Review Article

Comparing Translational Success Rates Across Medical Research Fields – A Combined Analysis of Literature and Clinical Trial Data

Gwen van de Wall¹, Astrid van Hattem¹, Joy Timmermans², Merel Ritskes-Hoitinga³,⁴, André Bleich⁵ and Cathalijn Leenaars⁵

¹Department for Health Evidence, Radboud Institute for Health Sciences, Radboud University Medical Centre, Nijmegen, The Netherlands; ²PXL University of Applied Sciences and Arts, Hasselt, Belgium; ³Faculty of Veterinary Medicine, Department of Population Health Sciences, Utrecht University, Utrecht, The Netherlands; ⁴AUGUST, Department of Clinical Medicine, Aarhus University, Aarhus, Denmark; ⁵Institute for Laboratory Animal Science, Hannover Medical School, Hannover, Germany

Abstract

Many interventions that show promising results in preclinical development do not pass clinical tests. Part of this may be explained by poor animal-to-human translation. Using animal models with low predictability for humans is neither ethical nor efficient. If translational success shows variation between medical research fields, analyses of common practices in these fields could identify factors contributing to successful translation. We assessed translational success rates in medical research fields using two approaches: through literature and clinical trial registers.

Literature: We comprehensively searched PubMed for pharmacology, neuroscience, cancer research, animal models, clinical trials, and translation. After screening, 117 review papers were included in this scoping review. Translational success rates were not different within pharmacology (72%), neuroscience (62%), and cancer research (69%).

Clinical trials: The fraction of phase-2 clinical trials with a positive outcome was used as a proxy (i.e., an indirect resemblance measure) for translational success. Trials were retrieved from the WHO trial register and categorized into medical research fields following the international classification of disease (ICD-10). Of the phase-2 trials analyzed, 65.2% were successful. Fields with the highest success rates were disorders of lipoprotein metabolism (86.0%) and epilepsy (85.0%). Fields with the lowest success rates were schizophrenia (45.4%) and pancreatic cancer (46.0%).

Our combined analyses suggest relevant differences in success rates between medical research fields. Based on the clinical trials, comparisons of practice, e.g., between epilepsy and schizophrenia, might identify factors that influence translational success.

1 Introduction

Preclinical research in animals is currently still standard practice for the early phases of medical research, being required by most local authorities before clinical testing. These requirements assume that animal models can accurately predict human outcomes. In other words, they assume high animal-to-human translational success rates. However, preceding work (Leenaars et al., 2019) showed high variability in translational success rates in pharmacological research, ranging from 0% to 100%. Several analyzed factors (e.g., animal model species) could individually not predict translational success rates. Because many studies show low animal-to-human translation, not all currently used animal models might be suitable to predict human outcomes.

Over the past decade, the costs of developing new medications have been increasing, and funders, health care providers, and patients have been criticizing these high medical costs. However, the return on investment in the pharmaceutical industry has dropped continuously, at least up until the SARS-CoV-2 pandemic. This decreases the interest of companies to invest...
in medical research and hinders the development of new treatments. A substantial part of the problem may be explained by poor animal-to-human translation.

If the overall translational success rate is low, using animal models to predict safety or efficacy in humans is not ethical, nor does it promote responsible research and innovation. In this respect, animal models that do not contribute to clinical developments should be the first to be replaced. Because of this, and because improving translational success rates may prevent uninformative research from being performed and benefit return on investment, it is important to investigate the translational value of animal models.

An often-suggested factor potentially limiting translational success is poor experimental design (Kola and Landis, 2004; Bolker, 2017). However, as long as experimental design is poorly reported (McNamara et al., 2016; Albersheim et al., 2021), it is difficult to analyze its effects on translational success directly. An indirect approach towards translational success analysis may be to compare common practice between medical research fields. A prerequisite for this is the presence of variation in translational success rates between medical research fields, which allows analysis of the differences and possible etiological factors. Therefore, as a first step, translational success rates between different medical research fields were compared using two complementary approaches.

Our first approach was a scoping literature review analyzing translational success in several research fields. The methods and criteria used are complementary to previous research (Leenaars et al., 2019). Because a full analysis of all medical research fields using this approach was not viable, we selected 3 fields as case studies: neuroscience, pharmacology, and cancer research. Neuroscience has been criticized for its low translational success rates (e.g., O’Collins et al., 2006). However, when searching the literature, we could not find evidence-based reviews confirming this for the entire field. Thus, this criticism seems to be based mostly on personal experience (Hyman, 2012) and has only been confirmed by systematic analyses of actual data for acute stroke treatments (O’Collins et al., 2006). Pharmacology provides a broad spectrum of research and was chosen because this is one of the largest fields of research and the basis for the largest part of the Leenaars et al. (2019) paper. While cancer research is more specialized, cancer has the second highest global disease burden (9% of the total global disease burden)4, and many efforts and large budgets are invested to find new therapies for cancer5.

Our second approach investigates the success of clinical trials by research field. Trial success of phase-2 clinical trials, defined as the replication of statistically positive effects from preclinical animal models in phase-2 clinical trials, was chosen as a proxy to analyze translational success. We use the term “proxy” to reflect “an entity or variable used to model or generate data assumed to resemble the data associated with another entity or variable that is typically more difficult to research.”6 Phase-2 clinical trials were selected because they are relatively comparable to (and have the same limitations as) animal experiments, both of which assess the safety and efficacy of treatments in a small and homogeneous study population. For the interpretation of the analyses, we assume that clinical trials are mostly based on successful preclinical research. Thus, the overall percentage of successful trial outcomes is assumed to correlate with translational success. While this assumption will not always hold, we do not expect relevant differences between research fields in, e.g., the number of trials started without animal research. Thus, this approach allows for an indirect comparison of success between fields. All fields of research were included and categorized according to the 10th version of the International Classification of Diseases (ICD-10).

2 Methods

2.1 Scoping literature review

No protocol was posted for this scoping review. An internal short protocol (in Dutch) was shared between 3 of the co-authors before the start of screening.

Search

We searched the PubMed database on January 26, 2021. The search consisted of 5 elements: Selected fields (neuroscience, pharmacology, and cancer research), animal models, clinical trials, translation, and review. These elements were combined with “AND”: Fields AND animal models AND clinical trials AND translation AND review.

The old Systematic Review Centre for Laboratory animal Experimentation (SYRCLE) animal search filters (updated after our search date (van der Mierden et al., 2022)) were used (Hooijmans et al., 2010) to search for the animal models. Furthermore, to search for clinical trials and translation, we adjusted previously published search strings (Leenaars et al., 2019). New search strings were developed for the fields. For each field, appropriate MeSH terms were selected, and synonyms were identified though Google searches. The full search string is provided in Appendix A of the supplementary file7. All retrieved references were downloaded and stored as an EndNote X9 (Clarivate Analytics, Philadelphia, PA, USA) database.

Selection of papers

Retrieved references were screened for suitability in two phases. Phase one consisted of title/abstract screening. Publications that were reviews, included both in vivo animal data (authors could refer to animal studies, in vivo studies or specific animal species) and clinical data, assessed animal-to-human translation, were written in a language the authors could read (i.e., English, Dutch, French, Italian, Spanish, German, Swedish, or Danish), were published from 2017 through 2019 (to supplement previous
research (Leenaars et al., 2019), and were in the research field of neuroscience, pharmacology, or cancer research were selected for further full-text screening. As there is overlap between the field of pharmacology with the other two fields, the pharmacology part of our literature study was restricted to “other pharmacology studies”, excluding pharmacological neuroscience or cancer research, which were assessed separately. Articles that complied with the criteria for title/abstract screening and provided sufficient evidence were included in the final analysis. Sufficient evidence was operationalized as describing the findings from at least 2 in vivo preclinical and 2 clinical references. Screening was performed by two independent researchers. In the title/abstract phase, discrepancies were resolved by a third independent researcher. In the full-text phase, discrepancies were resolved through discussion between the screeners.

Data extraction
We extracted data on bibliography, study design, animal models, and translational success rates from the included papers as summarized in Table 1.

We defined translational success as the replication of positive, negative, or neutral results from animal models in clinical trials. If translational success was available as or could easily be converted to a percentage, this value was used. Otherwise, translational success was operationalized as 0% for no translation (no concordance between animal models and clinical trials), 50% for partial translation (partial concordance), and 100% for complete translation (full concordance) of results from animal studies to human studies. This approach allowed us to include both continuous and non-continuous definitions of animal-to-human translation, increasing our overall sample size. Furthermore, it enabled the comparison of high and low success rates, while taking partial concordance (50%) into account. Moreover, it prevented loss of information when a paper included multiple treatment strategies for the same condition or one treatment for multiple conditions. If a paper included multiple animal-human comparisons, each treatment and condition was scored separately. Because we used the paper as our unit of analysis, we averaged these scores per paper. While this practice of transforming a partially ordinal variable (no/partial/full translational concordance for narrative reviews) into a seemingly continuous one (percentage correspondence) may decrease relevant variation in translational success rates, it prevents issues with correlated data within papers and at the same time allows for inclusion of all literature data into a single analysis.

Our references were categorized into 4 categories for size according to the number of references they included. Categories were pragmatically defined based on variation within the sample.

Analysis
Explorative analyses were performed in R (V4.1.2 “Bird Hippie”) with the Rstudio interface9. Translational success rates were summarized by research field (neuroscience, pharmacology, and cancer research). Results were visualized with the ggplot2 package10. Colorblind friendly color palettes were used for all figures (Wong, 2011).

2.2 Clinical trials
No protocol was posted for this scoping review. An internal short protocol (in Dutch) with a crude description of the analyses, was shared between 3 of the co-authors before the start of screening.

Retrieval of clinical trials
As explained in the introduction, we restricted our clinical trial analyses to phase-2 clinical trials. We retrieved completed phase-2 clinical trials from the World Health Organization International Clinical Trials Registry Platform11. Only trials with available results were included in our analyses. We restricted the analyses to trials that were completed before 2020, as a large part of the later trials were related to the SARS-CoV-2 pandemic. The pandemic massively accelerated research, and this period was not considered comparable to prepandemic research. No further restrictions were applied. Retrieved trials were saved as a Microsoft Access database.

Trial selection
Trials with available results with an indication corresponding to an ICD-10 code were all included. Terminated trials were only included if they were terminated due to meeting the primary endpoint before the planned end date, lack of efficacy, or safety concerns. All other terminated trials (e.g., those that were terminated because of a lack of participants or because of commercial interests) were excluded from the analyses. Trials that were still actively analyzing data or recruiting participants were only included if results from planned interim analyses (i.e., prespecified in the trial protocol) were available.

Data extraction
Data extracted from the clinical trials are shown in Table 2. The ICD (International Classification of Disease) is a hierarchi-
The third script counts the number of terminated, blinded, randomized, controlled, and successful trials for each subgroup, as created with the second script (see Section B.4). Next, this script calculates the percentages of these trials with certain characteristics for each subgroup, as the first subscript did for the overall set.

Explorative analyses were performed to compare all results between the subgroups. These graphical analyses were performed in R with the ggplot2 package. Colorblind friendly color palettes were used for all figures (Wong, 2011).

### 2.3 Evidence integration
For this study we used a segregated concurrent approach, using two complementary methods to assess animal-to-human translational success rates in different fields of medical research. Both approaches (literature and clinical trials) were performed simultaneously. For viability of the scoping review, a selection of research fields had to be made to assess translational success rates. For clinical trials, all available trials were included to assess as many different research fields as possible. This difference results in incomplete overlap. The field of pharmacology, as defined for the scoping review, is not comparable to the clinical trials, because pharmacology includes many diseases. The fields neuroscience and cancer research are better defined and can be compared directly between the ICD-10 chapters and the scoping review.

### 3 Results

#### 3.1 Scoping review

##### 3.1.1 Selection of papers

Our search in PubMed identified 11,032 papers. Of these, 880 abstracts seemed to fit the scope of this study. Of these, 763 papers were excluded from the final analysis, after which 117 papers were included in this review. The flow of references is shown in Figure 1.

##### 3.1.2 Characteristics of included papers

All 117 included papers were written in English. Of these, 53 papers investigated treatments in the field of pharmacology, 21 in the field of cancer research, and 43 in the field of neuroscience. In total, 4 papers described meta-analyses, 30 described systematic reviews, and 83 papers were narrative reviews, equally distributed over the fields.

As shown in Figure 2, mouse and rat models were most prevalent. The “other” category includes model species that were only referenced once: goat, miniature pig, cow, cat, rabbit, pig, chicken, frog and fish. 24 included papers mentioned preclinical in vivo or animal research without the specific animal species. Overall, papers seemed to include more pre-clinical references compared to clinical references (Fig. 3).

---

**Tab. 2: Extracted data**

<table>
<thead>
<tr>
<th>Category of data</th>
<th>Extracted data per included paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial design</td>
<td>Randomization, blinding, use of control</td>
</tr>
<tr>
<td>ICD-10</td>
<td>Researched disease categorized by ICD-10</td>
</tr>
<tr>
<td>Outcome</td>
<td>Trial success</td>
</tr>
<tr>
<td>Termination</td>
<td>If applicable, reason for termination</td>
</tr>
</tbody>
</table>

---

The hierarchical structure of the ICD-10 to group trials into disease groups including at least 50 trials, further described below. Each trial was scored as successful (results encouraging further clinical studies) or unsuccessful. When trial results were available, the authors’ conclusion was followed. In case of termination due to lack of efficacy or safety concerns, the trial was deemed unsuccessful. If the trial was terminated due to meeting its primary endpoint, it was scored as a successful trial.

**Analysis**

We analyzed all extracted data using R (V4.1.2 “Bird Hippie”) with the R studio interface. All scripts used the dplyr package V1.0.7. Full scripts can be found in Appendix B of the supplementary file as well as on GitHub (v1.0). The full script consists of three subscripts, separately described below.

The first subscript was written to summarize the full dataset. It counts the overall included number of terminated, blinded (single-blinded, double-blinded, and triple-blinded or more), randomized, controlled, and successful trials. Next, absolute numbers were converted to percentages using a custom function (see Section B.2).

The second subscript was written to group the trials by ICD-10 codes, according to the hierarchical structure, into groups of at least 50 trials (see Section B.3). Grouping started with counting the occurrence of each code. If the count was below 50, the code was checked with the code above itself in an alphabetical list. If these two codes had the same first 3 characters, and both had less than 50 counts, they were grouped together. For example, E78.0 (Hypercholesterolaemia) did not reach the minimum size of 50 trials and was grouped together with 5 other codes as code E78 (Disorders of lipoprotein metabolism), creating a group of 53 trials and was grouped together with 5 other codes as code E78.0 (Hypercholesterolaemia) did not reach the minimum size of 50 trials, they were grouped together.

The second subscript was written to group the trials by ICD-10 codes, according to the hierarchical structure, into groups of at least 50 trials (see Section B.3). Grouping started with counting the occurrence of each code. If the count was below 50, the code was checked with the code above itself in an alphabetical list. If these two codes had the same first 3 characters, and both had less than 50 counts, they were grouped together. For example, E78.0 (Hypercholesterolaemia) did not reach the minimum size of 50 trials and was grouped together with 5 other codes as code E78 (Disorders of lipoprotein metabolism), creating a group of 53 trials and was grouped together with 5 other codes as code E78.0 (Hypercholesterolaemia) did not reach the minimum size of 50 trials, they were grouped together.

---

12 https://CRAN.R-project.org/package=dplyr

3.1.3 Overall results
Across all papers, 100% translation was the most common score (k = 64, Fig. 4). Overall, the average translational success rate was 68%. However, a substantial number of papers showed translational success of 50% (k = 24) or 0% (k = 25). For all fields combined, for the 34 systematic reviews and meta-analyses, 17 papers showed high translational success (75%-100%), and 17 papers showed low translational success (0%-50%). For narrative reviews, the percentage of success (100%) seemed to be higher (Fig. 4).

Furthermore, papers that included more references (> 100) on average seemed to have lower success rates, and papers with only a few references seemed to have higher success rates (< 25) (Fig. 5).

3.1.4 Comparing translational success rates
Figure 6 shows the literature-based translational success rates across the fields of pharmacology, neuroscience, and cancer research. These results show that translational success rates in neuroscience overall are not worse than those in other fields, contrary to what was previously published (Azkona and Sanchez-Pernaute, 2022; O’Collins et al., 2006; Davies et al., 2020). Overall, no clear difference can be seen between these fields.

3.2 Clinical trials
3.2.1 Retrieval of clinical trials
Our search of the WHO Clinical Trial Registry retrieved 13,985 phase-2 clinical trials. 3,102 records were excluded due to termination before conclusive results, and 549 records did not have publicly available results. The flow of records is shown in Figure 7.
Fig. 2: Number of times each animal species was reported in the included papers

Fig. 3: Size of the analyzed meta-research; numbers of pre-clinical (x-axis) and clinical (y-axis) references per analyzed paper

Fig. 4: Translational success
Histogram with the number of papers included in our review by translational success rate. Bars are split by color reflecting review type. Note that data are largely based on narrative reviews, which necessitated transforming a partially ordinal variable (no/partial/full translational concordance) into a continuous variable (percentage correspondence) for inclusion. Refer to Section 2 for further information.

Fig. 5: Translational success and study size (number of included references)
Bars are split by color reflecting study size. Note: only one paper included more than 100 references. Note that data are largely based on narrative reviews, which necessitated transforming a partially ordinal variable (no/partial/full translational concordance) into a continuous variable (percentage correspondence) for inclusion. Refer to Section 2 for further information.
3.2.2 Overall results
Overall, 65.2% of the included clinical trials showed positive outcomes, of which 0.1% were terminated because of reaching the primary endpoint earlier than planned. 34.8% of trials showed negative outcomes (not encouraging further clinical studies: either no efficacy or adverse events), of which 7.6% were terminated earlier than planned due to lack of efficacy, and 3.3% were terminated due to safety concerns.

Regarding trial design, overall, 51.1% of trials were randomized and 40.9% of trials used a control (either active or placebo). In total, 40.2% of trials were randomized controlled trials (RCTs). Lastly, 39.9% of trials used some form of blinding.

3.2.3 Comparing medical research fields
We first analyzed all included clinical trials by overall ICD-10 chapter. These chapters include different diseases. For example, the ICD-10 codes for the diseases gastroenteritis and sexually transmitted disease were grouped together in the ICD-10 chapter “Infectious and parasitic diseases” with various other diseases. These ICD-10 chapters showed relatively little variation in success rates; they ranged from 53.5% to 80.0% (Fig. 8).

Next, we created smaller subgroups where possible; 6,778 out of 10,334 clinical trials could be grouped into 51 disease groups.

---

**Fig. 6: Reported translational success rates by field**
Note that data are largely based on narrative reviews, which necessitated transforming a partially ordinal variable (no/partial/full translational concordance) into a continuous variable (percentage correspondence) for inclusion. Refer to Section 2 for further information.

**Fig. 7: Flow chart of selection of clinical trials**
Fig. 8: Percentage positive outcomes in phase-2 trials (y-axis) across broad research fields (x-axis).

Fig. 9: Percentage positive outcomes in phase-2 trials (y-axis) across narrow research fields (x-axis).
of ≥ 50 trials (Fig. 9). Compared to the ICD-10 chapters, in lower-level ICD-10 categories, success rates varied more; they ranged from 45.5% to 86%. The highest success rates (top 5) were found in disorders of lipoprotein metabolism (86%), epilepsy (85.0%), influenza (83.9%), lymphoma (83.8%), and viral hepatitis C (83.2%). The lowest success rates (bottom 5) were seen in schizophrenia (45.4%), pancreatic cancer (46.0%), bladder cancer (47.3%), and liver cancer (49.1%). These results suggest relevant variability in clinical trial success rates across all medical research fields, specifically within cancer research.

Within the top 5 research fields, randomization of trials varied between 8.0% and 81.8%, use of controls ranged between 1.9% and 69.9%, and any form of blinding was used in 1.2% to 71.7% of trials. Within the bottom 5 fields, the factors ranged between 17.6% and 85.5% for randomization, 11.8% and 72.5% for use of controls, and 7.9% and 74.2% for use of blinding.

3.3 Evidence integration

Comparing both methods (scoping literature review and clinical trials) for the research fields of neuroscience (the ICD-10 chapters “Mental and behavioral disorders” and “Nervous system”) and cancer research (the ICD-10 chapter “Cancer”) shows similar results. For neuroscience, the average animal-to-human translational success rate in the literature was 62%, and the clinical trial proxy showed 61% positive outcomes. For cancer research, literature showed an average translational success rate of 69%, while the clinical trial proxy showed 62% positive outcomes. While both methods have their limitations, and while we can only compare two fields, this concordance between methods indicates that the results of our study are reliable enough as a basis for further research.

4 Discussion

In this study we explored differences in translational success rates across medical research fields using two approaches. We defined translational success as the replication of positive, negative, or neutral results from animal studies in clinical trials. Most of the time, negative results in animal studies are not followed by clinical trials. However, we identified a few examples of clinical trials after negative or mixed preclinical results, mainly in literature reviews (e.g., Giles et al., 2017; Sedláková et al., 2017; Micheli et al., 2018). Corresponding translational success rates varied.

Our first approach, a scoping review, showed only minor differences in average translational success rates (Fig. 6); the success rates in the fields of neuroscience, pharmacology, and cancer research were not clearly different. Interestingly, meta-research following more stringent methods, i.e., systematic reviews and meta-analyses seemed to show a lower success rate than less formal narrative reviews (Fig. 4). This may be due to the selection and/or publication bias in narrative reviews. Furthermore, larger (> 100 references included) studies showed lower success rates compared to smaller studies (< 25 references included). Generally, more systematic meta-research includes fewer references (in this study all systematic reviews included < 50 references) due to the high workload of this type of research. However, our results suggest bias in the outcomes of the less systematic narrative reviews. This indicates a need for more large systematic reviews and meta-analyses to obtain a better view of the actual current state of translational success rates. The view described in current literature may be overly optimistic because most included reviews are narrative, and based on our results these narrative reviews show higher success rates. However, mean calculated percentages from literature hardly deviate from percentages in the clinical trial proxy, discussed below.

Our second approach, using outcomes of phase-2 clinical trials as a proxy, showed larger variations in success rates compared to our literature results, ranging from 45.5% to 86.0%. Broader fields of research had less variation than narrow fields, possibly reflecting a large variation of research practice within widely defined research fields. The largest variation in success rates was found within cancer research, having both some of the lowest and some of the highest success rates. It is unclear why cancer research shows this high variation, but one possible explanation could be differences in research methodology. Therefore, cancer research could be ideal to further investigate differences between smaller research fields to identify factors contributing to high or low translational success. Overall, our observation of higher variation between smaller fields than between wider fields highlights the need to look at research fields on the smallest scale feasible for studies addressing differences in research practice.

Combined, these results show relevant variability in translational success in medical research. The average success rates identified for neuroscience and cancer research are similar with both methods used, around 65-75%. In contrast to the general opinion expressed by neuroscientists (e.g., O’Collins et al., 2006), this field as a whole does not show lower translation than other fields.

Our methods have several limitations. First, the ordinal categorization of translational outcomes as successful, partially successful, or not successful is more prone to bias compared to continuous outcomes. Besides, the categorization of outcomes was based on data expressed in the included papers and trial records only and performed by a single researcher. However, by scoring outcomes ordinally, partial correspondence between animal and human outcomes cannot be properly assessed. For the literature sample, the majority of included reviews was narrative, and there was not enough literature quantitatively discussing translational success rates to perform a fully continuous analysis per research field. For the clinical trials, it was not feasible to include pre-clinical animal data due to the high number of clinical trials assessed in this study. Future studies should investigate the pre-clinical and clinical data in some of the highest scoring and lowest scoring fields in more detail to validate our findings. However, this may be challenging as preclinical data is often not available in the public domain.
study should be interpreted as a first exploration of translational success in different research fields within the boundaries of currently available data.

Second, to ensure a decent sample size, the inclusion and exclusion criteria for our scoping review were broad. This increased the number of discrepancies in the full-text screening phase, as the criteria were open to interpretation. While the discrepancies were all eventually resolved by discussion between the screeners, this may limit the replicability of our screening. For future reviews, this might be prevented by using stricter inclusion and exclusion criteria. To maintain a sufficient sample size, some criteria should then be altered. For example, a larger range of publication dates can be included.

Third, another limitation of including narrative reviews is that these papers do not describe their methods and, therefore, their external validity cannot be properly evaluated. However, as there are still few systematic reviews, narrative reviews had to be included to reach an appropriate sample size.

Fourth, for our analysis of the clinical trials we used trial success as a proxy for translational success. We selected phase-2 trials because these trials investigate efficacy and safety of treatments in small groups of patients and are therefore more comparable to pre-clinical trials than clinical trials in other phases. To use clinical trial success as a proxy for successful translation, we assumed that each clinical trial was based on successful preclinical research. Interpretation of our clinical trial findings should take into account that this assumption is not always valid. Clinical trials may also be started without preceding animal research (e.g., for new indications for compounds already on the market). Besides, the outcome of a clinical trial depends not only on safety and efficacy of the tested intervention but also on factors such as appropriate trial design and participant treatment adherence. While we could not think of reasons to assume differences in these factors between fields, the results must be interpreted cautiously.

Last, for our analysis of the clinical trials, fields were categorized by ICD-10 code with a limit of at least 50 trials per group. Consequently, some fields of research that are quite different were grouped together. For example, gastroenteritis and sexually transmitted diseases were grouped together under “Infectious and parasitic diseases”, and inguinal hernia and peritonitis under “Diseases of the digestive system”. Most of the groups comprising quite different diseases had average trial success rates around 60-70%, which is similar to the overall average success rate of 65%. This indicates the importance of investigating research fields separately, focusing on fields where sufficient data are available.

Because of these limitations, future studies should investigate the pre-clinical and clinical data in some of the highest and lowest scoring fields in more detail to validate our findings. However, this may be challenging as preclinical data is often not available in the public domain. Therefore, this study should be interpreted as a first exploration of translational success in different research fields within the boundaries of currently available data.

Despite these limitations we believe this study carries value. To our knowledge, it is the first study formally comparing translational success rates between medical research fields. Fields were compared using two quantitative approaches, with concordant results. The resulting data allows us to make recommendations for future meta-research, as described below. Besides, it shows which fields of medical research are most in need of improving translational success, i.e., schizophrenia, pancreatic cancer, bladder cancer, colon cancer, and liver cancer. Efforts to improve animal models, or to replace them by more informative alternative approaches, should focus on these fields of research first to improve translational value and reduce unnecessary clinical and preclinical research.

We performed these studies because of our interest in improving animal to human translation. Investigating common practices in fields with high and low translational success rates might identify, e.g., experimental designs that can lead to either translational success or failure. Due to insufficient reporting of study design, it is still not feasible to assess the relationship between design and translational success rates from literature directly. Therefore, we propose this indirect approach. In this work, we identified several research fields with relatively low and high success rates. The fields of schizophrenia, bladder cancer, and colon cancer could serve as case studies for fields with translational failure as these are fairly large research fields (> 75 clinical trials included in this research) with relatively low success rates. For translational success, we would suggest analyzing sleep disorders, chronic lymphocytic B-cell leukemia, and diabetes mellitus type 2, which are of fair size (> 75 clinical trials included in this research) and have high success rates. While this study only analyzed a few experimental design parameters in human studies, the results indicate sufficient variation in practice between fields to make the indirect approach viable.

To conclude, this study provides useful insight into translational success rates across medical research fields. It may serve as a basis for future research into factors that influence translational success rates and can serve as a basis to prioritize efforts in replacing and reducing animal research.

References


**Conflict of interest**

The authors of this paper have no conflict of interest to declare.

**Data availability**

All presented data is available through OSF (doi:10.17605/OSF.IO/EZAN6) as well as GitHub\(^2\).

**Acknowledgements**

This work was funded by the Federal State of Lower Saxony (R2N) and the Radboud University Medical Center.