

Data Science: Bridging the gap between controlled experiment and real-world data?

CONTENTS

Introduction _____	2
Background _____	3
Survey Highlights _____	5
Survey Insights _____	16
The Future _____	27

INTRODUCTION

Could data science be about to revolutionize the regulatory approval of new drugs?

The biopharmaceutical and healthcare industries now collect more data than ever before due to advances in the variety of information sources combined with the ability to store vast quantities of diverse data. Sophisticated machine learning (ML) and artificial intelligence (AI) techniques allow us to access and analyze any combination of a multitude of data sources. The way that traditional controlled sources are viewed is being adapted in light of new evidence that emerges from real-world data. A recent Deloitte survey¹ found that 90 percent of biopharma companies are making significant investments in real-world evidence capabilities to drive drug development and meet regulatory requirements. Real-world evidence (RWE) has historically been used for post-marketing endorsement and in pricing and reimbursement negotiations. But could data science offer an opportunity to fundamentally shift this paradigm, leading to better and more affordable medications being approved on the basis of RWE?

Cytel created and ran a survey asking respondents from our audience about the potential of data science approaches in the sector. The survey was designed as a qualitative pulse check, rather than to claim any statistically significant findings. Many of the responses were obtained from delegates at conferences our team attended in June 2018, and we acknowledge that our pool of respondents at the PSI (Statisticians in the Pharmaceutical Industry) 2018 and PhUSE US Connect 2018 events will likely have interest in certain methods and topics.

The respondents were asked for demographic information (company type and job function) then asked to respond to specific questions about:

-  **The main goals and current sources for data science approaches within their organizations**
-  **The key opportunities that data science presents for the industry**
-  **The key barriers and challenges faced by the industry to maximize the potential for healthcare**

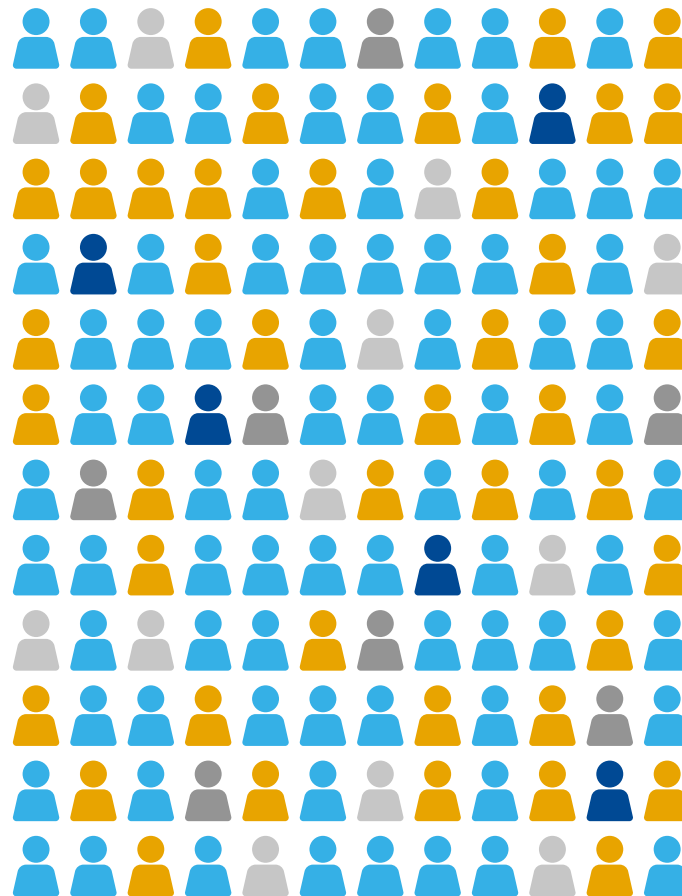
Respondents were also asked about how data science approaches on real-world data might impact the regulatory approval of new drugs and which therapeutic areas would most likely benefit. We believe we have gained a useful insight from the opinions of respondents of the survey, which reveal how the evolution of data science could lead to a revolution of real-world data being accepted as robust evidence for regulatory approval of new drugs.

Survey Respondents

144 respondents

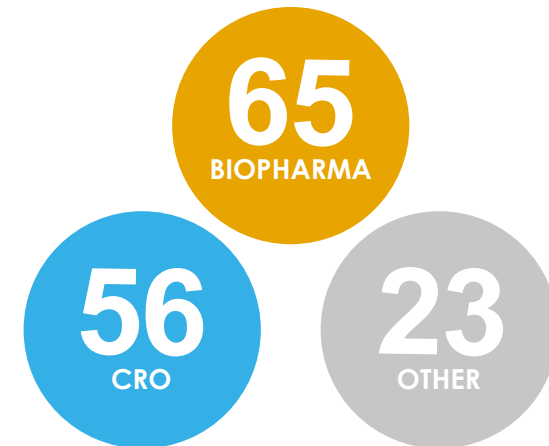
Split by function

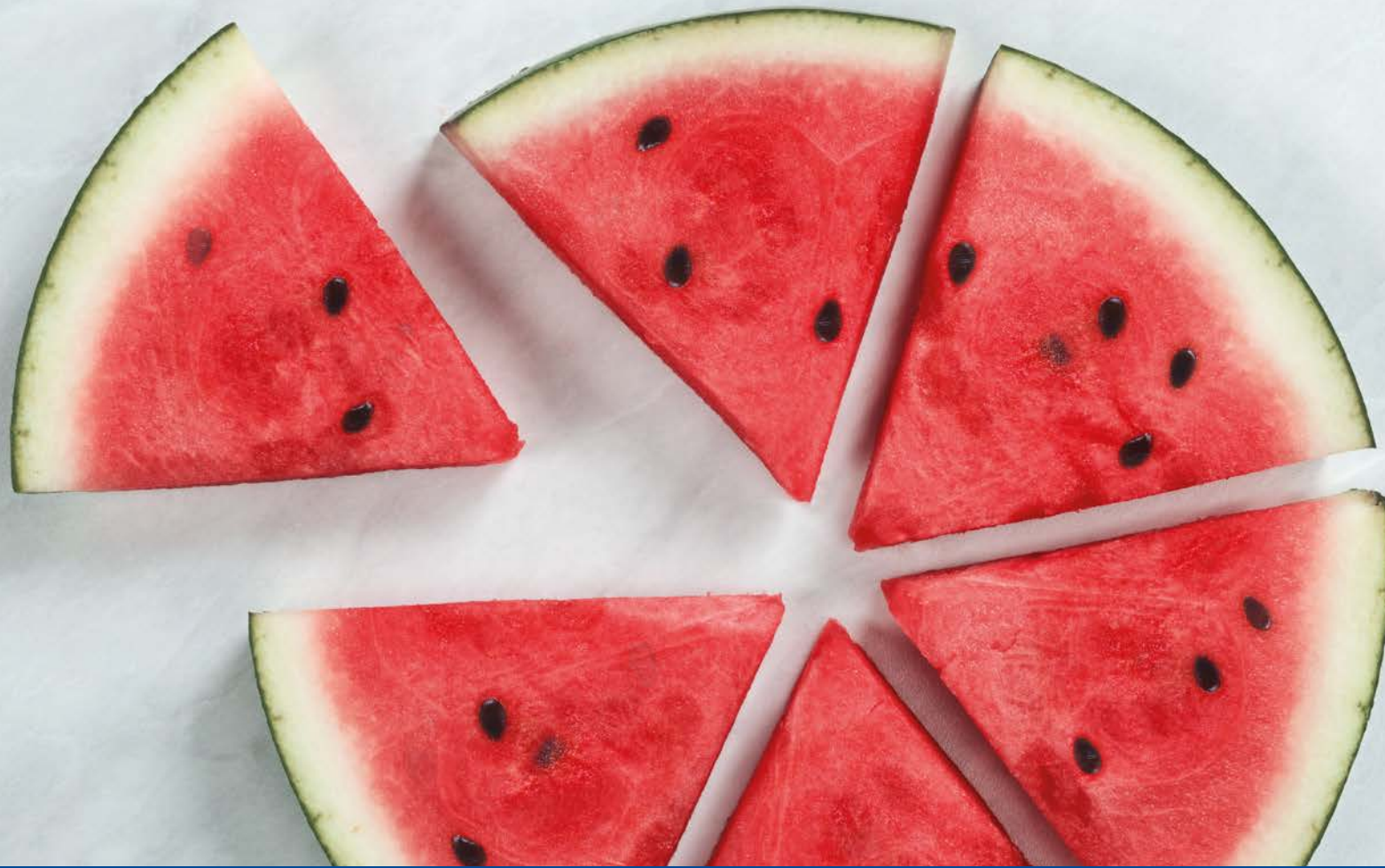
- 78** Biostatistics
- 42** Programming
- 7** Data Management
- 5** Data Science
- 12** Other



144 respondents

Split by organizations





SURVEY HIGHLIGHTS

Less than 1 in 7 of all respondents suggested a definition of data science

“An ‘art’ which borrows from statistics and machine learning tools to extract reliable knowledge from a data source”

“Getting meaningful information from the pool of data”

“To extract the facts and truth hidden inside the large data which will give solution to today's medical problems”

“An efficient way to handle the large amount of scientific data”

“Data science is using predictive indicators to drive patient safety and better outcomes”

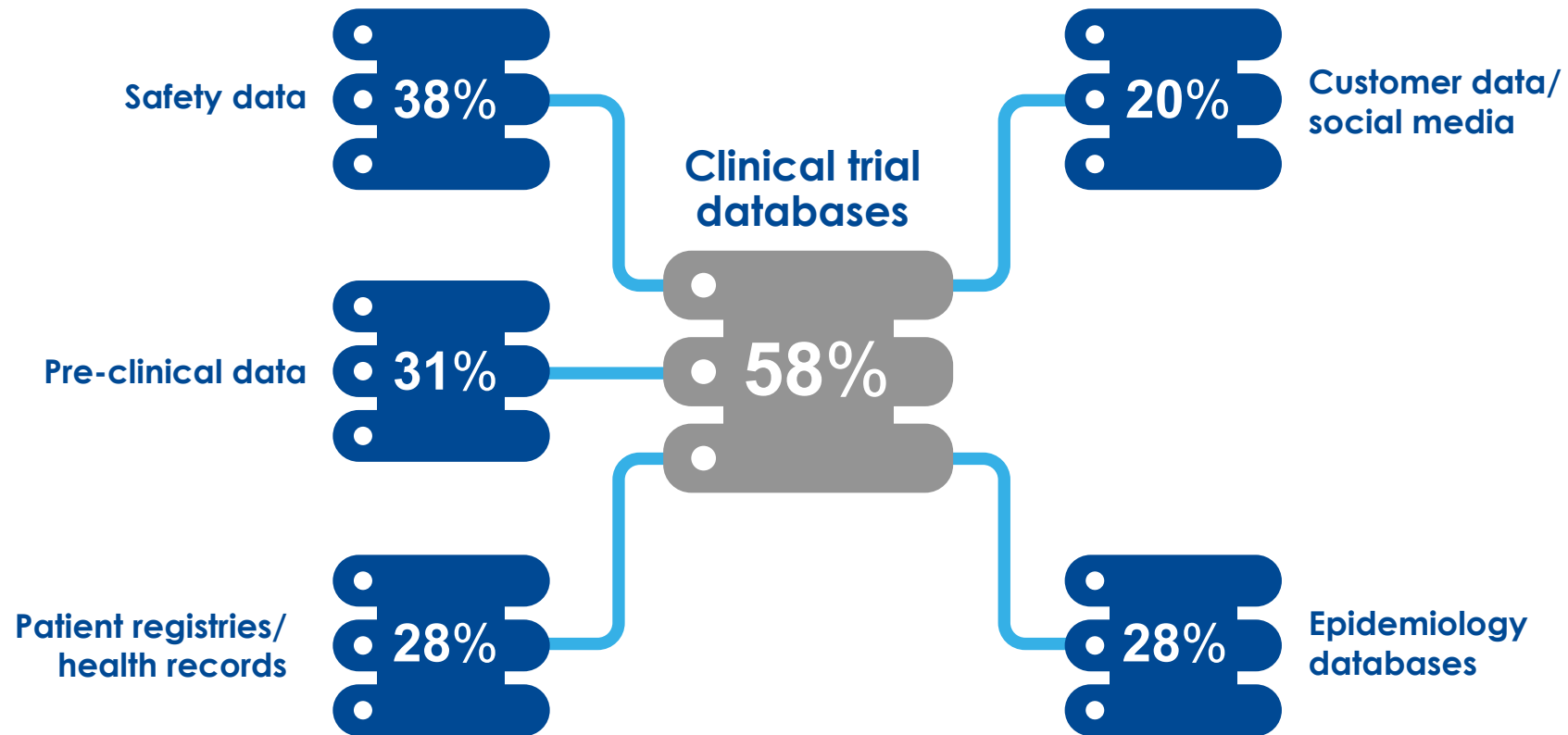
“Child of Machine Learning and Statistics”



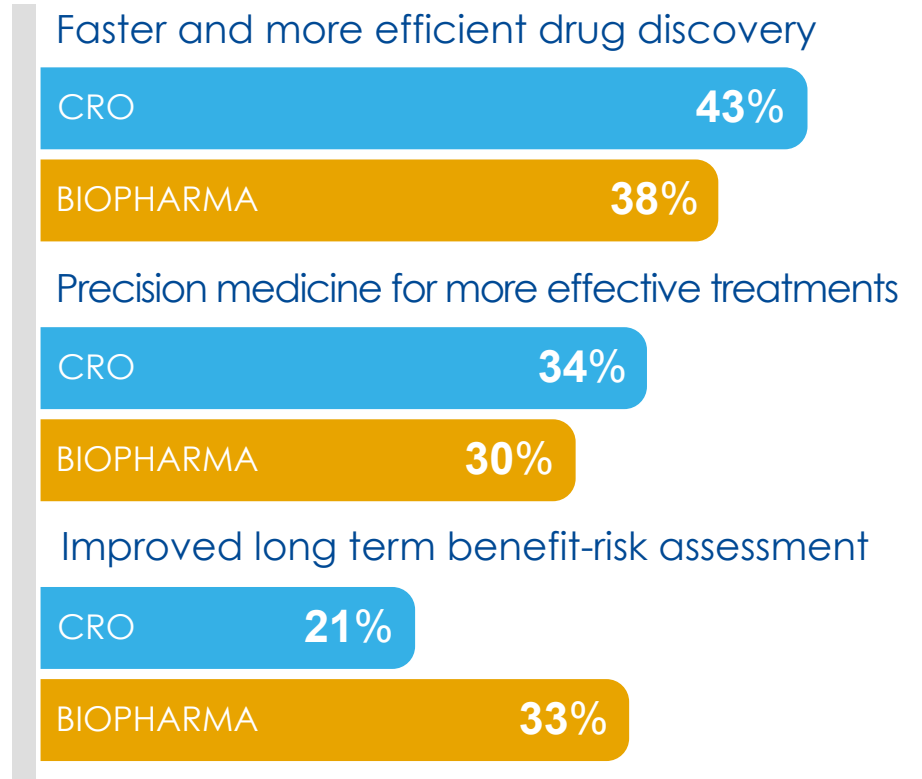
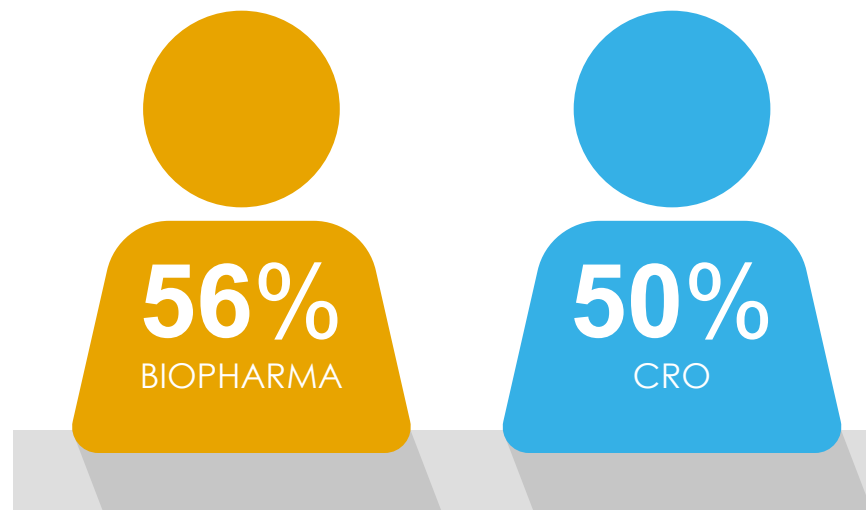


Percentage of respondents with a data science department

Data science is increasingly being used across a broader range of contemporary data sources

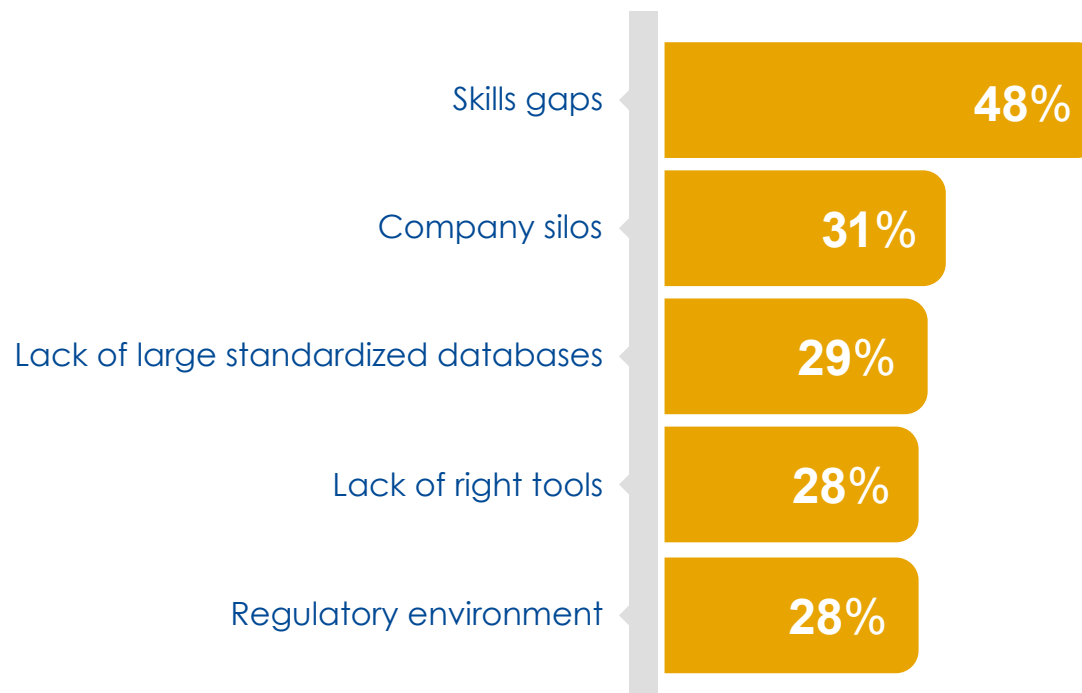


The majority of respondents agree that improved clinical trial design is a goal that will benefit most from data science



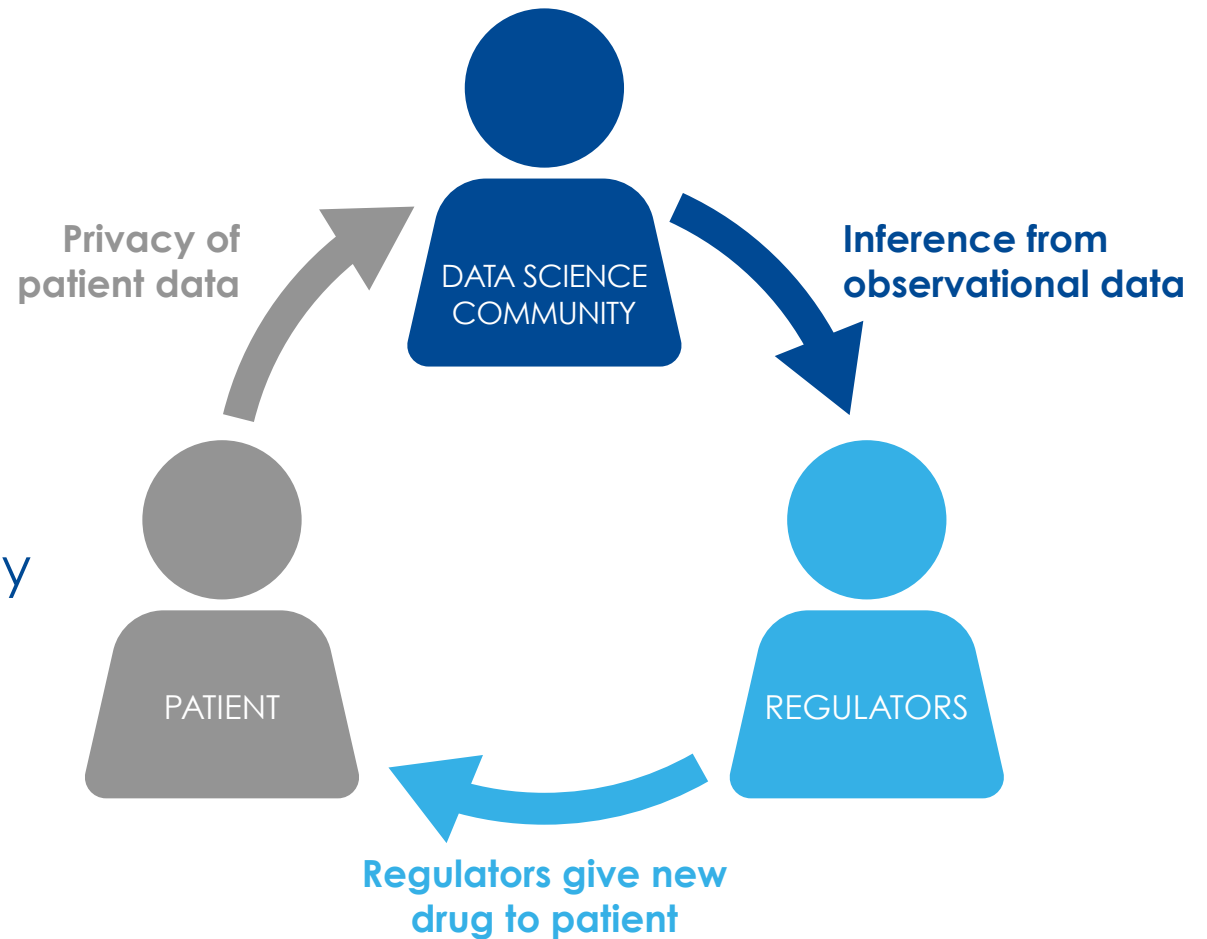
Percentages represent number of responders who selected that option – multiple responses were allowed

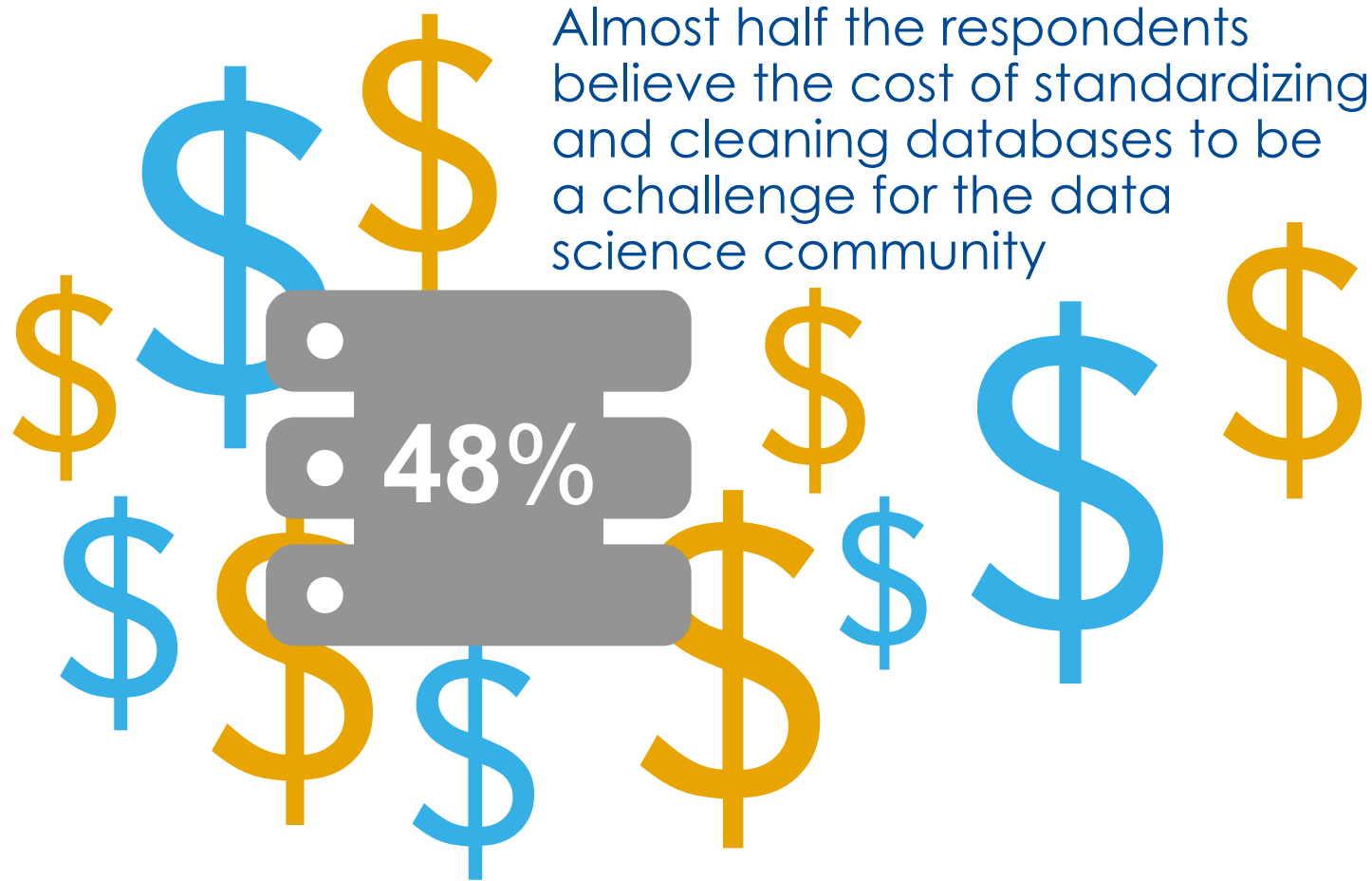
Skill gaps are cited by the almost half of all respondents as a key barrier to greater use of data science



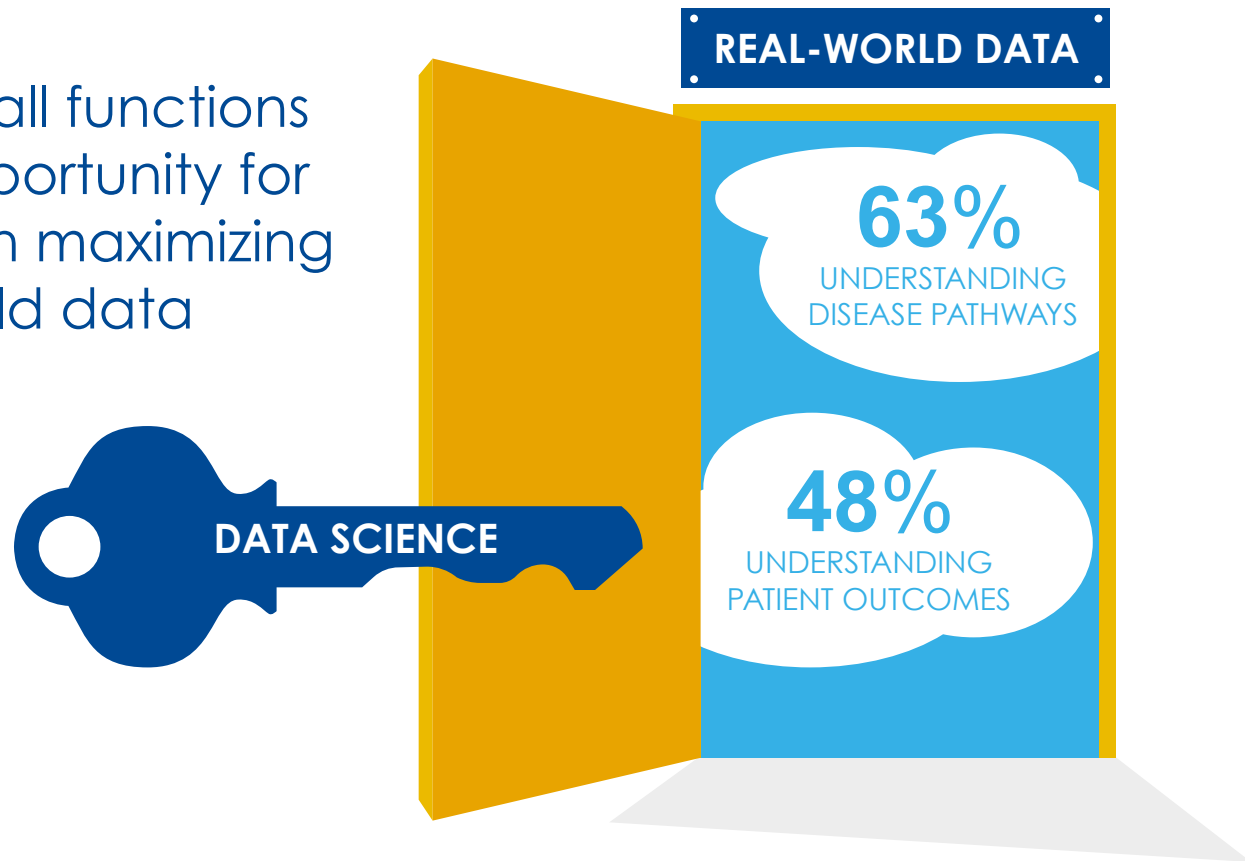
Percentages represent number of responders who selected that option – multiple responses were allowed

Providing trusted solutions is a key challenge that needs to be addressed by the data science community

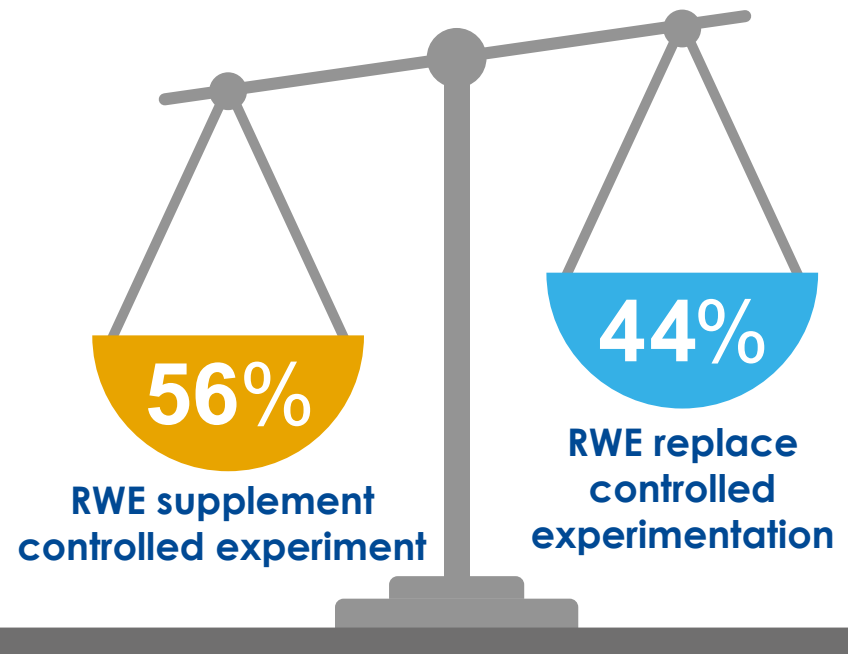


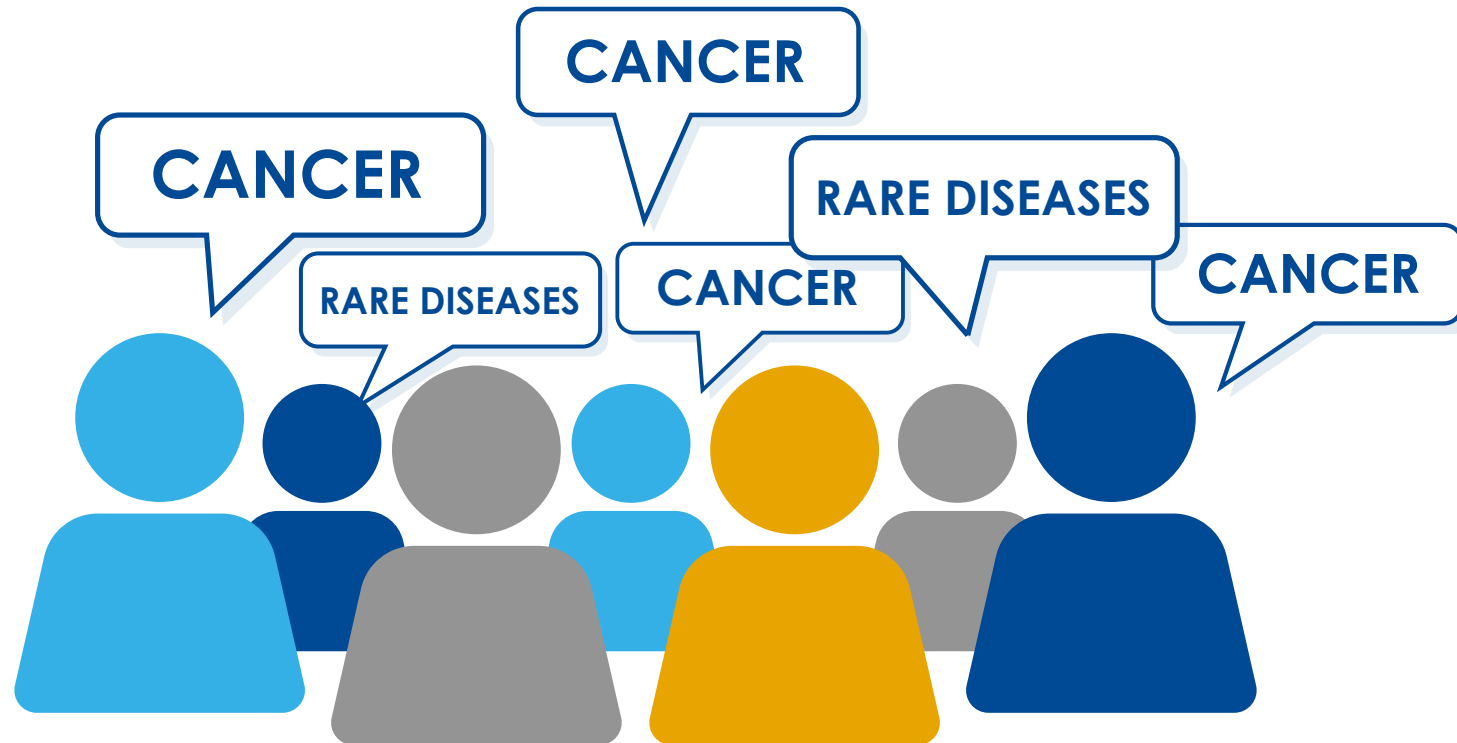


Respondents across all functions perceive the key opportunity for data science to be in maximizing the value of real world data



Data science approaches on real world evidence may even replace controlled experimentation



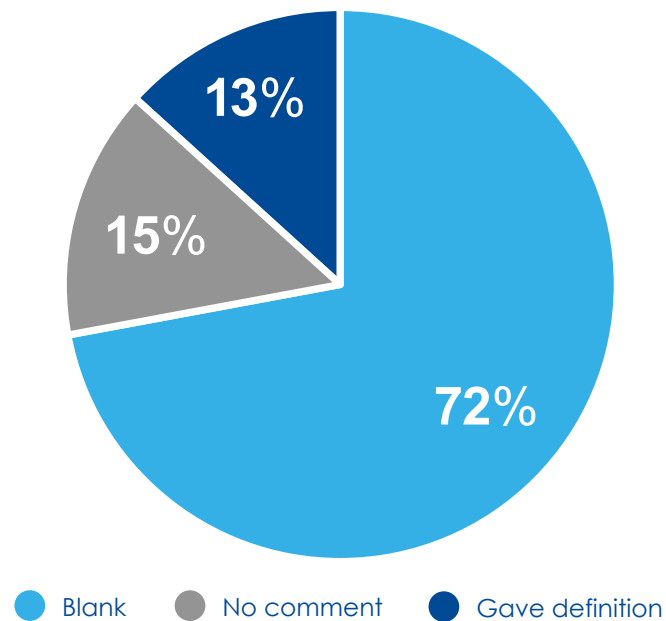


Oncology is a key therapeutic area where real-world evidence could be accepted by regulators to replace controlled experiment for the market approval of new drugs

SURVEY INSIGHTS

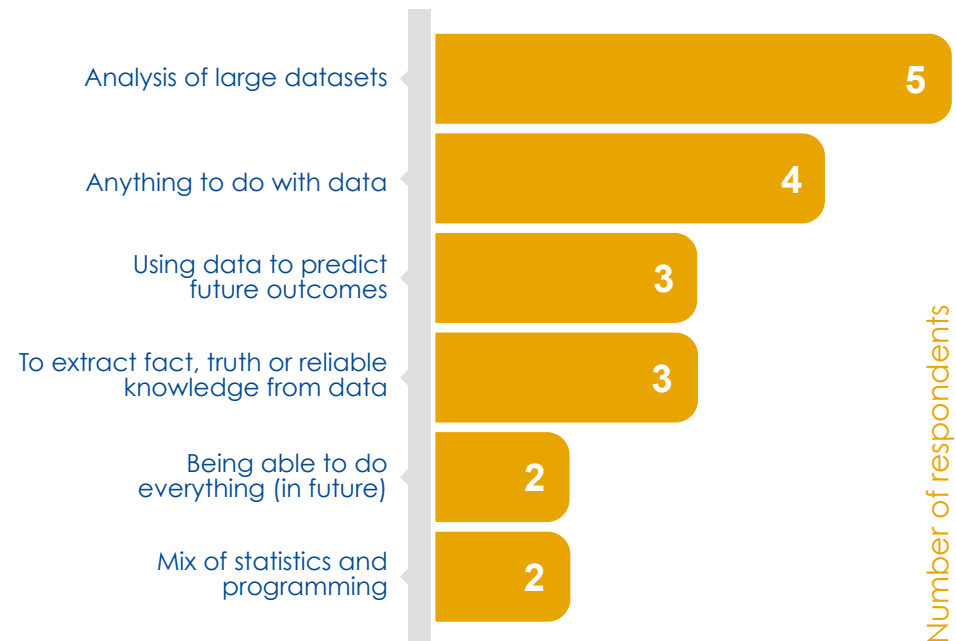
Only **13%** of respondents volunteered a definition of 'data science'

Do you have your own definition of data science?



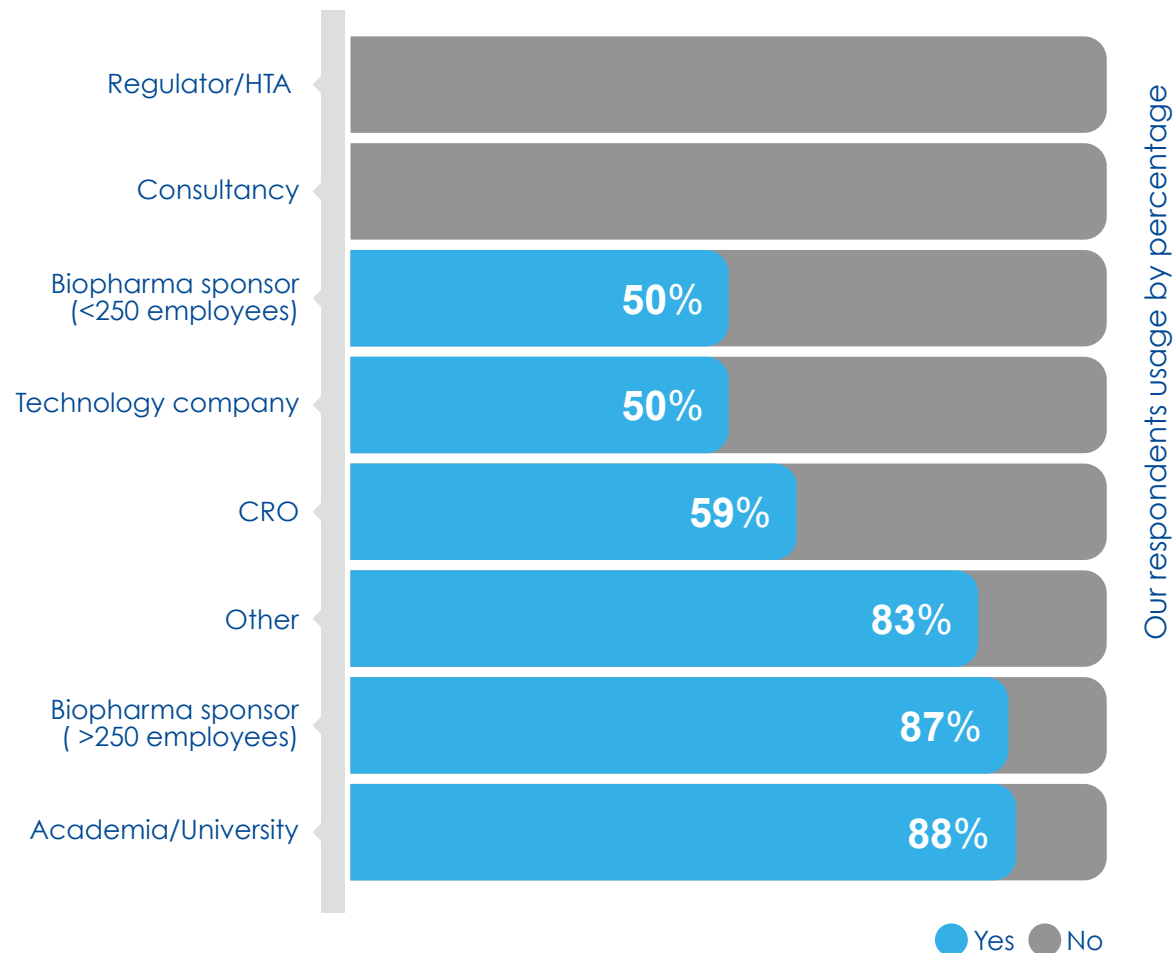
Those offering a definition did not reveal a unanimous understanding of the data science approach.

Themes from respondents' definitions of data science



Responses varied between analysis of large data sets to interpretation and extracting more reliable evidence.

Over **70%** of respondents reported that their organizations have a dedicated data science group



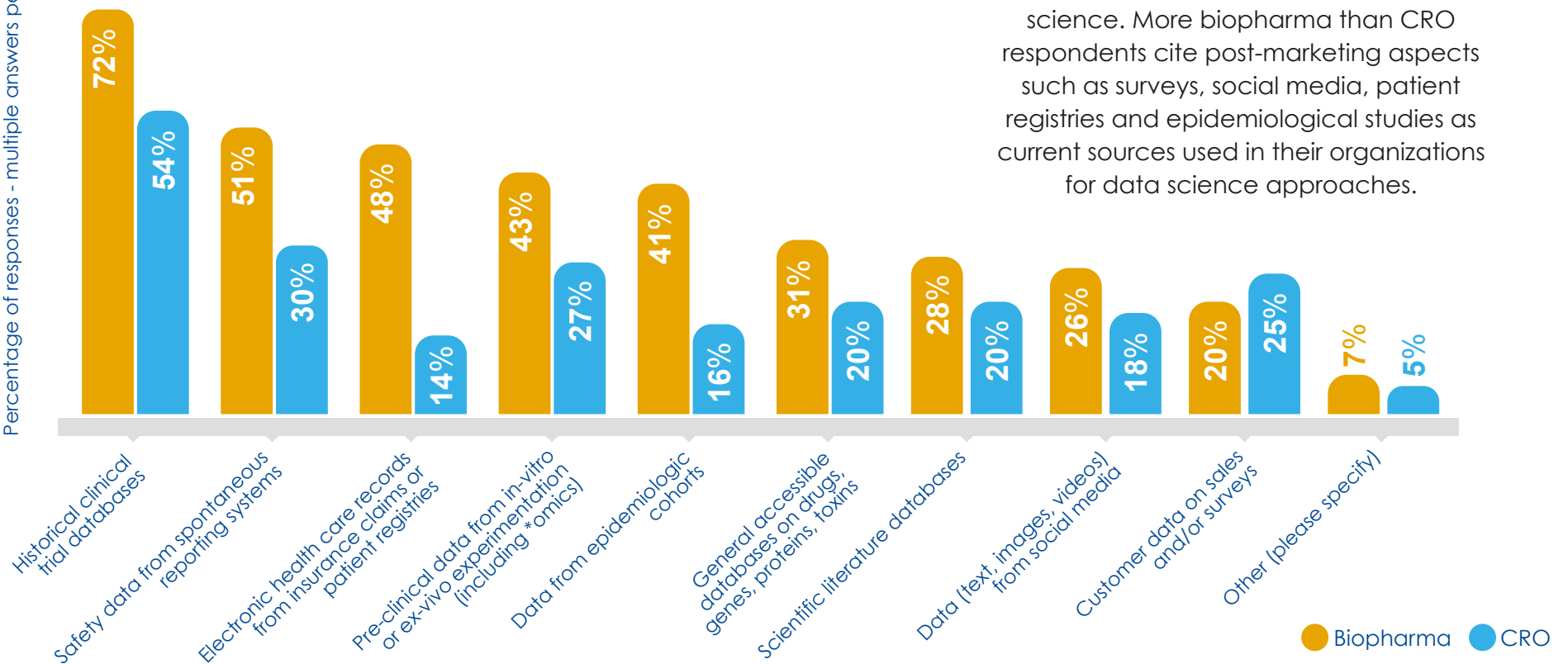
Does your organization have a dedicated data science group?

There is a higher proportion of organizations with dedicated teams in large biopharma companies than in service organizations (CROs, technology companies, and consultancies). Respondents from academia and university environments also reflected a high proportion of dedicated data science teams.

Historical clinical trial databases are the most cited source for data science approaches

Which data sources are currently used at your company for data science approaches?

Percentage of responses - multiple answers permitted

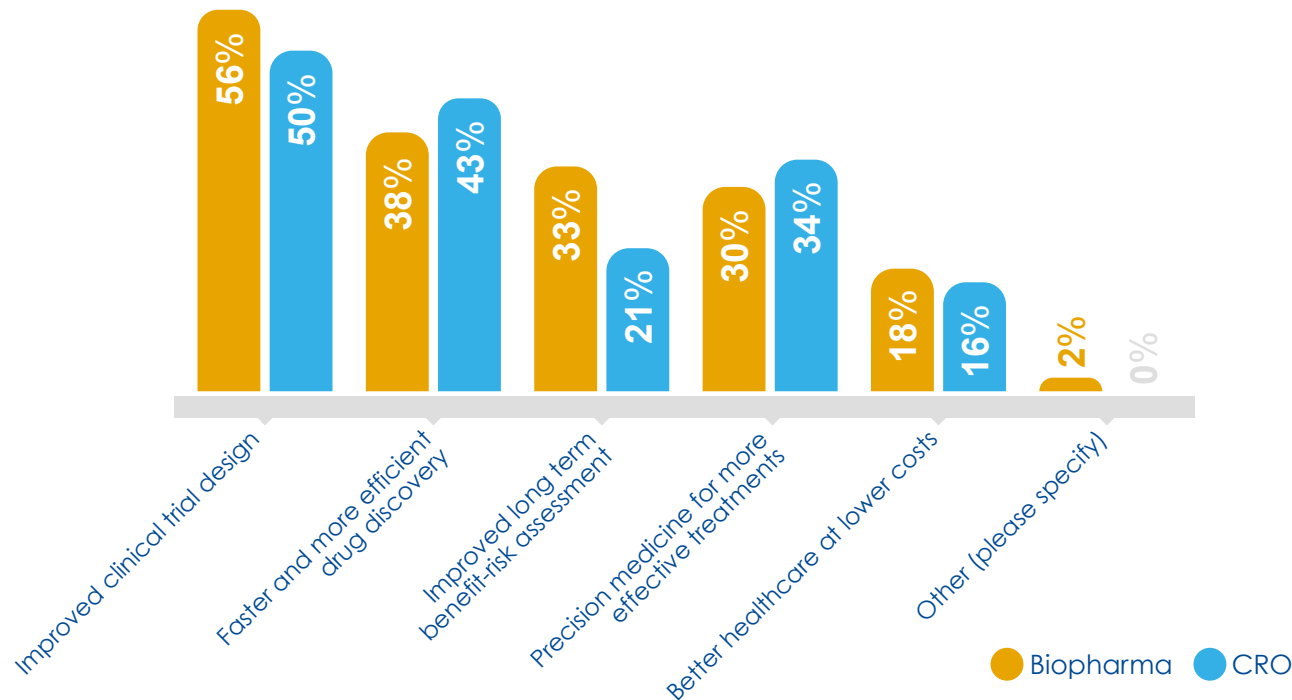


In addition to clinical trial databases, pre-clinical and safety data also emerge amongst the current sources for data science. More biopharma than CRO respondents cite post-marketing aspects such as surveys, social media, patient registries and epidemiological studies as current sources used in their organizations for data science approaches.

Respondents agree that organizational goals to improve clinical trial design are likely to benefit the most from data science approaches

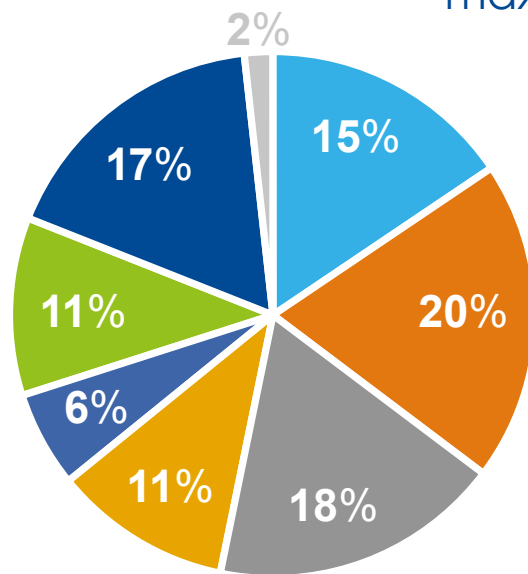
Which goals do you think will benefit most from data science approaches?

Data science is also regarded as an opportunity to advance more efficient drug discovery in order to provide more effective treatments. Providing better healthcare at lower costs and improvements in the long-term risk-benefit ratio were cited more frequently by biopharma respondents.



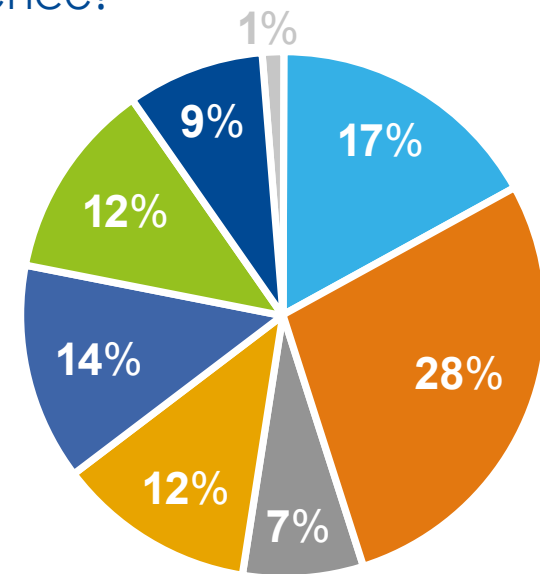
A key barrier for the industry to maximize the potential of data science is a skills gap

What are the key barriers the industry faces in maximizing the potential of data science?



Biostatisticians

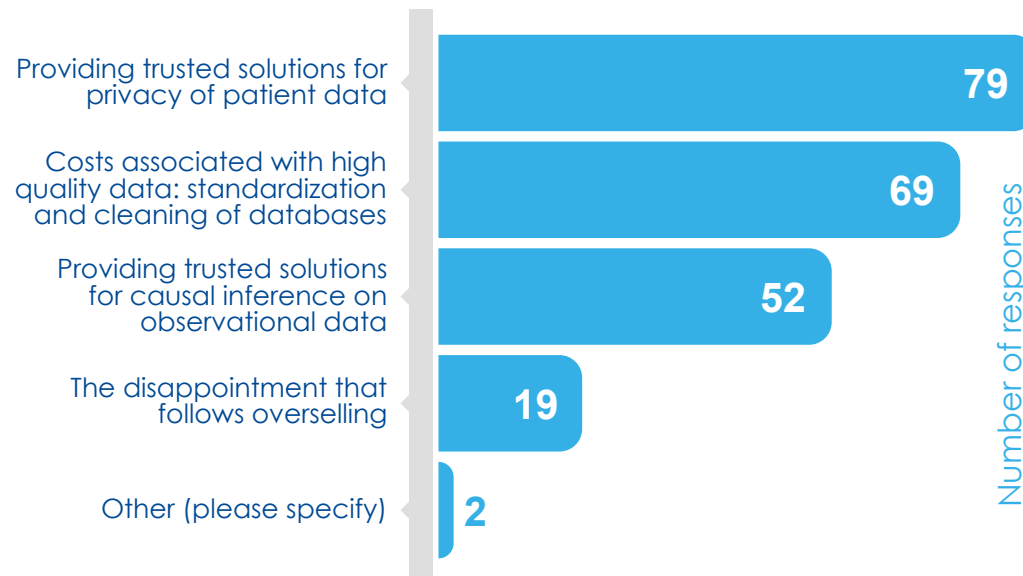
- Company silos
- Skills gaps
- Regulatory environment
- Lack of strategic vision
- Averseness to risk
- Lack of the right tools
- Lack of large standardized databases
- Other



Programmers

Multiple responses permitted. Skills gaps were cited the most as a barrier by both biostatistician and programmer respondents. The application of new approaches such as ML and AI will require new experts and disciplines to work alongside biostatisticians and programmers. In addition, the regulatory environment and the lack of large, standardized databases were widely noted by biostatisticians as an issue, and both biostatisticians and programmers often identified company silos as an issue.

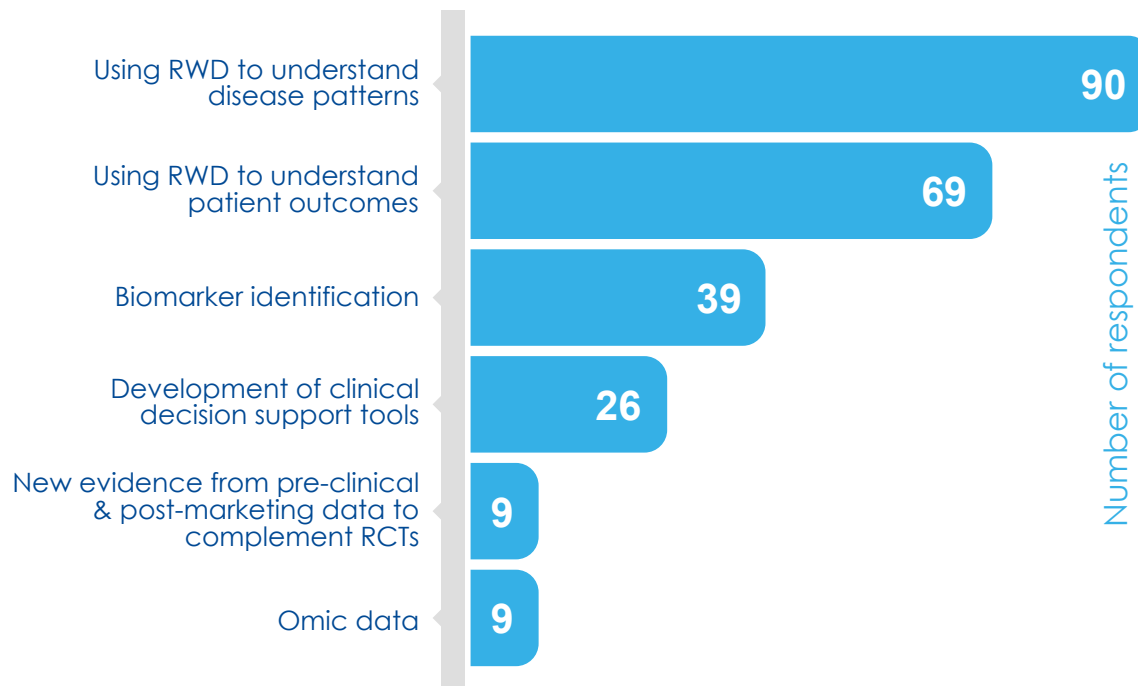
The key challenge faced by the data science community is one of trust



What are the key challenges the data science community has to overcome to maximize its potential for healthcare?

Trust emerged as the key challenge to be overcome, from the perspective of providing trusted solutions for the privacy of patient data, as well as providing trusted solutions for causal inference on observational data. A climate of enhanced privacy laws on the use of patient data may be driving this view. The costs associated with high-quality data were also concern for almost half the respondents.

The key opportunity for data science is in accessing real-world data (RWD) to better understand disease patterns and patient outcomes



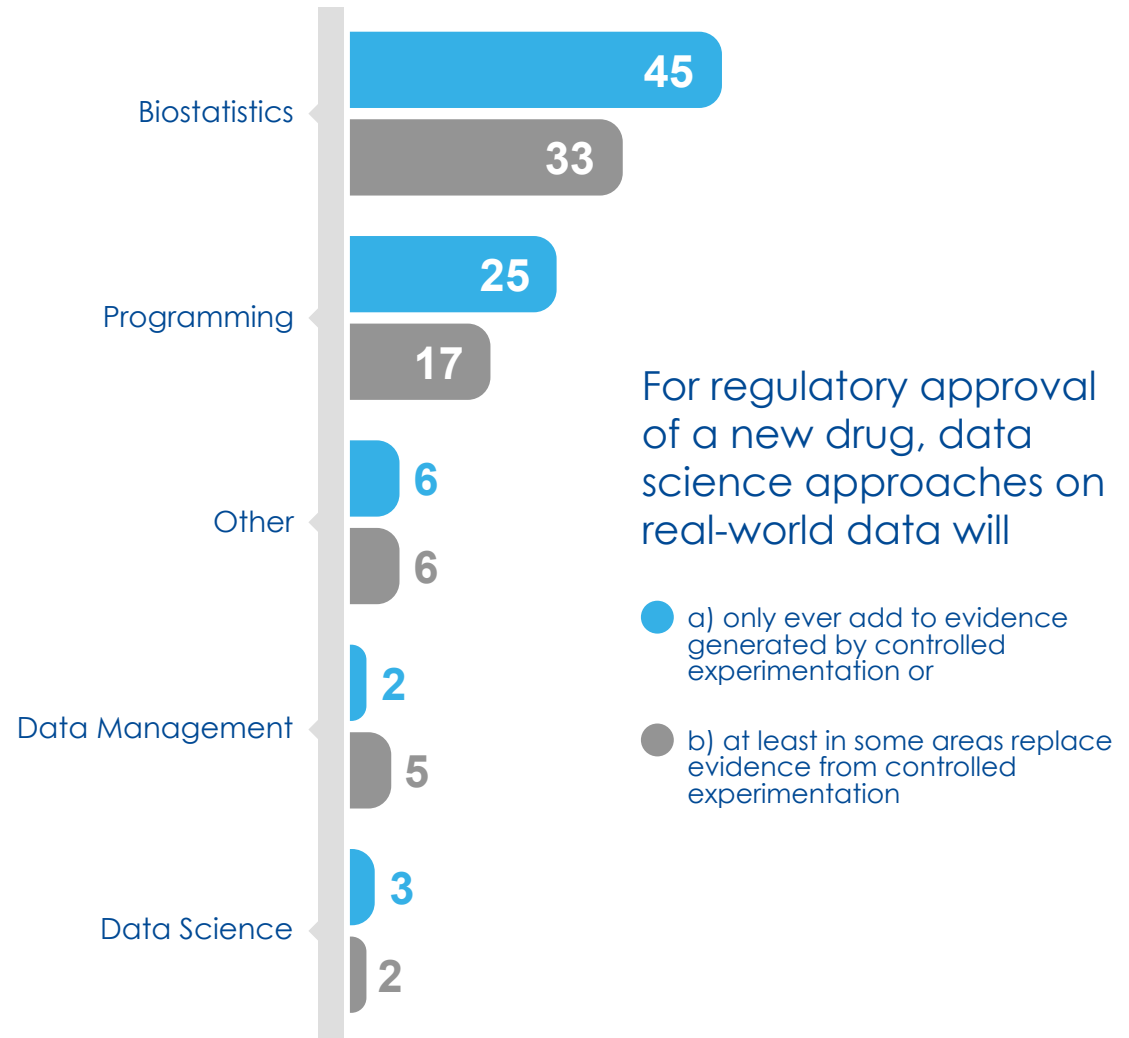
Which areas do you think represent the key opportunities for data science?

9 out of 10 respondents identified real-world data as a key opportunity for data science. Data science techniques may be particularly useful in accessing, processing, and interpreting the volume and variety of real-world data sources.

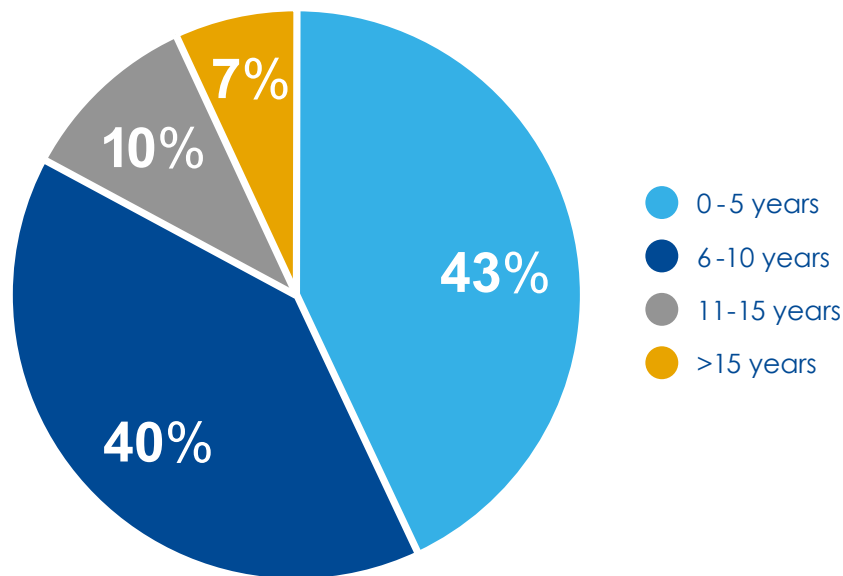
Just under half of all respondents felt that data science approaches on real-world data could at least in some areas replace controlled experimentation for the regulatory approval of a new drug

Nearly half of biostatistics respondents felt data science could lead to RWE replacing controlled experiment in some areas. This represents a shift from traditional statistical techniques being employed in the controlled clinical trial environment. In some areas this is an important distinction and further research into the perceptions of the group about the areas that this may occur could be the topic of a future survey.

Real-world evidence may not only complement but also improve clinical trial design to reflect disease patterns and patient outcomes more accurately.



"The transformation will take place within the next decade"



Responses of the 44% who believed that data science approaches on RWE could at least in some areas replace evidence from controlled experimentation for approval of a new drug.

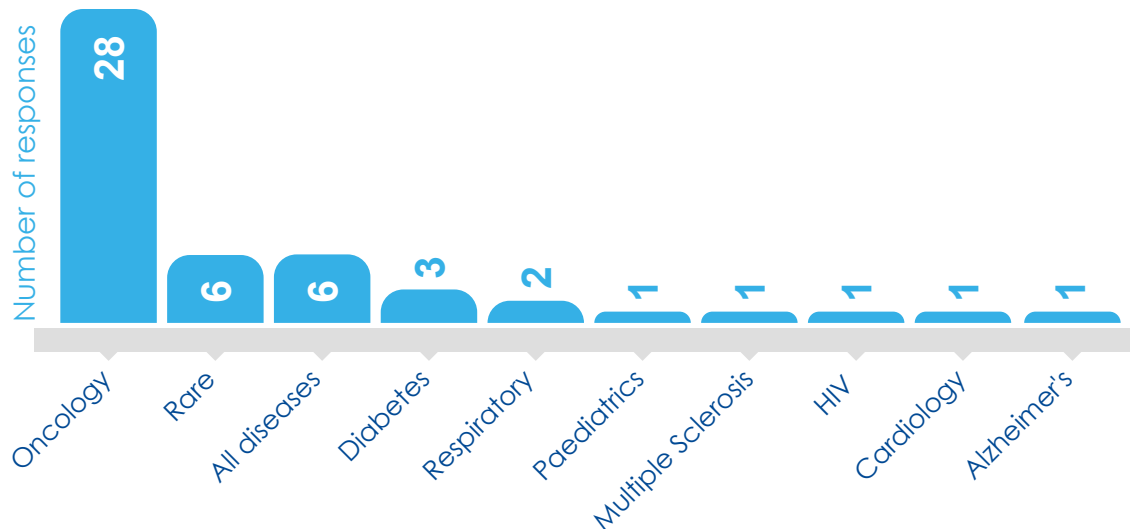
What timeframe do you envisage real-world data generated evidence being accepted by regulators for market approval of a new drug?

This question was posed only to the segment of respondents who believed that data science approaches on RWE could in some cases replace controlled experimentation for regulatory approval (63 of 144). 80% of these predicted that regulators may accept real-world data generated evidence for the market approval of a new drug within the next 10 years.

40% of this segment of respondents suggested that this could even occur within 5 years. If these predictions become a reality, then data science could lead to a transformation of how evidence is graded and regarded by regulators in determining the marketing authorization of new drugs.

Oncology and rare diseases are key areas where data science could exploit real-world data (RWD) to produce evidence that would be accepted by regulators for market approval of new drugs

Which disease areas could RWD replace controlled experiment for regulatory approval of a new drug?



Responses of the 44% who believed that data science approaches on RWE could at least in some areas replace evidence from controlled experimentation for approval of a new drug.

The segment of respondents who believed that data science approaches on RWE could in some cases replace controlled experimentation for regulatory approval, overwhelmingly cited oncology as the disease area where this was most likely to occur. Given the difficulty in recruiting patients to oncology trials, real-world evidence could provide vital validation of new treatments which struggle to gain market authorization and funding. New sources of real-world evidence can help understand the efficacy of new products in the real world clinical setting, and allow patients to make more informed treatment decisions.

There are specific challenges in oncology development that the use of RWE could help to overcome including:

- Greater understanding of biomarkers and advances in targeted therapies mean that it may not be ethical to randomize patients.
- Pivotal trials testing oncology products may also have crossover designs making it statistically difficult to show increments in survival between groups.

The future: Questions for discussion

This survey has revealed the existing commitment that organizations have made in applying data science approaches. It has also highlighted that respondents across functions believe that exploiting real world data represents the greatest opportunity for data science approaches to unlock. This could fundamentally shift the current drug development and approval paradigm. However, there are many issues to resolve, and questions to answer to enable this shift to be realized in practice.

Below, we outline some of the questions that as an industry we need to debate and address:

- ❓ In order to fulfill the potential of data science, do organizations need to integrate the existing skills of programmers and biostatisticians to overcome the reported company silos and maximize existing resources as well as develop the new skill sets required in machine learning and AI?
- ❓ With an increasing amount of drug development knowledge shifting to CROs, could there be potential for CROs to provide the expertise and capacity to oversee and develop sponsors' data science teams?
- ❓ How in practice are regulators going to incorporate using real-world evidence as part of the approval of a new drug?
- ❓ Without approval, how 'real-world' is the data gathered on new drugs prior to marketing authorization being granted?
- ❓ What confidence level will regulators need to have in evidence generated by predictive analytics and AI?
- ❓ How can patient privacy be protected whilst exploiting the valuable insights that real-world patient data could offer into disease outcomes?

A definition of data science?

In the publication, *50 years of Data Science*², David Donoho discusses the concepts of 'Lesser Data Science' and 'Greater Data Science'. His summary of 'Greater Data Science' in our opinion, is a worthy definition for consideration:

“Data Science is the science of learning from data; it studies the methods involved in the analysis and processing of data and proposes technology to improve methods in an evidence-based manner.”

Naturally, as the field of data science develops and matures so too will the definition. Cytel is looking forward to helping to shape this conversation, and collaborating with our industry colleagues to usher in this new paradigm.

References

- 1) *Getting real with real-world evidence (RWE): Deloitte Insights.*
- 2) Donoho, D. (2017). *50 Years of Data Science. Journal of Computational and Graphical Statistics, 26(4), pp.745-766.*

Cytel is shaping the future of drug development. Our solutions bring together technology, data science, data analytics, and services to drive superior outcomes and results for clinical trial sponsors.

Learn more at www.cytel.com/services/data-science

