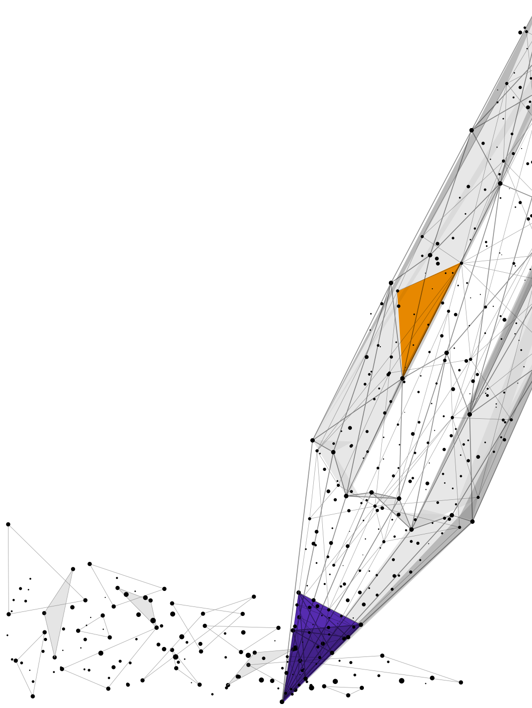
Avoidable Experiment Expenditure

Examining a major issue affecting life science productivity

BenchSci

Contents



Executive summary

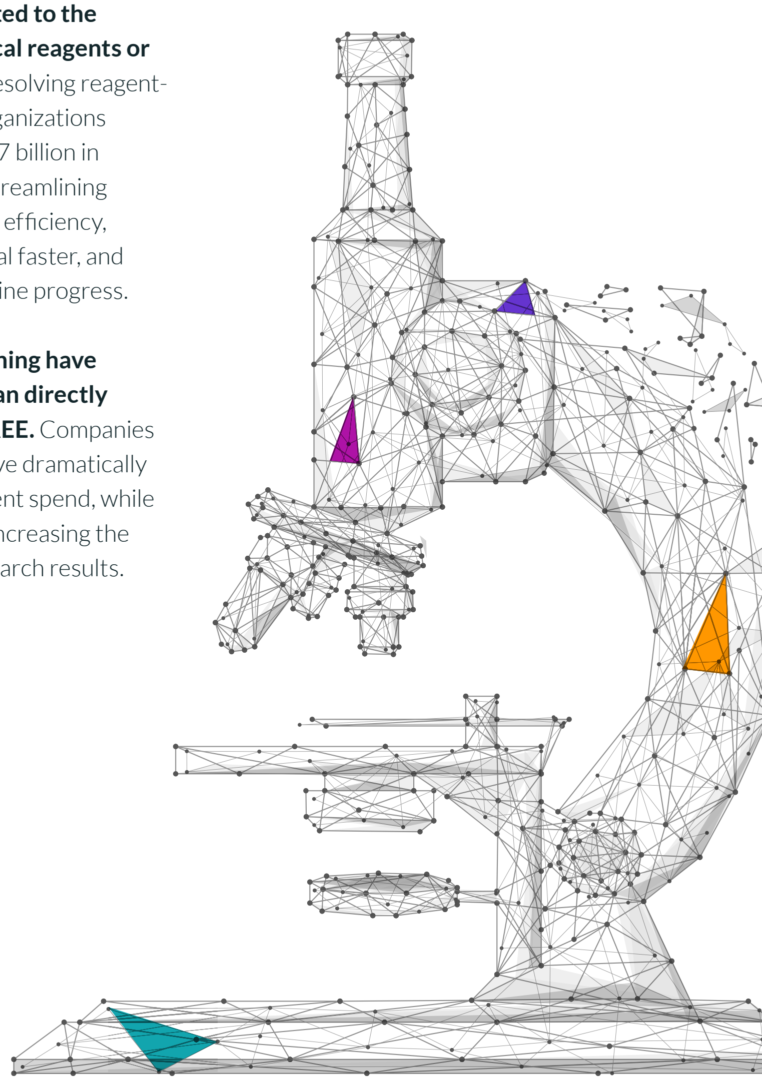
The pharmaceutical industry is facing an increasing number of productivity challenges when bringing a new drug to market. Over the last decade, pharmaceutical R&D costs have been steadily increasing, while the rate at which drugs are approved has remained roughly constant. As of 2018, it cost approximately $2.17 billion (all monetary figures are in USD) to bring a new drug to market, almost double the $1.18 billion per-asset costs reported in 2010.[1] In the same time span, annual forecast peak sales were cut in half, from $816 million in 2010 to $407 million in 2018.[1] This has resulted in a decrease in returns of 8.2 percentage points, from 10.1% in 2010, to 1.9% in 2018.[1]

Drug development isn't only costing more, it's also taking longer. The average project length increased from 9.7 years in the 1990s to 10-15 years through the 2000s and 2010s.[2,3] Factors commonly attributed to these trends include clinical trial inefficiencies, such as patient enrollment competition and more stringent selection criteria, and the industry's shift into more difficult, high-risk, high-reward research areas like oncology.

Our data suggests that Avoidable Experiment Expenditure (AEE) is an overlooked source of unnecessary spend and effort, significantly contributing to the increase in project length and fall in returns. Considering that ~42.9% of overall spend on drug development goes toward preclinical R&D, this issue deserves more attention.[4] Avoidable Experiment Expenditure refers to all inefficiencies and productivity challenges in designing and carrying out preclinical experiments. Experiments are the foundation of preclinical research and development, however, irreproducibility rates in preclinical experiments exceed 50%, costing the industry nearly $48 billion annually.[5]

Over a third of AEE, or more than $17 billion, can be attributed to the ineffectiveness of biological reagents or reference materials.[5] By resolving reagent-related AEE, life science organizations could potentially recoup $17 billion in unnecessary spend while streamlining operations, improving R&D efficiency, bringing drugs to clinical trial faster, and generally accelerating pipeline progress.

Advances in machine learning have yielded technology that can directly address reagent-related AEE. Companies that use this technology have dramatically reduced unnecessary reagent spend, while accelerating research and increasing the reproducibility of their research results.



Introduction

In recent years, pharmaceutical organizations have been confronted with a number of challenges that have greatly impacted their productivity and ROI. Patent expiration has caused a decline in branded drug sales, as consumers favor less expensive generic drugs. Increasing competition has caused many organizations to shift their focus to more high-risk, high-reward projects, with the hope that operating in a more specialized market will increase their share of sales and offset the extra expenses from a higher rate of unproductive experiments. More stringent regulatory requirements have also had a hand in slowing drug development. Some of these challenges are unavoidable; patent expiration ensures drug prices are kept in check, competition is fundamental to our economies, and regulations protect both companies and consumers from harmful practices. There are, however, issues which can be addressed.

Despite their importance,

up to

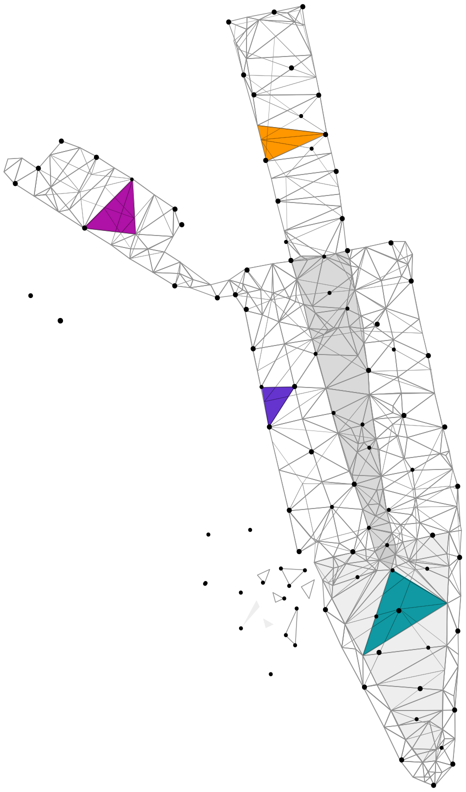
50%

of experiments in preclinical R&D are unproductive [5]

One such issue, which has attracted increasing attention, is Avoidable Experiment Expenditure, or AEE, which affects every organization in the pharmaceutical industry. Avoidable Experiment Expenditure refers to inefficiencies and productivity challenges that scientists and pharmaceutical organizations encounter during the preclinical phase. Experiments are vital to preclinical R&D; they are the method by which assets are advanced to clinical trials, thus scientific integrity is key. Yet despite their importance, up to 50% of experiments in preclinical R&D are unproductive (i.e. unable to advance assets).[5]

There are multiple factors that contribute to AEE, and are inherent challenges of any drug discovery effort. These include inappropriate reagents, poor experimental design, unreliable or variable protocols, lack of details in reporting, and lack of transparency in internal data.[5] This paper will focus on the impact of inappropriate biological reagents on AEE.

INTRODUCTION



Commonly used biological reagents, including antibodies, recombinant proteins, RNAi, and CRISPR, as well as model systems like cell lines, are sourced from biological origins, such as animals and patient samples. As a result, the performance of biological reagents can vary depending on the biological systems in which they are both created and used. Reagent selection is challenging—experimental context must be taken into account, data for which is scattered throughout many sources. This is exacerbated by unreliability in reagent data, as well as in reagents themselves; it is estimated that up to 50% of reagents do not work as intended.[6]

In fact, over 36% of all unproductive experiments in preclinical R&D can be attributed to inappropriate reagents.[5] Current standard processes for selecting reagents are flawed and unreliable at best, as we will discuss further. However, the general perception is that the exploratory nature of preclinical R&D makes it inherently inefficient, thus the selection and use of inappropriate reagents is deemed an unavoidable and necessary part of the scientific process.

> *Inefficiencies in preclinical experiments have a direct and measurable impact on an organization's productivity.*

The inefficiencies in preclinical experiments have a direct and measurable impact on an organization's productivity, resulting in longer research times, more spend on drug development, and increased downstream risk. We will elaborate on each of these pain points in the following sections, based on our work with over 40,000 scientists in more than 4,300 institutions, including 15 of the top 20 pharmaceutical companies.

We assert these inefficiencies are avoidable. With more effective experimental protocols, access to better information, improved reagent selection, and more informed experimental design, issues such as unproductive experiment rates and proportions of material waste could be reduced. We are already seeing signs that these concepts are being applied and starting to reverse the decline in productivity.[7] By extending their application throughout the preclinical research phase, beginning with reagent selection, organizations could further improve productivity, reduce spend, increase the success rate of preclinical experiments, and decrease downstream risk. At the end of this paper we propose a solution in the use of data aggregation and machine learning to provide scientists with productivity-enhancing software that increases their efficiency.

Challenges of selecting antibodies

Amongst the various biological reagents for preclinical experiments, antibodies are arguably the most commonly used. However, finding the right antibody is far from an easy task.

For a given protein, for example, PD-L1, there are 67 vendors that sell a total of over 2,300 antibodies with distinct product names. In a typical manual search process, scientists need to visit numerous vendor websites just to find these products. At the same time, each vendor only provides their in-house validation data, which are often limited to a single experimental condition, thus limiting the applicability of such information and making product comparisons challenging.

Making things worse is the fact that many vendors would purchase products from an OEM (original equipment manufacturer) and relabel those products as their own, effectively polluting the total product pool with ones that are, in fact, identical to each other but have different product names. Our data suggests 20-30% of products are relabeled. In addition, conventional search engines do not take protein synonyms into account. As a result, searching for PD-L1 will not show products for CD274, even though both names refer to the same protein. Together these factors add additional barriers between scientists and the potential antibody products they need.

Research takes longer

Time is an important factor to take into account when considering avenues to improve pharmaceutical R&D. The faster scientists can get the results to advance their research, the sooner new drugs can enter the market to help patients. Of course, quicker results cannot be prioritized over scientifically valid ones, so a way must be found to expedite results without compromising safety and efficacy.

Unlike in clinical phases, there is no FDA equivalent that regulates the standards around how biological reagents are developed or used. There is an abundance of reagent data available online through databases and vendors, however, it is generally disorganized and often incomplete, unvalidated, inaccurate, or biased.[5, 6, 8] In addition, valuable information may be overlooked due to being buried in images which don't register on a web browser search. Vendors confuse matters even further by purchasing reagents from each other and reselling them under their own product name.[6]

12

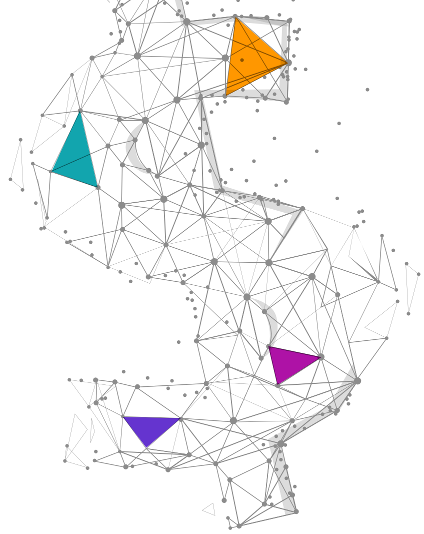
Sorting through this data manually can be frustrating, and occupy days of scientists' time, with no guarantee they're seeing all relevant information due to search engine technology limitations. Because of the high levels of uncertainty in the data, as well as the fact that up to 50% of reagents simply don't perform the function they are intended to, most researchers have adopted a "shotgun" approach, wherein they will purchase multiple different reagents for a singular purpose in order to increase the chances of success.[6] This leads to more time spent validating reagents, which incurs additional material costs. The entire process can take weeks to months. Even then there is no guarantee of productive results, and unreliable or misidentified reagents can result in irreproducible experiments and wasted years of hard work. Suboptimal reagents—those that work but aren't ideal—are also problematic. Antibodies that produce a weak signal, for example, have been known to cause research projects to get stuck for a year or more.

One way to avoid this is to improve the organization and accuracy of the information scientists use to design their experiments. There have been efforts to establish standards for reporting biomedical data, such as the NIH's policy mandating the authentication of biological reagents.[9] However, in order to be effective, standards must be agreed upon and followed by the entire international life science community.

There are some tools, such as PubMed and Google Scholar, that can assist scientists in locating the information they need for reagent selection and experimental design. However, these aren't much more than specialized web browsers for scientific literature. They are still reliant on keyword searches, which may not understand synonymous terms for reagents, protein targets, connections via relevant biology, and disease/phenotype descriptors. Moreover, supporting evidence provided, including images of validation assays, are frequently unlabeled and sometimes misappropriated. This is where we see an opportunity to apply machine learning to decode and present data from scientific literature in a more sophisticated, more relevant way. Based on our research and direct feedback from scientists, sufficient improvements to the searchability, accuracy, and completeness of available reagent data could accelerate project timelines by months.

Resources associated with antibody validation

In addition to the antibodies themselves, many resources are required to validate antibodies. For example, to perform genetic validation, which involves testing antibodies using gene knockout or knockdown, tissues would have to be first extracted from genetically modified animals, or RNA interference experiments need to first be performed in cell lines. The gene knockout or knockdown would then have to be verified through sequencing techniques. Together with the time it takes to grow the animals and cells, it can take up to weeks of scientists' time just to prepare the samples for a single antibody validation experiment.



Research costs more

Avoidable Experiment Expenditure (AEE), and the reagent selection process in particular, directly impact an organization's spend in multiple ways, and each instance compounds the others. For example, longer research times due to manual reagent selection processes mean more spend on scientist salary per project. Unreliable reagent data forces scientists to use the "shotgun" approach, purchasing more reagents than necessary and increasing materials spend. Then, more time must be spent testing and validating those unnecessary reagents, further increasing spend on salary, as well as on the additional materials required to run these tests.

Pharmaceutical organizations spend up to $35 million per asset on biological reagent material costs.[9] Our data shows that, as a consequence of AEE, a significant percentage of that $35 million is spent on unproductive experiments. For example:

15-20%, or
$5.25 to $7 million
is spent on single-purchase reagents (these are reagents that scientists only purchase once, due to finding them inferior for experiments and not worth purchasing a second time)[10]

20-30%, or
$7 to $10.5 million
is spent on redundant testing (which includes repeated validation experiments for the same reagents, due to factors such as them being unknowingly relabeled and lack of communication between research groups)[10]

Up to 25%
of custom reagent orders
are not necessary, as they have commercial equivalents that scientists did not find nor test[10]

The financial impact does not end with the reagents. Reagents may be a cause of failure but are only a subcomponent of an entire experiment's material and costs. From our industry measures, every dollar of reagent spend correlates to $1.14 of lost hard cost expenditure attributed to reagent failure.[10] This additional cost includes media, cell lines, tissues, and glassware, and excludes operational expenditure such as equipment time. Therefore for every $10 million in reagent spend, potentially $11.4 million is also lost in materials accompanying a failed reagent.

However, the greatest unaccounted cost is the human capital expenditure. This includes the time researching, designing, executing, and analyzing experiments that failed to yield a productive result due to an inappropriate reagent.

These issues most impact researchers in therapeutic areas for which finding reagents is a particular challenge, including oncology, immunology, neurology, and rare diseases.

Downstream risk is increased

The average rate of new drug applications submitted to the FDA has been increasing over the last five years, but clinical success rates are decreasing. Between 1983 and 1994 the overall clinical success rate was at 21.5%. This decreased to 11.8% between 1995 and 2007, and decreased further to 9.6% between 2006 and 2015.[4] As of 2020, it has been reported that 95% of drugs that enter clinical trials are unsuccessful.[11]

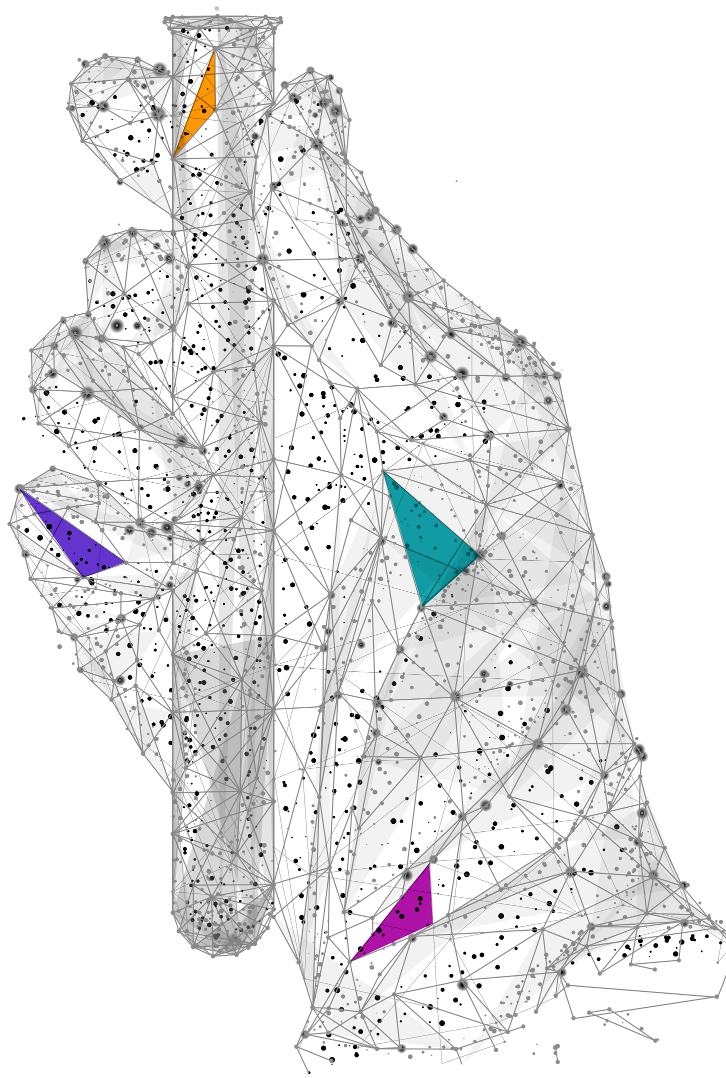
This could be due, in part, to more stringent regulatory requirements imposed by the FDA. Other contributors may include methodological flaws and poor experimental design in preclinical in vitro and in vivo animal studies.

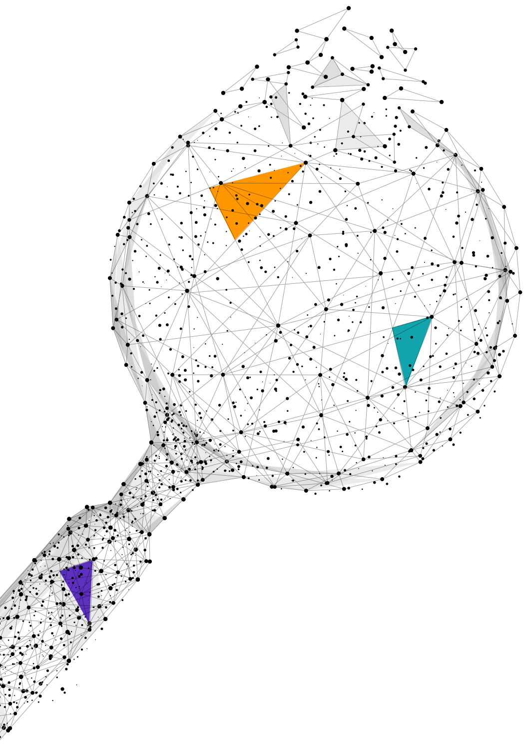
It is, however, a simple fact that clinical trials are built upon preclinical research. As an example, biomarkers identified and validated during the preclinical stage are used as primary endpoints to measure the disease states in clinical studies. Poorly characterized biomarkers can have lasting detrimental effects into the later stages of drug development. Better reagents lead to better understanding of biology, including better biomarkers.

It's unclear what percentage of unsuccessful clinical trials are due to unreliable or irreproducible preclinical results, as this is difficult to measure, but based on the high rate of attrition in clinical stages we can infer that it is not insignificant.[12] It is a logical conclusion, then, that by ensuring preclinical experiments are designed properly, using the appropriate reagents, one would see a higher success rate once targets proceed to clinical trial. In an environment where only 5% of drugs that enter clinical trials are successful, this type of assurance is incredibly valuable.[11]

95%

of drugs that enter clinical trials are unsuccessful[11]





Measuring impacts and finding solutions

As we've established, Avoidable Experiment Expenditure (AEE) has a significant impact on the scientific progress and ROI of pharmaceutical organizations. The first step to solving a problem is finding a way to quantify it. As of now, very little is being done to track productivity in preclinical experiments. We suggest it would be wise for the industry to align on standards to measure this, as well as the impacts of AEE, beginning with reagent-related AEE. Once we know the scope and scale of the problem, we can measure the impact of solutions that address it by helping scientists make more informed decisions.

The challenge, however, lies in data curation. With so many disparate sources of information strewn across the internet, it's impossible to sort through and organize it all manually. This is where machine learning becomes invaluable. If trained to analyze the data as a human researcher would, an AI can decode data from the world's scientific literature in a more sophisticated, more relevant way, and present it in an intuitive, user-friendly interface. It can do this in a small fraction of the time it would take a human scientist to do.

With the backing of Gradient Ventures, Google's AI fund, **BenchSci** has been exploring this potential for the past three years. Using proprietary machine learning models trained by PhD biologists, BenchSci has decoded over 10.8 million scientific articles and 18.1 million experimental data points, as well as data for over 28.5 million products from more than 261 vendors, to provide scientists with the best data available for selecting reagents. This information facilitates better, more informed experimental design, negating the need for a "shotgun" approach to reagent selection, in which only a third, or less, of purchased reagents perform as intended. The goal, simply put, is this: perform fewer experiments and get better results, faster.

As an ancillary benefit, organizing the world's reagent data allows for the extension of this technology to procurement data, thus providing a clearer picture of reagent purchase cost-effectiveness within an organization. Without this technology, organizations cannot accurately identify reagents within hundreds of thousands of rows of purchasing data, nor connect these reagents to targets and projects and determine their effectiveness. Organizing the world's reagent data with AI has enabled this capability, allowing customers of BenchSci to quantify its impact. Companies that use BenchSci's AI-Assisted Reagent Selection platform see significant reductions in reagent-related AEE, resulting in improved productivity, which translates directly to increased ROI.

Conclusions

The productivity challenges faced in the pharmaceutical industry have a significant impact on project timelines, ROI, and scientific progress in general. Although there is no single cause for these challenges, Avoidable Experiment Expenditure (AEE) is an overlooked contributor where there are opportunities to make a significant impact.

One such opportunity lies in improving the reagent selection process. By providing scientists with more accurate, better organized reagent data to improve their decisions, it is possible to reduce the time needed to search for and validate reagents by weeks to months, while also reducing irreproducible and unproductive experiments. This allows organizations to reduce unnecessary spend, both on hard materials and scientist capacity, which can be reinvested into exploring additional targets and assets. As an added benefit, more reliable preclinical results improve productivity and reduce downstream risk.

The sheer amount of reagent data available makes it impossible to effectively curate manually. However, by utilizing sophisticated machine learning models, it is possible to teach an artificial intelligence to do the heavy lifting for us. A properly trained AI can process scientific literature at a much quicker rate than a human scientist, and present that information in an intuitive, relevant way. BenchSci is the only company in the world currently pursuing this avenue.

For more information about BenchSci and our mission to exponentially increase the speed and quality of life-saving research, please **visit our website** or **contact us**.

hello@benchsci.com

www.benchsci.com

/benchsci

@BenchSci

References

1. *Ten years on, Measuring the return from pharmaceutical innovation 2019*. Deloitte. 2019. **https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/life-sciences-health-care/deloitte-uk-ten-years-on-measuring-return-on-pharma-innovation-report-2019.pdf**

2. DiMasi, Joseph A., Henry G. Grabowski, and Ronald W. Hansen. *Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs*. Journal of Health Economics. North-Holland, February 12, 2016. **https://www.sciencedirect.com/science/article/abs/pii/S0167629616000291?via%3Dihub**

3. Pammolli, Fabio, Laura Magazzini, and Massimo Riccaboni. *The Productivity Crisis in Pharmaceutical R&D*. Nature News. Nature Publishing Group, June 1, 2011. **https://www.nature.com/articles/nrd3405**

4. Norman, Gail A. Van. *Drugs, Devices, and the FDA: Part 1: An Overview of Approval Processes for Drugs*. JACC: Basic to Translational Science. Elsevier, April 25, 2016. **https://www.sciencedirect.com/science/article/pii/S2452302X1600036X#bib11**

5. Freedman, Leonard P., Iain M. Cockburn, and Timothy S. Simcoe. *The Economics of Reproducibility in Preclinical Research*. PLOS Biology 13, no. 6 (September 2015). **https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002165**

   *Freedman et al. report figures for the US only. We have extrapolated their data to reflect the fact that the US accounts for 58% of global pharmaceutical R&D. See: **https://www.abpi.org.uk/facts-and-figures/science-and-innovation/worldwide-pharmaceutical-company-rd-expenditure-by-country/**

6. Baker, Monya. *Reproducibility Crisis: Blame It on the Antibodies*. Nature News. Nature Publishing Group. May 19, 2015. **https://www.nature.com/news/reproducibility-crisis-blame-it-on-the-antibodies-1.17586**

7. Ringel, Michael S., Jack W. Scannell, Mathias Baedeker, and Ulrik Schulze. *Breaking Eroom's Law*. Nature News. Nature Publishing Group, April 16, 2020. **https://www.nature.com/articles/d41573-020-00059-3**

8. *Artificial intelligence turns to antibody selection*. Nature News. Nature Publishing Group. Accessed May 12, 2020. **https://www.nature.com/articles/d42473-019-00331-0**

9. Lauer, Mike. *Open Mike*. National Institutes of Health. U.S. Department of Health and Human Services, January 29, 2016. **https://nexus.od.nih.gov/all/2016/01/29/authentication-of-key-biological-andor-chemical-resources-in-nih-grant-applications/**

10. BenchSci internal data and research

11. Begley, C. Glenn, and Lee M. Ellis. *Raise Standards for Preclinical Cancer Research*. Nature News. Nature Publishing Group, March 28, 2012. **https://www.nature.com/articles/483531a?linkId=33568136**

12. Seyhan, Attila A. *Lost in Translation: the Valley of Death across Preclinical and Clinical Divide – Identification of Problems and Overcoming Obstacles*. Translational Medicine Communications 4, no. 1 (2019). **https://transmedcomms.biomedcentral.com/articles/10.1186/s41231-019-0050-7**