, J.-S., Kimmelman, J., Bell, J., Bramson, J., Ilkow, C., et al. (2017). IMproving Preclinical Reporting in Oncolytic Virus Experiments (IMPROVE). Open Science Framework. https://osf.io/j2dwm/.

OMTO, Volume *14*

Supplemental Information

Assessing the Completeness of Reporting in Preclinical Oncolytic Virus Therapy Studies

Dean A. Fergusson, Neil L. Wesch, Garvin J. Leung, Jenna L. MacNeil, Isidora Conic, Justin Presseau, Kelly D. Cobey, Jean-Simon Diallo, Rebecca Auer, Jonathan Kimmelman, Natasha Kekre, Nader El-Sayes, Ramya Krishnan, Brian A. Keller, Carolina Ilkow, and Manoj M. Lalu

Supplementary Online Material

Table S1. Study design and publication characteristics.

Question	Response	Frequency
Indicate the corresponding author's country of residence.	United States	72
	China	63
	Canada	16
	Japan	15
	Finland	11
	South Korea	10
	United Kingdom	9
	Spain	8
	Germany	7
	France	6
	Russia	5
	Sweden	3
	Austria	2
	Denmark	2
	India	2
	Belgium	1
	Ireland	1
	Netherlands	1
	Saudi Arabia	1
	Switzerland	1
Indicate the year of publication.	2016	169
	2017	67
Indicate all contributing forms of research funding.	Government	196
	Academic	107
	Foundation / Charity	94
	Industry	39
	Other (individual donations)	12
	No mention of funding	5
	Explicitly unfunded	2
What is the journal of publication?	Oncotarget	29
	Molecular Therapy – Oncolytics	23
	Gene Therapy	12
	Molecular Therapy	10
	Cancer Gene Therapy	10
	OncoImmunology	9
	Clinical Cancer Research	7
	Molecular Cancer Therapeutics	7
	PLoS One	6
	Human Gene Therapy	5
	Neuro-Oncology	5
	Nature Communications	4

	Cancer Research	4
	Vaccine	4
	Cancer Letters	4
	Journal of Controlled Release	3
	Scientific Reports	3
	International Journal of Oncology	3
	International Journal of Cancer	3
	Oncology Reports	3
	Journal of Virology	3
	Viruses	3
	Cancer Biology	2
	Acta Biomaterialia	2
	Virus Research	2
	Cancer Science	2
	Journal of Cancer	2
	ACS Nano	2
	Journal of Translational Medicine	2
	Blood	2
	Journal of Pharmacological Sciences	2
	Journal of General Virology	2
	Cancer Immunology Immunotherapy	2
	Apoptosis	2
	Virology	2
	Molecular Medicine Reports	1
	Molecular Medicine	1
	Medical Oncology	1
	Journal of Cellular and Molecular Medicine	1
	Journal of Gene Medicine	1
	Journal of the National Cancer Institute	1
	Cancer Immunology Research	1
	Therapeutics and Clinical Risk Management	1
	Surgery	1
	The Journal of Immunology	1
	The Journal of Nuclear Medicine	1
	Virus Genes	1
	Veterinary and Comparative Oncology	1
	Applied Materials	1
	Molecular Oncology	1
	Drug Design, Development and Therapy	1
	OncoTargets and Therapy	1
	Molecular Therapy — Methods	1

	Acta Pharmacologica Sinica	1
	Translational Oncology	1
	Breast Cancer Research	1
	American Journal of Cancer Research	1
	Stem Cell Reports	1
	Oncolytic Virotherapy	1
	Cellular Immunology	1
	Frontiers in Oncology	1
	BMC Cancer	1
	Reproductive Sciences	1
	Oncogene	1
	Biochemical and Biophysical Research Communications	1
	International Immunopharmacology	1
	Journal of Ovarian Research	1
	International Journal of Molecular Sciences	1
	Journal of Immunology Research	1
	Oral Oncology	1
	Tumor Biology	1
	Cellular Oncology	1
	Cell Reports	1
	EBioMedicine	1
	Journal of Experimental & Clinical Cancer Research	1
	The Canadian Journal of Veterinary Research	1
	Neurotherapeutics	1
	Head	1
	Molecular Genetics, Microbiology and Virology	1
	Current Gene Therapy	1
	Gut	1
	Russian Journal of Genetics: Applied Research	1
	International Journal of Clinical and Experimental Medicine	1
	Human Gene Therapy Clinical Development	1
	Journal of Biomedical Science	1
Which of the following outcomes were investigated?	Cancer size/burden	204
	Survival	116
	Viral Infectivity	109
	Anti-tumor response	94
	Cytokine response	55
	Tumor specificity/biodistribution	52
From the title alone, could you tell this was a preclinical article?	Yes	76
	No	160

Table S2. Epidemiology of viral, cancer, and animal models.

Question	Response	Frequency
What is the viral platform?	Adenovirus	104
	Vaccinia virus	34
	Herpes simplex virus	27
	Vesicular stomatitis virus	15
	Measles virus	11
	Newcastle disease virus	9
	Reovirus	8
	Myxoma virus	4
	Alphavirus	3
	Cytomegalovirus	3
	Maraba virus	3
	Parvovirus	3
	Lentivirus	2
	Picornavirus	2
	Tanapoxvirus	2
	Canine distemper virus	1
	Chicken infectious anemia virus	1
	Foot-and-mouth disease virus	1
	Human papilloma virus	1
	Influenza virus	1
	Lymphocytic choriomeningitis virus	1
	Mengovirus	1
	Mumps virus	1
	Poliovirus	1
	Semliki forest virus	1
	Tobacco mosaic virus	1
Is the virus replicating?	Yes	182
	No	31
	Unsure	26
What is the type of viral therapy?	Monotherapy	152
	Monotherapy + combination therapy	74
	Combination therapy	10
What is the route of viral administration?	Intratumoural	176
	Not intratumoural	77
	Unsure	7

Question	Response	Frequency
What type of cancer was induced?	Colorectal	42
	Skin	35
	Breast	33
	Liver	28
	Brain	25
	Lung	25
	Ovary	16
	Blood	15
	Pancreas	15
	Muscle	10
	Prostate	10
	Bladder	7
	Bone marrow	5
	Gastric	5
	Kidney	5
	Head and neck	4
	Bone	3
	Cervical	3
	Esophagus	2
	Mesothelium	2
	Nervous system	1
	Teratoma	1
	Thyroid	1
What was the method of cancer/tumor induction?	Implantation	235
	Transgenic	3
	Chemical	0
Indicate the type of implant.	Xenograft	153
	Syngeneic	99
What is the site of the tumor?	Heterotopic	170
	Orthotopic	83
	Unsure	6

Question	Response	Frequency	Number Used
What was the type of animal(s) used?	Mouse	231	25516
	Rat	5	516
	Hamster	4	NA
	Non-human primate	1	12

What type of animal model was used?	Immunodeficient	134
	Humanized	3
	Partially immunocompromised	16
	Unsure	5
Was ethics approval for animal use explicitly reported?	Yes	172
	No	64
Was the housing and handling of animals reported?	Yes	104
	No	132

Table S3. Reporting assessment results. The frequency of reporting of all 21 items deconstructed and operationalized from the National Institutes of Health preclinical reporting guidelines across the all included studies. These recommendations are separated into seven distinct domains.

Domain	Item Description	N=236 n (%) reported
Standards	Community based reporting guidelines listed	1 (0.4)
Replicates (biological vs. technical)	Results substantiated by repetition under a range of conditions	225 (95)
	Number of subjects stated for all experimental outcomes	25 (17)
	Number of measurements per subject for one experimental outcome stated	105 (44)
	Number of measurements per subject for all experimental outcomes stated	2 (1)
Statistics	List of the total number of subjects used in each experimental group	148 (63)
	List of all statistical tests used	225 (95)
	Definition of the measure of central tendency	203 (86)
	Definition of the measure of dispersion and precision	206 (87)
Randomization	Random group assignment reported	103 (44)
	Description of the method of random group assignment	2 (1)
Blinding	Experimenters blinded to group allocation during conduct of the experiment	7 (3)
	Experimenters blinded to group allocation during result assessment	15 (6)
Sample Size Estimation	Description of an *a priori* primary outcome	12 (5)
	Sample size computed during study design	10 (4)
	Description of the method of sample size determination	4 (2)
Inclusion and Exclusion Criteria	Total number of animals procured for the experiment reported	7 (3)
	Description of the criteria used for the exclusion of any data or subjects	11 (5)
	Description of any outcomes that were measured and not reported in the results section	228 (97)
	Pilot or preliminary studies performed and listed	27 (11)
	Null or negative outcomes included in the results	96 (41)

Table S4. Exploratory comparative analysis. A comparison in the completeness of reporting between *Oncotarget,* which was delisted by MEDLINE in 2016,[33] compared to all other articles included in our review.

		Oncotarget N=29	Other N=207
Domain	**Description**	**n (%) Reported**	
Standards	Reporting guidelines	0 (0)	1 (0)
Replicates	Range of conditions	29 (100)	192 (93)
	Number of subject for one outcome	15 (52)	133 (64)
	Number of subjects for all outcomes	4 (14)	21 (10)
	Number of measurements for one outcome	10 (34)	95 (46)
	Number of measurements for all outcomes	0 (0)	2 (1)
Statistics	Statistical tests used	25 (86)	200 (97)
	Measure of central tendency	24 (83)	179 (86)
	Measure of dispersion	23 (79)	183 (88)
Randomization	Multiple experimental groups	29 (100)	205 (99)
	Random group assignment	15 (52)	89 (43)
	Method of randomization	0 (0)	2 (1)
Blinding	Group allocation blinding	3 (10)	4 (2)
	Result assessment blinding	2 (7)	13 (6)
Sample Size Estimation	Primary outcome	2 (7)	10 (5)
	Sample size calculation	2 (7)	10 (5)
	Mathematical method used	0 (0)	4 (2)
Inclusion and Exclusion Criteria	Total number of subjects	0 (0)	7 (3)
	Data/subjects/results exclusion	0 (0)	11 (5)
	No result omission	29 (100)	199 (96)
	Preliminary studies	1 (3)	26 (13)
	Null/negative results	9 (31)	87 (42)