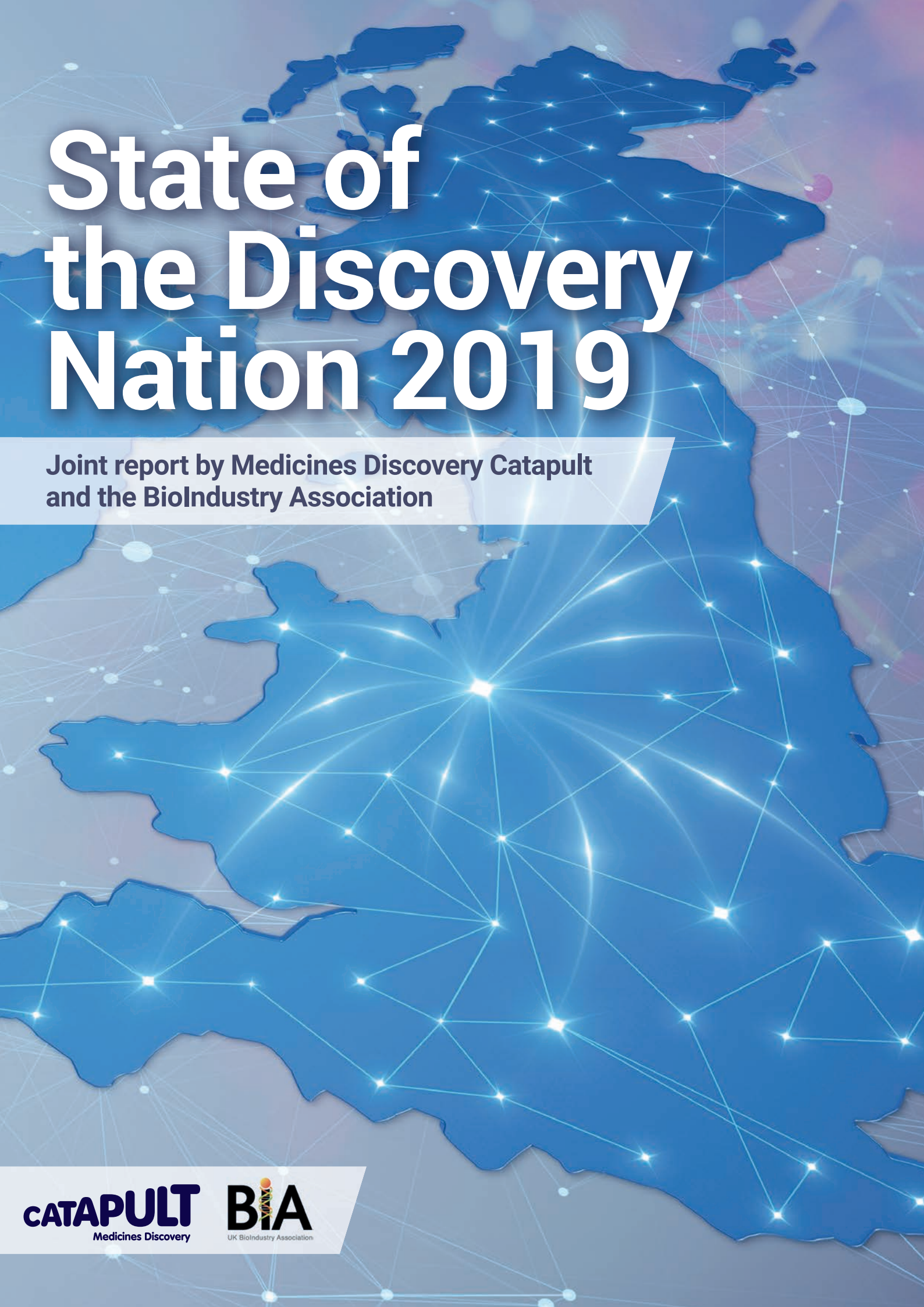


State of the Discovery Nation 2019

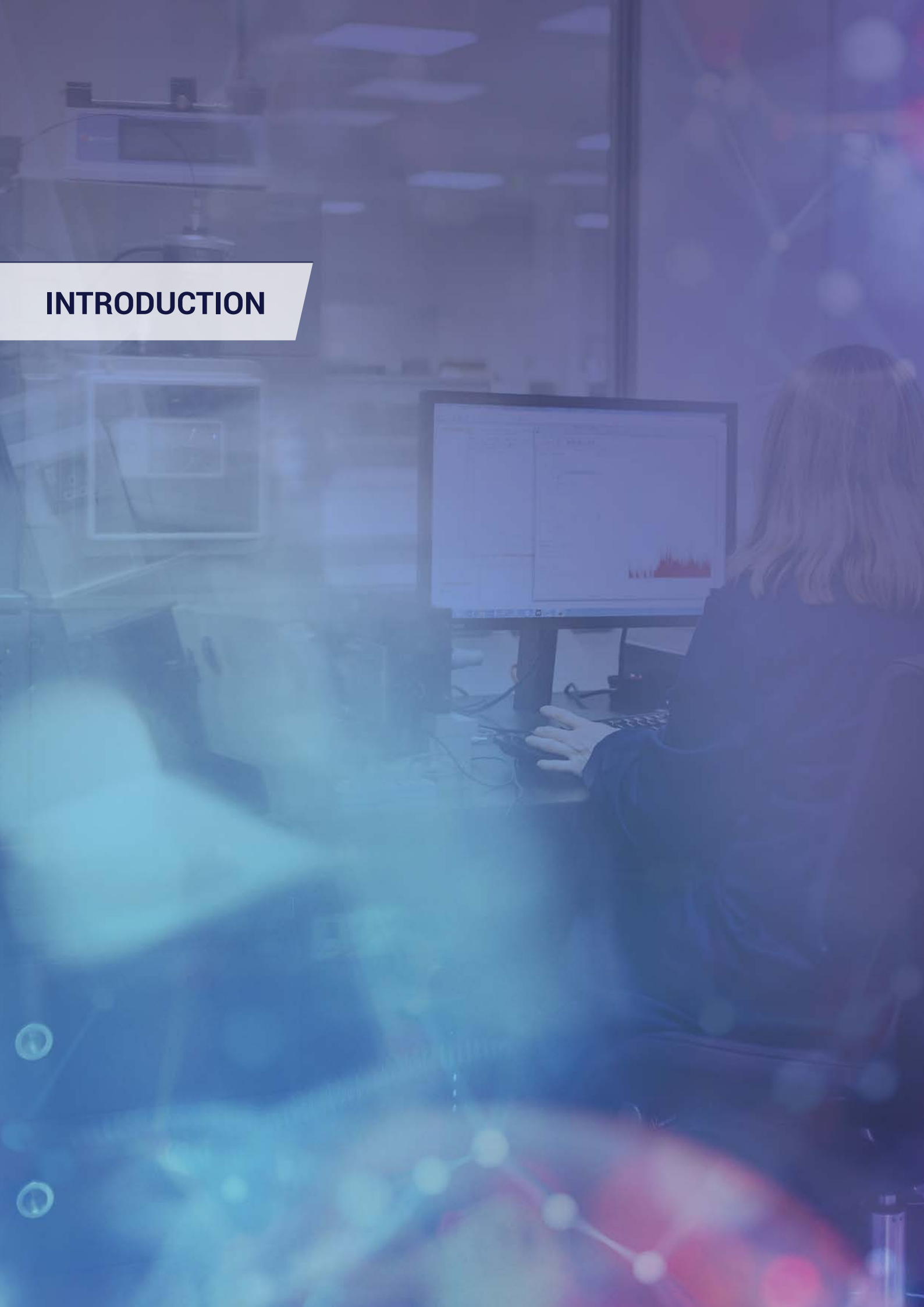
Joint report by Medicines Discovery Catapult and the BioIndustry Association

CATAPULT Medicines Discovery

BIA UK BioIndustry Association

CONTENTS

INTRODUCTION



This second State of the Discovery Nation report provides key insights from Medicines Discovery Catapult's (MDC) market research, surveys, interviews, and experience from the drug discovery community.

MDC has two main purposes:

1) Identifying, industrialising and driving the adoption of technologies and methods that will improve productivity and predictability of medicines discovery.

Two areas of Research and Development (R&D) selected to industrialise new tools are Humanised preclinical models with combinations of advanced bioanalytical tools and Artificial Intelligence (AI)/Machine Learning systems.

2) Small or medium-sized enterprises (SMEs) and translational academics require efficient access to UK infrastructure, both public and private, to support their R&D. MDC has therefore developed 3 platforms:

- **Virtual R&D:** Brokered access to outsourced expertise and experimental support
- **Samples & Data:** Brokered access to consented patient samples and clinical data sets
- **Syndicates:** Managed consortia of medical research charities and other providers to affect portfolios of patient-driven discovery projects

Figure 1

The first State of the Discovery Nation report[1] reviewed the key issues with medicines discovery productivity and identified key strengths in the UK's science and talent base, as well as challenges in translation and commercialisation.

This second report — based on new survey and interview data — focusses on two scientific areas of MDC focus:

- AI for drug discovery
- Complex (preclinical) cell models (CCMs)

It also reviews the size and shape of the sector and captures the most recent views expressed by the community.

KEY INSIGHTS

- **80% of SMEs in the community are Service and Supply companies, accounting for 90% of employment**
- **60% of companies have fewer than 5 staff; 80% have fewer than 20 staff**
- **70% of drug assets are in cancer, anti-infectives, or central nervous system (CNS)**
- **Cancer is the strongest therapeutic area for UK companies**
- **Companies working in anti-infectives face a challenging market**
- **AI and cell and gene therapies were deemed the hottest areas of 2018**
- **Generally companies are calling for the government to prioritise direct financial support rather than infrastructure support**
- **90% of companies need AI; currently 75% of AI spend is on data access and curation**
- **CCMs show much promise to reduce animal usage but need validation before large-scale use**

This SME market review is intended to deepen the understanding of this vital sector, as well as adding insight on how MDC can continue to best support the community.

MDC will continue to conduct research to listen to the sector and to shape its strategy in response to evidence.

[1] https://md.catapult.org.uk/resources/report-state-of-the-discovery-nation-2018/

[2] *'Service' is used in this document to mean companies delivering services for companies developing medical assets, such as Advisory (e.g., Market Analysis/Information Consultants/Communications/Specialist Consultants with the Regulatory Expertise [organisations]); Outsourcing (e.g., Clinical Research Organisation + Contract Manufacturing/Research Organisations); and Materials Supply companies*

[3] *'Core' is used in this document to mean companies actively researching and developing their own human therapeutic products*

SME MARKET OVERVIEW

KEY MESSAGES

- There are 1,500 UK medicines discovery SME companies split 80% Service[2] and Supply; 20% Core[3] companies
- UK medicines discovery SME companies employ 21,000 people, with 8,000 in research-outsourcing companies
- Advisory companies are the most common; 80% of these employ fewer than 5 people
- Core companies employ only 2,500 staff; fewer than 5% have over 50 staff
- Over half of sector staff are in larger service companies with over 50 staff
- Therapies other than small molecules now represent over half of all Core companies
- 70% of assets are in oncology, anti-infectives, or CNS

This overview focusses on companies actively researching and developing their own human therapeutic products (referred to as the Biopharmaceutical Core or 'Core' sector to match the Office of Life Science (OLS) nomenclature[4]), together with the specialist service and supply chain companies that support medicinal product developers of all size (referred to as the Biopharmaceutical Service and Supply or 'Service and Supply' sector). This cohort excludes cell and gene therapy companies, because these are supported by the Cell and Gene Therapy Catapult.

Analysis is based on publicly available data from the OLS, Companies House (2017 data), company websites and MDC internal data. Analysis shows that the market moves quickly, with companies starting-up and closing each year, and some growing rapidly. This snapshot is, therefore, only an estimation of the picture in 2018 and will evolve each year.

Company numbers and types

The Biopharmaceutical sector has 1,500[5] medicines discovery SMEs, split between 1,200 Service and Supply companies (80%) and 300 Core companies (20%). This is represented in Figure 2 which highlights the key segments for the Service and Supply sector. The largest segment is Advisory (combining 300 general advisory companies and 80 regulatory advisory companies), followed by Outsourcing and Materials Supply. Core companies are a small minority of the total company numbers. The BioIndustry Association's (BIA) vision for the sector in 2025[6] shows that the UK science base should support 50 more biotech companies at early clinical stage.

The UK science base should support 50 more biotech companies at early clinical stage

300 Core companies, 1,200 Service and Supply

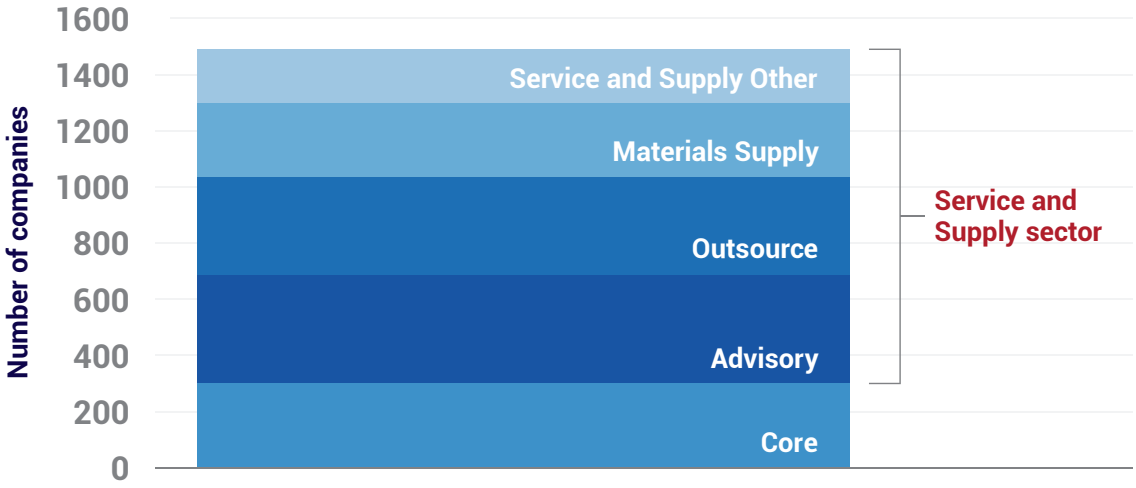


Figure 2

[4] *https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/707072/strength-and-opportunity-2017-bioscience-technology.pdf*
[5] *Estimates have been rounded to reflect the volatility in company numbers*
[6] *A vision for the UK life science sector in 2025, BIA*

As cell and gene therapy companies currently account for about a quarter of biotech asset-developing companies, this would imply net growth of MDC's Core group by around 35-40 early clinical companies over the next 6 years, meaning a much higher number of start-ups in early discovery.

Company staff numbers

Core and Advisory companies have lower staff numbers than Outsourcing and Supply companies. Our estimates show that of the 21,000 staff in the MDC's SME community, 8,000 are working for Outsourcing companies and 5,000 for Material Supply companies (Figure 3).

SME employment is mainly outsourcing and materials supply

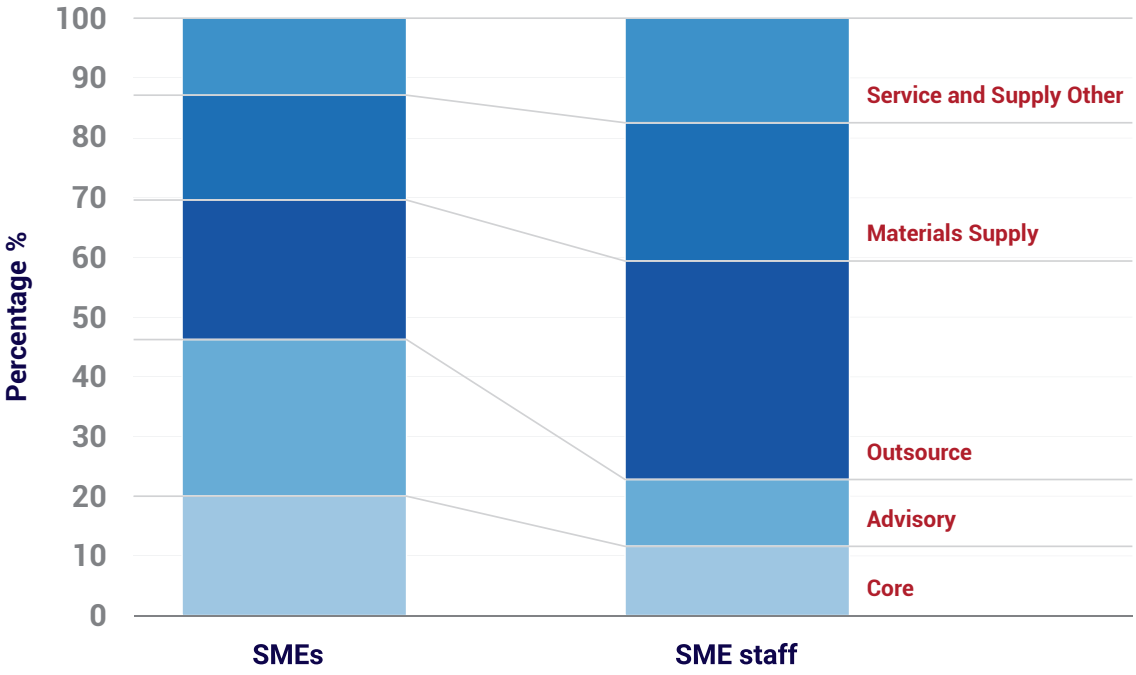


Figure 3

The Service and Supply sector

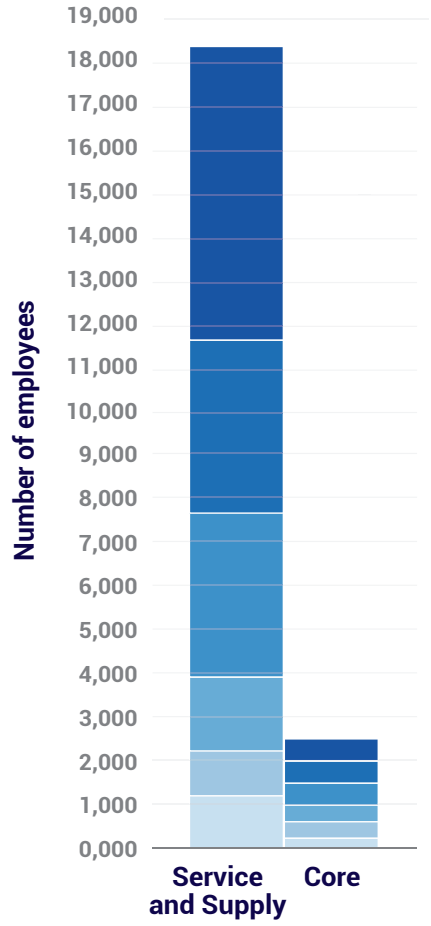
The Service and Supply sector is comprised of SMEs which offer services and goods to SMEs and larger companies. An estimated 1,200 Service and Supply SMEs exist, with three segments being responsible for over 80% of all the companies:

- Advisory — 33%
- Outsourcing — 29%
- Material Supply — 22%

A breakdown of company size (based on the number of employees) for the Service and Supply sector is shown in Figure 4.

Most of the companies in the sector have 4 or fewer employees. This is driven primarily by the Advisory segment where 80% of this group of companies employ fewer than 5 people. The small size of companies makes efficiency in networking paramount for this group, as they must balance exploring opportunities and collaborations against delivery of contracts.

Total employment:
Core vs Service and Supply



Company staff numbers range:
Core vs Service and Supply

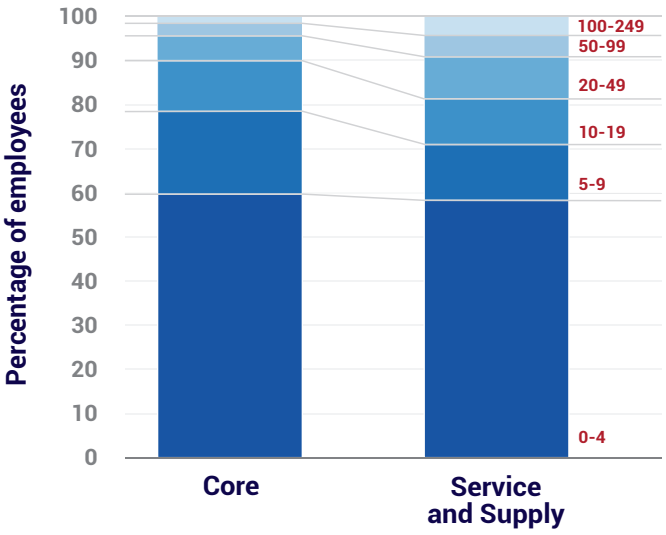


Figure 4

However, there is a higher proportion of larger SME companies in the Service and Supply sector, meaning that half of the sector's staff work in Service and Supply companies with over 50 staff. These larger SME companies are mainly contract research organisations (CROs) and Material Supply companies, where company success is often associated with headcount growth. The BIA's 2025 benchmarking showed that the UK could support an additional 33,000 biotech jobs.

This analysis suggests that significant jobs growth within the medicines discovery sector is likely to come from larger SME Service and Supply companies.

The Core sector

The Core sector comprises 300 SMEs that are researching and developing their own prescription pharma products. The Core sector (excluding cell and gene therapies) is largely made of four molecule types, all supported by the MDC: small molecule; therapeutic protein; vaccine; and antibodies. Overall, the Core sector is dominated by small molecule research companies that represent over 60% of Core SMEs.

Small molecules are no longer the majority therapeutic modality in UK SMEs medicines discovery

There is increasing diversity in therapeutic modalities. For example, if you add in the UK's 100 cell and gene therapy asset companies[7] (CGT), small molecules are no longer the majority of the UK's medicines discovery and cell and gene Core companies (Figure 5). All of these companies are now supported by Catapults.

Increasing diversity in therapeutic modalities

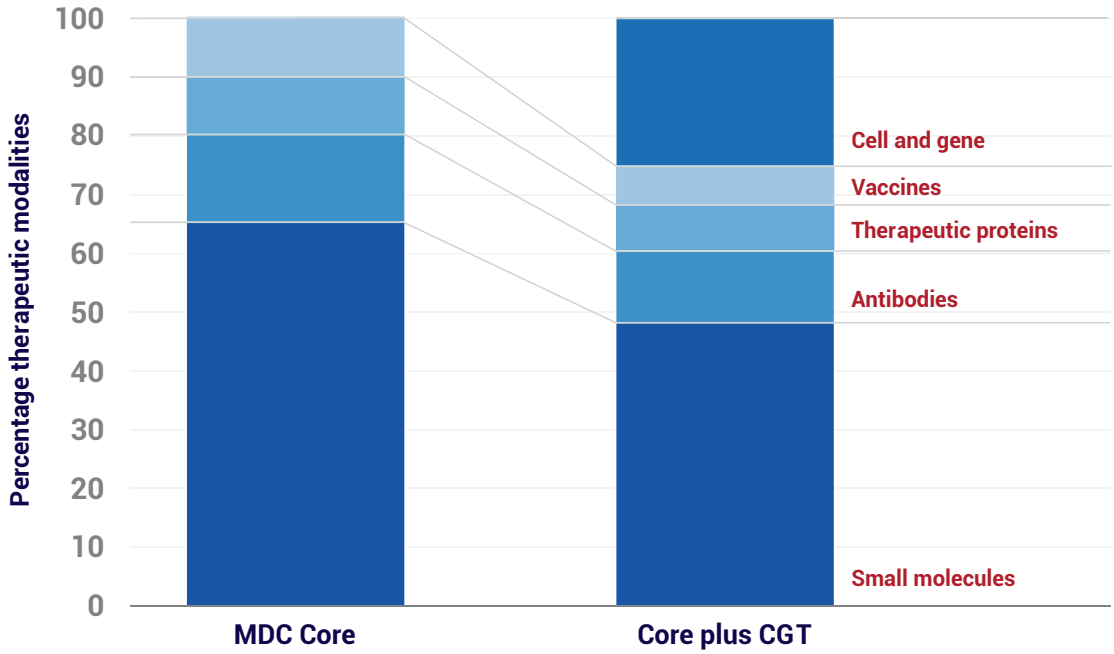


Figure 5

Like the Service and Supply Sector, the majority of SMEs in the Core sector employ 4 or fewer people (Figure 4). This shows that their development work is being outsourced, requiring a functioning market for services and supplies to allow Core staff to build and manage complex research programmes.

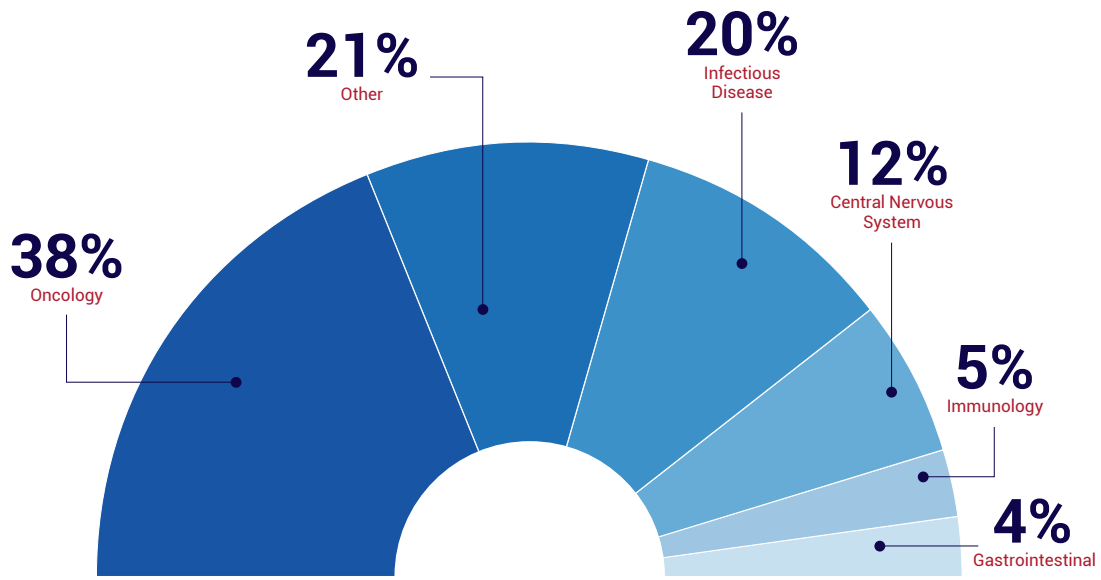
The majority of SMEs in the Core sector employ 4 or fewer people

[7] *NB, this number is taken directly from the OLS data and not reviewed further as these companies are in MDC's scope*

Unlike the Service and Supply Sector, there are very few Core companies with over 20 staff, due to the tendency for smaller UK companies to be acquired after key milestones are met. This would suggest that Core companies focus on scientific project progress rather than organisational growth, whereas Service and Supply companies are more likely to need additional people to deliver more services to maintain long-term productivity. It also reinforces BIA and MDC's concerns that the UK market for growth capital is not allowing UK SMEs to build in size locally.

Cancer, Anti-infectives and CNS make 70% of the assets in development

Further analysis using information from Global Data suggests that UK SMEs are currently developing over 930 drug assets. The most common therapy area was Oncology (349 drugs), followed by Infectious Diseases (182) and then CNS (115) (Figure 6). This is different from the global large pharma company late-stage assets mix identified by Deloitte[8], which also has oncology as the largest group but has fewer infectious disease and more respiratory and immunology assets.



Core Biopharma - Drugs by Therapy Area

Figure 6

Source: Global Data

[8] *Unlocking R&D productivity. Measuring the return from pharmaceutical innovation 2018. Deloitte Centre for Health Solutions, 2018*

ASSET OWNER SNAPSHOT

KEY MESSAGES

- **Cancer cluster is strong and supported with world leading science, funding, and coordination**
- **Anti-infectives cluster is large but faces commercialisation challenges**
- **CNS is heterogenous, with Dementia showing particular potential**
- **Respiratory and musculoskeletal disease have relatively few preclinical companies**

Snapshot of five major asset areas

Nearly half of the SMEs in this subset are focussed on oncology

Deeper analysis of Core SMEs was performed across five prioritised therapy areas: Oncology, including Immuno-Oncology; Anti-infectives; CNS, including Alzheimer's Disease and Pain; Musculoskeletal disorders; and Respiratory[9]. More detailed analysis of a subset of these SMEs, which were confirmed as active across one of the five therapy areas, are defined here as 'Asset Owners'. Nearly half of the SMEs are focussed on oncology, with an equal balance between preclinical companies and those with assets in the clinic (Figure 7). A strong understanding of the biology of cancer, and the ability to connect clinical unmet need with basic science, provides a solid basis for translation, company formation, and growth. This is in line with evidence from the State of the Discovery Nation 2018 report[10], where 100 UK experts rated the strength of UK basic science and translational medicine in oncology top of the 12 therapy areas examined.

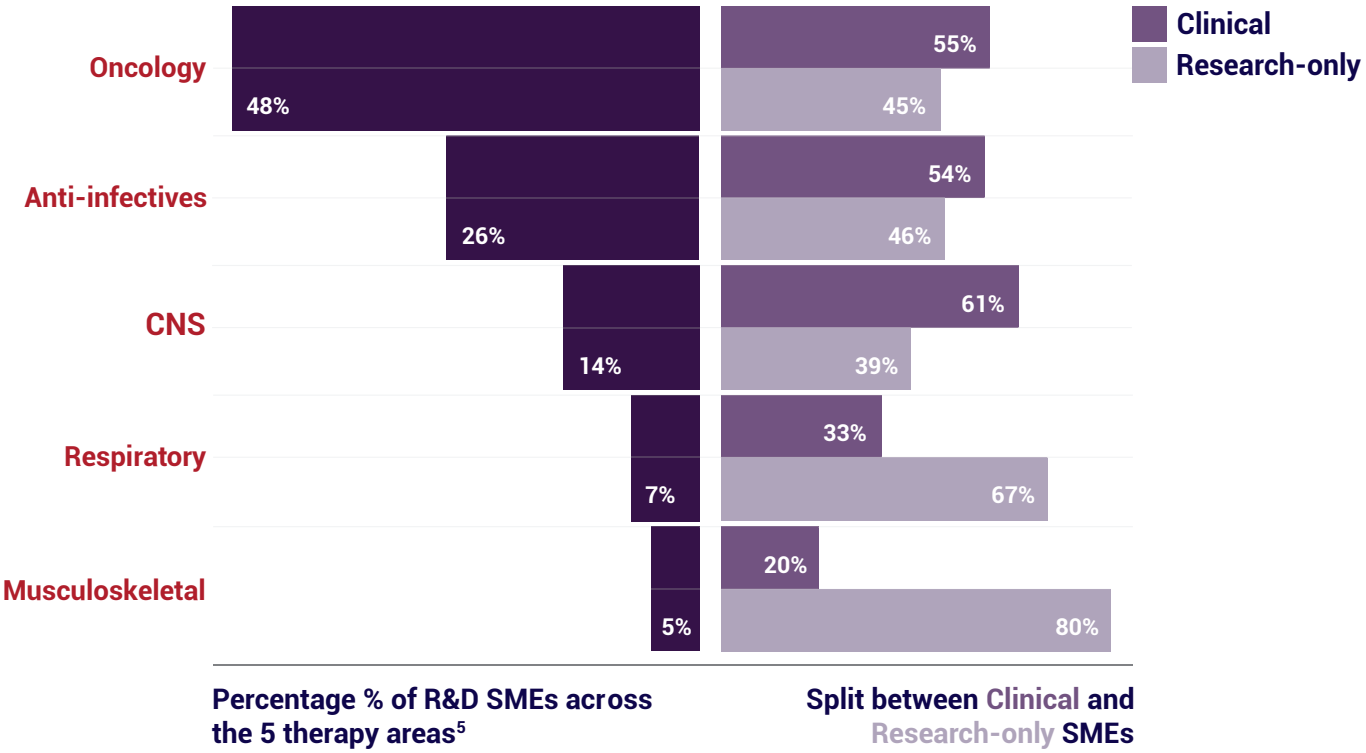


Figure 7

[9] In addition to searching publicly available company information, two main data sources were used: UK Biotech Company List2, and BioPharm Insight Database3. A standard set of search criteria was used, including a limit of 10 drugs in the portfolio to ensure we focussed on SME community of companies. Search criteria included: i) key words for the Therapy Area, ii) UK Head Quarters, and iii) ≤10 drugs in the portfolio across all Phase of Discovery and Development. Companies with no news or external communications in the last 3 years were excluded, unless there was clear evidence from other sources that the company was still solvent and operating. Technology platform companies with projects in >1 Therapy Areas, are included in each of the Therapy Area numbers. http://biopharmguy.com/links/country-united-kingdom-all-location.php, BioPharm Insight (www.biopharminsight.com)

Cancer has an accessible network of experts that helps small companies

Important drivers for this strength are the presence of a cohesive research community, led by Cancer Research UK (CRUK), and the commercial opportunity for new oncology drugs. Large charities like CRUK increase the strategic coordination of research, and via initiatives such as the National Cancer Research Institute, attempt to make research and clinical experts more accessible. CRUK's public donations increase the funds available for oncology research and early development in academia and in SMEs. The continuing commercial success of oncology companies and drugs through late-stage development and reimbursement attracts venture capital and pharma investment in this therapy area. As a result, the UK oncology SME sector is amongst the best funded and performing in the UK, despite the ongoing challenges faced, as illustrated by the recent £65m Series B funding of Artios Pharma.

The lack of commercial pull in anti-infectives means support is needed all the way to adoption

Although there are significant number of anti-infective companies evenly spread across the preclinical and clinical phases, the growth drivers for the anti-infectives UK SME sector are very different. The preclinical research part of this sector is growing. This is driven by government-led public awareness of the threat posed by antibiotic resistant microbes, organisations such as the Antimicrobial Resistance (AMR) Centre, and government and charity funding (e.g., National Institute for Health Research, UK Research and Innovation, and the Wellcome Trust). This funding is complementary to the academic strength in basic anti-infective science identified in the State of the Discovery Nation 2018 report[10].

In contrast to oncology drugs, uncertain pricing and the adoption of new anti-infective treatments as a 'last resort' means they are perceived to have a low market potential. This makes securing venture capital funding challenging. Licensing and partnerships are rare as many major pharma companies have stopped late-stage development and commercialisation of anti-infectives. The State of the Discovery Nation 2018 report[10] reflected this situation, as the anti-infectives area was rated weak in translational medicine, suggesting that anti-infectives SME landscape could benefit from translational support. However, the lack of commercial pull means that translational support is needed not just across the 'valley of death' but also all the way to adoption, or there is a risk of molecules and companies stalling at later phases. In this context, the UK government's commitment to "develop and test new models for national purchasing arrangements that de-link the price paid for antimicrobials from the volumes sold"[11] could make a real impact.

[10] *https://md.catapult.org.uk/resources/report-state-of-the-discovery-nation-2018/*
[11] *https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/773130/uk-amr-5-year-national-action-plan.pdf*

CNS may be an area where UK SMEs can grow in strength

Drawing conclusions from the relatively small number of CNS SMEs is complicated by the very diverse nature of this therapy area. It includes Pain, Psychiatry, Neurodegenerative diseases and Dementia, with companies often working on more than one indication. However, there is commercial and public interest for new products in these diseases. Dementia stands out most prominently (nearly 40% of these companies) and, with growing charity and academic sector coordination (e.g., the Dementia Discovery Fund, Alzheimer's Research UK, and UK Dementia Research Institute), may be an area where UK SMEs can grow in strength.

UK's respiratory musculoskeletal areas have fewer preclinical companies

The Respiratory and Musculoskeletal SME sectors pose a different challenge, as few SMEs were identified in these sectors, one-third of which are preclinical. There is a risk that the UK is losing the critical mass of translation experience in these two therapy areas.The small community of clinical stage SMEs in these areas could be difficult to sustain over time due to a lack of local preclinical companies producing assets for clinical research. Unlike anti-infectives, major pharma companies are still investing in both respiratory and musculoskeletal, creating commercial pull for products. However, based on this snapshot, the UK's Respiratory and Musculoskeletal companies may need support to meet the global demand.

There is a risk that the UK is losing the critical mass of translation experience in these two therapy areas

In summary, the UK R&D SME landscape across five therapy areas indicates substantial differences in numbers of early- and late-stage SMEs across the therapy areas. Oncology is the only area with a global scale cluster, with potential in CNS, scale issues appearing in preclinical musculoskeletal and respiratory, and market issues in anti-infectives.

A predictable market attracts investors and provides the much-needed pull for SMEs to translate vital new medicines. As seen in oncology, it appears that a strong coordinating organisation or a government-driven initiative, which brings together scientific and clinical expertise and is able to provide funding, can provide essential network access and strategy development benefits to smaller companies. CNS, musculoskeletal and respiratory combine commercial pull and an active community of experts and companies that could benefit from support.

THE SME POINT OF VIEW

KEY MESSAGES

- AI and cell and gene therapies are deemed the hottest areas
- Mixed overall sentiment, with significant Brexit concerns
- Data science is needed across all areas
- CCMs are needed, especially in oncology and CNS
- About two-thirds awareness and two-thirds interest in MDC's themes
- SMEs do not recommend the current infrastructure support in general
- MDC's 'Net Promoter Score' is increasing — the SME community is increasingly likely to recommend working with MDC
- MDC should continue to do more collaboration building and enable access to 'science industrialisation' services
- 75% recommend government support around funding and tax incentives

An online survey and in-depth interviews were conducted to understand the opportunities and challenges felt by companies in the sector, and to determine what support they were asking for.

Half of respondents to our online survey[12] felt that AI and cell and gene therapies were the areas of discovery that had increased most in value in 2018. There was less agreement on the innovation that has decreased the most in value, with a quarter identifying small molecules or high-throughput screening. The sector's feelings about the future were mixed, with 80% viewing Brexit negatively, and recruitment issues being the most commonly expected consequence.

Data science is needed across all areas, and CCMs were particularly sought after in oncology and CNS disorders. There is increasing recognition for MDC: 69% of companies were interested in MDC's themes, and companies are increasingly willing to recommend working with MDC to their peers. SMEs asked for a range of support from MDC including access to high-risk science and collaboration building. SMEs did not recommend most of the current UK non-financial support mechanisms, with three quarters recommending government support via grant funding and tax incentives.

Sector recognition of MDC

Most respondents were aware of the range of support that MDC can offer and felt that MDC areas of expertise were of interest to them. The community is increasingly positive about MDC: its 'Net Promoter Score' was the most improved of 13 business support structures[13], and it is now third highest ranked, behind Innovate UK's grants and the Medicines and Healthcare Products Regulatory Agency (MHRA).

The major requests of MDC (Figure 8) were:

- **Collaboration:** connecting and linking collaborators, and brokering introductions between larger and smaller players
- **Access to science:** providing niche technologies, and access to datasets and screening libraries
- **New models of engagement:** engaging earlier with companies and being able to invest in joint projects

[12] *In 2018 the MDC and BIA repeated its annual online survey of the community to understand the opportunities and challenges encountered by companies in the sector, and what support was being requested.*
[13] *Academic Health Science Centres, University technology transfer offices, University contracting department, Genomics England, Academic Health Science Networks, The Knowledge Transfer Network, NICE, The Cell & Gene Therapy Catapult, NOCRI and the CRNs for clinical trials, National Biologics Manufacturing Centre, MDC, Innovate UK's competitive grants, MHRA*

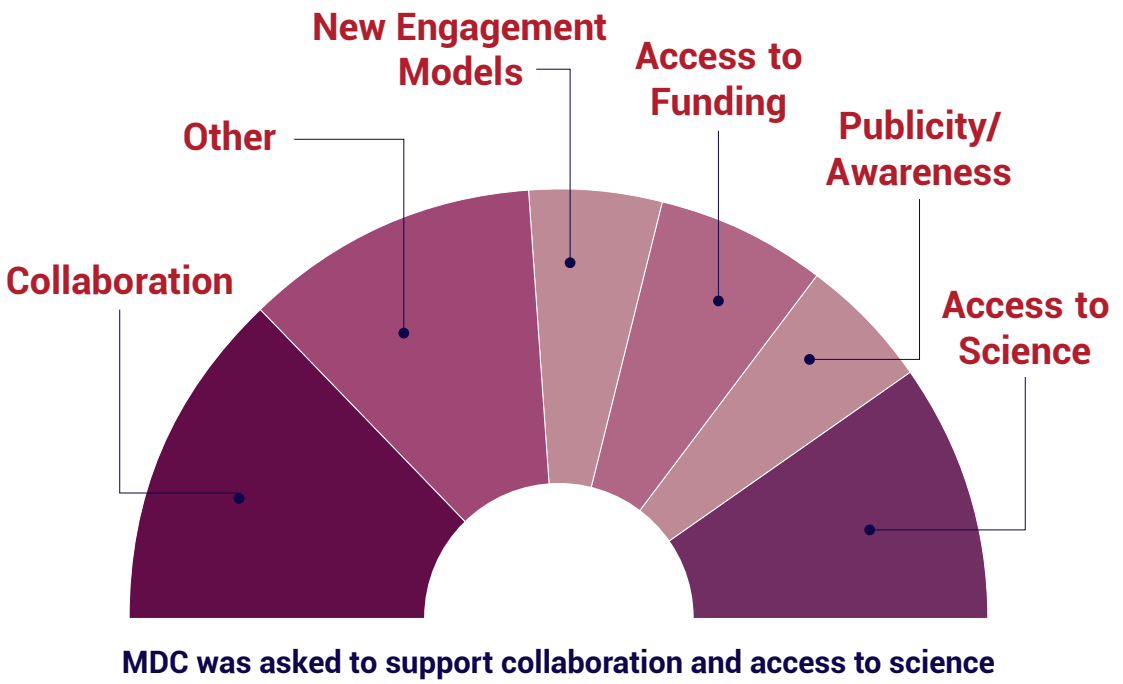


Figure 8

A small number commented that MDC could potentially compete with the supply companies that it is here to support. MDC's goal is to provide a unique combination of capabilities that the community cannot otherwise readily access. MDC has multiple collaborations with supply companies to co-develop new products for them to adopt or access, thereby supporting the continued success of UK CROs.

Broader government support for the sector

75% of recommendations were for government stimulus funding, covering increased grant funding, tax incentives, and more efficient funding mechanisms. The next largest theme ('making connections across the sector') was only raised by 6%, and no other area got more than one mention.

The sector noted concern over the move from ringfenced Health and Life Sciences to 'Smart' Innovate UK competitions, which has decreased the opportunity for SMEs to access medicines discovery funding.

A big pharma view: fund SMEs and their substrate too

When the themes of this report were discussed with large pharma drug discovery developers, one responded with insights into the large company issues around drug development, which is given here to complement the priorities of smaller companies.

"…The field still does not understand the fundamental biology of many diseases well enough to validate drug targets: this will come from increased understanding of disease at the cellular level, both mechanistically and longitudinally and across populations. It will require access to well curated patient cohorts with samples, omics technologies and data management. Cancer has demonstrated the progress that can be made when the cause of a disease (genetic mutations) is identified and becomes affordable to study at massive scale. How can the research community replicate this in cardiac disease or CNS disease by mapping the signatures of disease at a cellular level? The infrastructure and technologies that drive this understanding may not be in small companies and are often better suited to precompetitive, charitable, or government funding, but they are the bedrock that allow innovation and SMEs to start up. These datasets are the input for AI and set the standard for cell models and should not be neglected as a long-term investment…"

AI IN DRUG DISCOVERY

KEY MESSAGES

- AI is now a core part of drug discovery across all areas
- AI is most valuable when applied to a portfolio alongside domain expertise
- The industry wants benchmarks and comparisons between AI systems
- Small companies prefer to sell assets; large companies prefer partnerships
- AI budgets are growing, and pharma companies are moving AI in-house
- Automated data generation or validation is highly valuable
- The right data are key: acquisition and preparation of data take 75% of resources
- Pharma data are valuable but difficult for SMEs to access
- MDC can help by making high-value datasets available and linking the sector

UK AI in the drug discovery sector is in a good position. Increasing end-user (pharma/biotech) acceptance of the value of AI has enabled investment in people, data, and algorithm development, and has led to financial and organisational successes. 90% of UK SMEs are using or need data science in drug discovery. However, there is a trend towards in-sourcing and partnerships that smaller AI and biotech companies will need to navigate. The field is still being held back by the availability of well-curated, large-scale data, and models to make pharma data available on acceptable terms; it is a challenge, but possible. Ways to benchmark the performance of different AI approaches would be welcomed. Small companies are being encouraged to look into pharma's 'pain points' of data curation and document/literature mining, natural language processing, as well as building less specialised interfaces to their products to help a broader group of users. MDC can help both by making datasets available and collaborating to validate SME approaches, and organisationally by increasing customer and producer awareness.

The field is still being held back by the availability of well-curated, large-scale data, and models to make pharma data available

AI user needs

AI is now a core part of SME and large-company drug discovery. While computational chemistry has been using data science and machine learning techniques for years, increased confidence in the field, increased input data quantities, and increased availability of tools and computing resource mean that AI approaches are being used in all areas of modern drug discovery. 90% of UK SMEs needed data science, with half of these specifically needing AI and machine learning. Lead generation and optimisation were the most frequently cited area of data science need followed by Target ID to clinical research. Within the techniques underlying data science, data mining (including AI and machine learning) was most cited, more so than data acquisition, filtering, curation and information extraction from unstructured sources (Figure 9).

AI approaches are being used in all areas of modern drug discovery

90% of UK SMEs needed data science

Areas of data science need

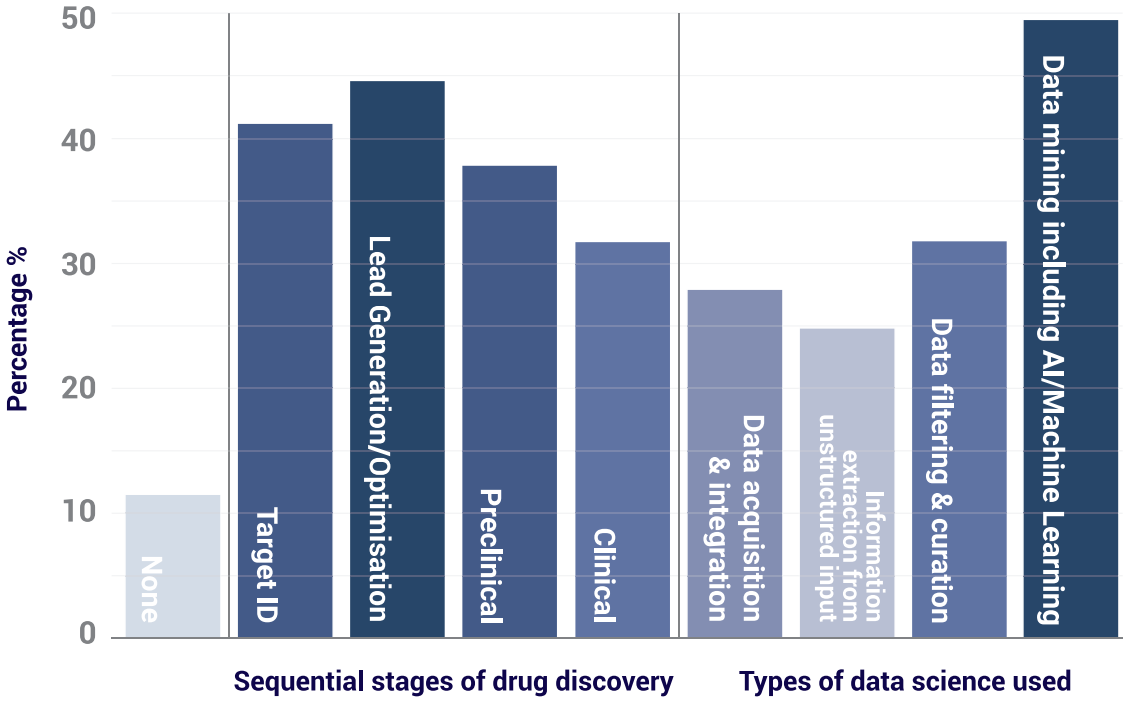


Figure 9

This relative de-prioritisation of data acquisition and integration does not match with the many requests for improved data access, which may indicate that data access is a governance and operational issue for SMEs, not a lack of data science.

AI is now seen as a core part of industrial drug discovery, but measuring impact and validation are seen as highly complex issues

There is a need to improve decision making across target identification and validation, drug discovery and design including absorption, distribution, metabolism and excretion (ADME), efficacy, toxicity, and process improvement. AI is also used for clinical trial optimisation, but this is out of scope for this report.

50% of respondents observed that we do not yet have enough data to prove the later clinical benefit of using AI in the early stages of drug development, due to the time delay and potential confounding factors. Analyses, such as AstraZeneca's 5R paper[14], will start to inform such debate as benchmarking data accumulates. In the interim, validation of AI methodologies (through test datasets, pilot projects, partnerships, expert use, and relevant open competitions) is highly important.

[14] *Morgan P et al. Impact of a five-dimensional framework on R&D productivity at AstraZeneca. Nature Reviews Drug Discovery 2018; 17:167–181*

In the context of uncertain validation, differing views on business models are held. 40% of smaller companies believe that they could create and identify their own in-silico assets and commercialise these to pharma. 50% of larger companies prefer a business model of partnership and close integration or consulting on in-house programmes. Pharma companies preferred to use external SMEs as a technical resource when a technology was unavailable in-house. SMEs were more interested in selling a high-value product e.g., chemical equity to pharma companies.

A continuation of the current mix of fee-for-service and risk-sharing work is expected by all. However, pharma customers and leaders increasingly see AI as a core competency. They are also seeing higher budgets and ease of internal approval for AI programmes, meaning a higher proportion of skills and work brought in-house is likely.

Validation of AI methodologies is highly important

The community is sceptical of black-box algorithms but shows increased interest in models that can inform either the model's own accuracy or the underlying biology that is being represented. Users have more confidence in AI models that give outputs that are interpretable by the scientist, as opposed to outputting a single score. There are real opportunities for companies to create simpler user interfaces and functionality that is accessible to bench scientists without needing an informatician to interpret or guide.

There are real opportunities for companies to create user interfaces and functionality that is accessible to bench scientists

Automation and the value of well-prepared data

Automated synthesis and analysis of molecules designed *in-silico* could enable both rapid validation of the performance of the assets and rapid feedback to improve the algorithms. However, the integration of AI design and automated synthesis in production mode is seen as years away, with current systems at proof-of-concept stage. Any algorithm must be able to show excellent reproducibility across a range of datasets and at large scale as the pharma discovery process is already highly automated and optimised.

Data availability is a key enabler of AI in drug discovery. While there are some AI approaches which make the most of sparse or incomplete data, all data science approaches benefit from increased amounts of relevant, well-curated and linked data. Data curation and preparation accounts for 75% of the budget for data science, with only a small proportion taken in the development and use of the algorithms. There were particular requests for data associated with ADME and toxicity, synthesis prediction, compound activity, and linked genetic-phenotypic datasets. Perceptions of data value have increased over the last five years, and so achieving value for money in data generation, acquisition or sharing is essential.

Data curation and preparation accounts for 75% of the budget for data science

SMEs accessing pharma companies' data

83% of pharma companies felt that the perceived value of their in-house datasets had increased, and for most this was a barrier to sharing their data. IMI[15] initiatives have been set up enabling pharma companies to share data, but most UK SMEs are not part of these initiatives and do not have their own in-house databases to contribute.

There was a willingness in pharma companies to try new models of sharing data that do not involve giving away all of the value or confidential assets within the data. This approach could allow algorithms to be trained on data without the data being exposed to the algorithm developer, for example. This approach was considered to be difficult to implement, but possible, and represented a chance for SMEs to access some of the value in pharma datasets on reasonable terms for both parties. In particular, the sector needs methods of sharing pharma company bioactivity data that do not expose precise biological targets (either explicitly or implicitly) or compound structures, while still retaining enough information to inform machine learning.

[15] *Innovative Medicines Initiative, https://www.imi.europa.eu/*

Issues with valuation of AI

It is difficult to assess the value of an AI approach. Ideally, there is a validation set of data, unseen by the algorithm developer, which the customer can run the algorithm on to assess performance against the standard. While there are open competitions built around this model, most respondents felt that this was not covering the need in drug discovery. This may be due to the highly specialised areas of prediction that AI is now reaching, or the lack of validation data. This is a market failing, preventing customers from independently verifying the performance of their products.

It is difficult to assess the value of an AI approach

Some respondents were sceptical about the value of algorithms alone, noting that the publication and sharing of many AI techniques, in particular deep learning toolkits, mean that the tool is rarely differentiated. However, others were convinced that valuable innovation in algorithms continues. Areas where data science has already seen extensive use, such as molecular modelling, are less innovative for data science and more about data access and appropriate use of the approaches. The less well-covered fields of data extraction and improvement, and natural language processing, were seen as more innovative and highly valuable in curating and enabling the use of public and private scientific and clinical literature.

There is no consensus on certain approaches falling in or out of fashion, although there was increased use of ensemble or multi-algorithmic approaches. The field remains fluid, with different algorithms being used successfully, but with difficulties in comparing results. In addition, AI that helps the user understand the basis or accuracy of its predictions is more valuable, both to the informatic and bench science community.

The UK's position and SME advice for MDC and broader government

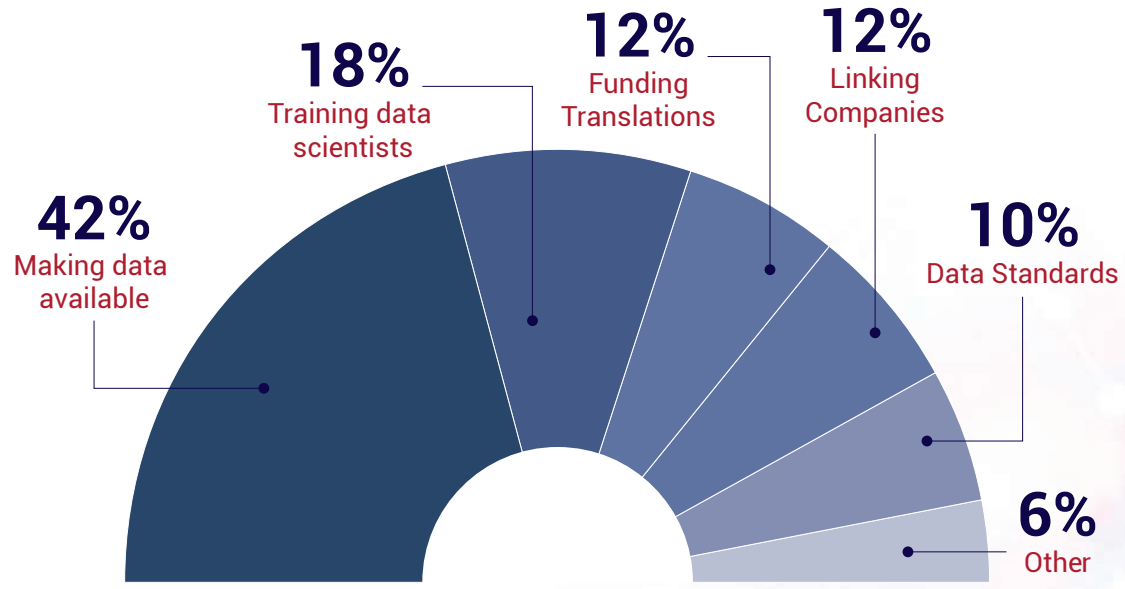
The UK was seen positively in the international context, with a strong mix of data science and drug discovery expertise enabling small and growing AI in drug-discovery companies to compete internationally, despite lower capital availability than in the US.

The success of BenevolentAI, Exscientia and Genomics plc, amongst others, demonstrate that AI in drug-discovery companies can grow and thrive in the UK. AI companies feel more hindered by the lack of available talent and affordable data than by a lack of investment capital. Despite the difficulty for investors in assessing a combination of tech and biotech business fundamentals, this is a 'hot' sector, and the field is considered to be the area of medicines discovery that has most increased in value over the last year.

AI in drug-discovery companies can grow and thrive in the UK

MDC was asked to work on collecting and making available datasets around ADME, toxicity, and linked genetic and phenotypic data

The survey demonstrated that MDC is well known and supported by the AI Drug Discovery community. Collecting and making available datasets around ADME, toxicity, and linked genetic and phenotypic data was identified as a focus area for MDC. The fast changes in the demand and supply for AI in drug discovery led to requests for MDC to organise events and networking to support the community, improve mutual understanding between producers and customers, and matchmake deals nationally and internationally. MDC's role in advising companies on formation and early stage was valued, and collaborations with SMEs have brought additional funds and resources to the sector.



Making data available is the most valued government support for UK data science

Figure 10 — recommendations for support from broader government

Over 40% of respondents want broader government support in increasing data access to support the adoption of AI for drug discovery (Figure 10). Industry also called for 'industrial' training of data scientists and focussed funding for the translation of new academic approaches, suggesting that many of the nation's most impactful interventions only require limited public sector investment on priority areas. This activity, to enable AI in drug discovery, is well suited to be delivered via MDC as a trusted central provider that can leverage both private and public investment.

COMPLEX CELL MODELS

KEY MESSAGES

- CCMs are a valuable preclinical research enhancement, if used correctly
- Improving reproducibility and developing validation data are current priorities
- Adoption is small-scale and experimental, with regulation the enabler for growth
- UK SMEs have potential, reputation and skills to use CCMs well
- Need to validate models and support pre-sales, to bring together the community and advise
- The field is broad, so MDC should prioritise high-risk, high-value areas

CCMs cover a broad field aiming to improve on the predictability of current two-dimensional (2D) and non-human-derived cell models, while managing increased assay difficulty and cost.

CCMs are technological approaches to improve the predictions from cell models, for example by organising the cells into three-dimensional (3D) structures, mimicking differentiated human tissues, approximating fluid composition and flow, and linking different tissue types into a model system.

The use of CCMs is being driven by a combination of decreased trust in the translational value of animal models and increased availability of data to support the validity of complex human cell models. Landmark publications from the Institute of Cancer Research and the Hubrecht Institute[16] have linked the predictions from CCMs with clinical response in patients. There is an increasing awareness of the issues with 2D models, such as cells on flat surfaces being less physiologically relevant or single populations of cells lacking cell-cell interactions. In this context, value comes from appropriate use and understanding the challenges and limitations of CCMs.

Areas of complex cell model needs

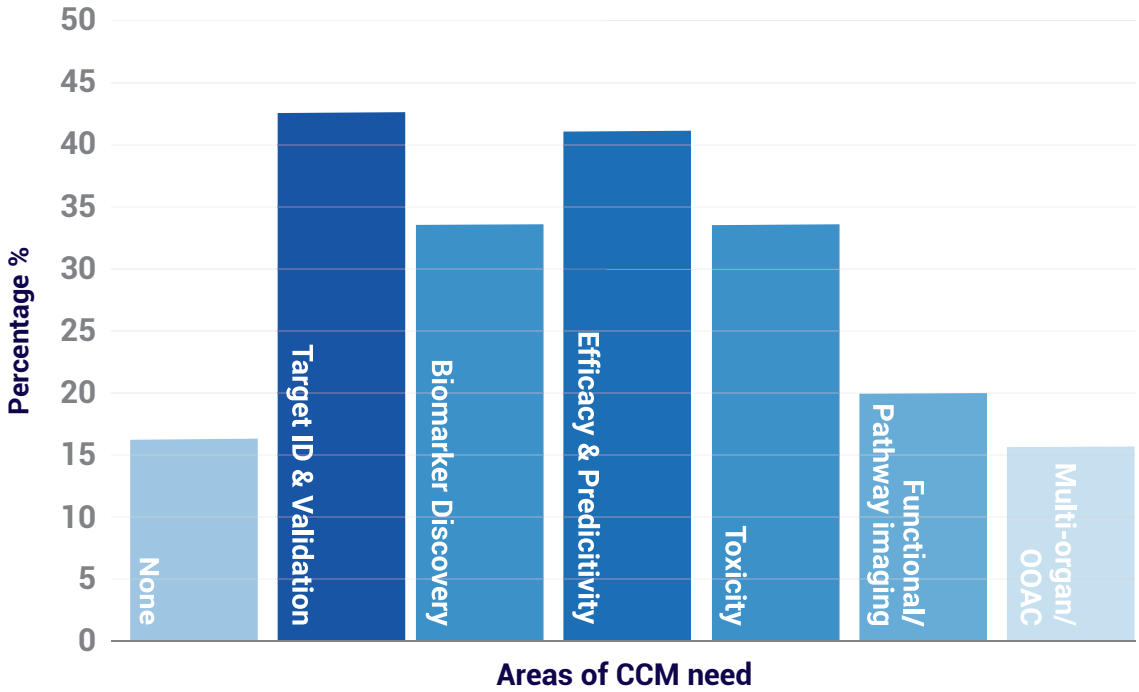


Figure 11

[16] *http://science.sciencemag.org/content/359/6378/920 and https://www.ncbi.nlm.nih.gov/pubmed/29224780*

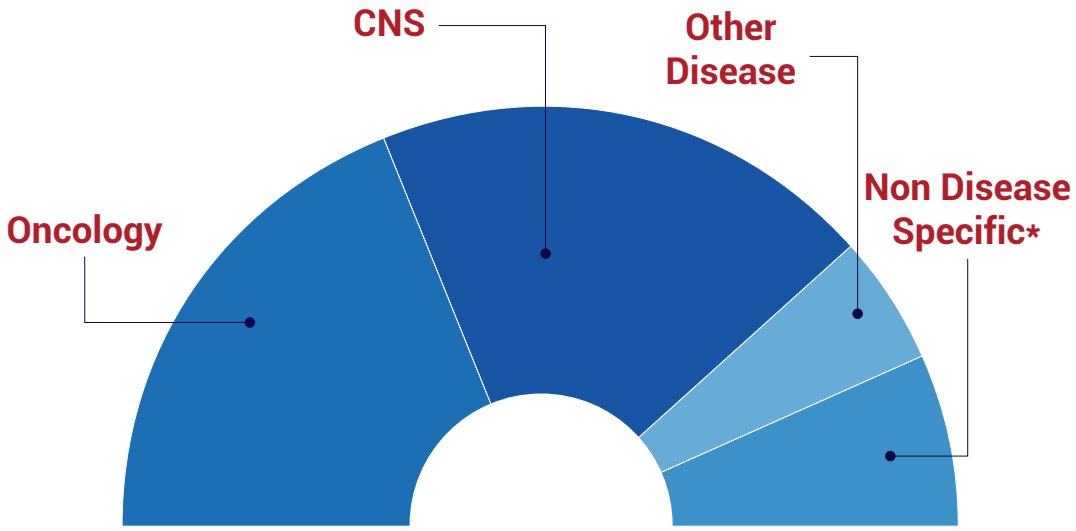
CCMs are needed by the vast majority (84%) of companies for target identification and validation followed by drug efficacy, biomarker validation, and toxicity (Figure 11). There was less need for CCMs in functional and pathway imaging, and for multi-organ or organ-on-a-chip (OOAC) approaches. Interviews identified that OOAC systems require further validation and industrialisation before they can be adopted into widespread use.

Experts producing, commercialising and using CCMs emphasised that the use of these approaches must balance the increased complexity and cost of setting-up, maintaining and interrogating the CCM with the increased accuracy of predicting drug efficacy and safety. Drug discovery scientists need a model that is validated, reproducible, and easy to use, despite the increased complexity. Excellent data skills are needed to manage and interpret the increased and novel outputs from new models.

CCMs are needed by the vast majority (84%) of companies for target identification and validation

Excellent data skills are needed to manage and interpret the increased and novel outputs from new models

Users need a model that is validated, reproducible, and easy to use



Oncology and CNS are seen to be the areas where CCMs could have most impact

Figure 12

*Non disease specific (e.g. Safety/Toxicology, Immunology)

Early areas of commercial interest include toxicity, oncology, immunology, and neurology. CCMs were considered to be most impactful in oncology and CNS disorders (Figure 12). This mirrors the current technological advancement and commercial opportunity for cancer models, and the strengths of CNS-humanised models versus their animal comparators for complex biological diseases.

Current usage is small-scale and experimental, with regulatory support expected to be the major growth driver. Broad innovation continues but customers value usability over novelty, so industrialisation is a key hurdle to overcome. The UK has a good reputation in the field, but companies have received less support than the US. MDC was asked to support companies in demonstrating and validating models in selected areas with high commercial interest that are too risky for company funding, and to link-up the community of users and suppliers.

CCMs user priorities

Identified core needs include: predictivity, reproducibility, validation, and ease of interpretation. The driving need is improved predictivity across the stages of drug discovery, or between the CCM and clinical performance. Quantitative *in-vitro* to *in-vivo* extrapolation should be improved by the use of complex human models. This should enable segmentation and ranking of molecules, and improved understanding of therapeutic margins and *in vivo* potency, as well as adding data to increase the understanding of the disease. Results should be consistent across multiple human donors, consistent over time and reflect the mature human phenotype.

The core needs of users were: predictivity, reproducibility, validation, and ease of interpretation

Although there is no definitive publication or data that 'proves' a model, a reasonable number of widely accepted reference compounds should have been run on multiple occasions with consistent results across molecular and functional endpoints. The current gold standard of validation is checking if the drug that was active or toxic in the clinical setting matches with a signal in the CCM. Ideally, the model's expression profile should match clinical samples, or the CCM should give the correct prediction in situations when the 2D model does not correlate with the clinical result, but this is not commonly done. The validation standard used by CROs is understandably 'good enough' rather than 'beyond all doubt', so a wider initiative is needed to increase the validity of CCMs and to compare across competing models. The costs and associated risks are too high to allow CROs to develop new models from scratch.

A wider initiative is needed to increase the validity of CCMs

3D models can give more information, leading to increased data size and complexity, as well as new data points that do not have equivalence in prior assays. This requires links to advanced analytical approaches for interrogation of models that can measure conventional markers in 3D. Skilled analytical, data and biological scientists are, therefore, needed to interpret novel findings, even when the models achieve the predictivity and reproducibility mentioned before.

Skilled data scientists and biological scientists are needed to interpret novel findings

CCMs are currently better proven in certain areas than others. Early usage was expected in drug safety testing as it is common across all drugs, with cardiovascular and hepatic toxicity as the initial examples. Primary cells or cells derived from inducible pluripotent stem cells (iPSC) can be used as long as the model is reproducible and well-validated for a given end-point. Cancer CCMs have more competition from animal models such as xenografts, but there has been a range of success across the tumour types, and particular interest in modelling the tumour environment. Improved immunological and neurological CCM models are in demand, due to the issues in translating data from existing models to the clinic.

Immunology and neurology CCM models are in demand

Pricing & regulators drive adoption

The development of regulatory guidelines is a main driver of adoption and standardisation of complex assays. Within pharma, there is a trend towards the increasing adoption of CCMs, but this is more at the experimentation stage rather than large-scale use. Broader adoption will be driven by the data that are being requested by regulators, and by reduction in the cost and time needed to validate and adopt a CCM. Pharma have their own standard processes for *in vitro* testing, which will take evidence, time and money to change. It is accepted that CCMs will not replace animals in the next five years, but they should reduce the number of animals being used.

Broader adoption will be driven by the data that are being requested by regulators

Given that preclinical budgets are not increasing, CCM funding will most likely be at the cost of animal budgets. The price of the model is important and current pricing suggests they are being used in the later stages of preclinical drug discovery. However, as the models are not yet seen to be fully industrialised, the current sales by UK SMEs are smaller test projects costing tens of thousands for tens of vials[17].

Precise pricing is complex as the comparison of value is not yet objectively accepted, so pricing follows the current assumptions of value: more than 2D and less than *in-vivo.*

[17]*https://cn-bio.com/innovate-uk-awards-grant-to-cn-bio-innovations-and-astrazeneca-to-use-organs-on-chips-to-evaluate-the-use-of-existing-drugs-to-treat-fatty-liver-disease/*

Innovation in CCMs

There is significant innovation in CCM design and supply, both in the development of new models and in the improvement in performance of current models. Much of this innovation is being carried out within universities, often in collaboration with industry partners. Cost remains important, as well as usability via improved robustness and repeatability. Vendors were felt to have neglected usability in the past, and customer feedback was leading to simplification of the culture process and reducing the time to maturity. Some respondents noted that transfer of organoids between institutions was not yet commonplace, partly through concerns that the protocols would not replicate. CCMs are currently low-to-medium throughput by nature of their complexity, shape, and fluid component, and that scaling-up may not be completely solvable, which will define the point of the process where they are used to evaluate drug candidates.

Gene editing (e.g. CRISPR) is now affordable for CCMs and represents an opportunity for producing disease-like phenotypes. Innovations in 3D-extracellular matrices are being researched by CROs to give increased control and relevance. Cell differentiation and maturation remains a challenge and a focus in areas such as neurology. There was little commercial demand for complex, multi-organ systems, and also significant challenges in delivery of systems combining more than 2-3 organ models. However, OOAC technologies are continuing to mature, and the use of flow is important for a variety of organ types. Far-reaching IP ownership of protocols (e.g., stem cell differentiation and CRISPR) is one of the major blockers to innovation being exploited commercially.

Innovations in matrices and fluid flow can drive up OOAC impact

UK opportunities and calls for MDC action

The UK has CCM expertise of international quality. A combination of robust scientific approaches and an experienced community in models and drug discovery, in general means that the reputation of UK CCMs is high. There are companies in the UK who could take advantage of the unmet need for CCMs, for example iPSC suppliers or CROs, and there are UK academic institutes developing novel complex models. However, other countries have invested strategically in OOAC and cell models, such as the US National Institute of Health (NIH) Tissue Chip for Drug Screening program[18], and the UK will need similar support to remain at the forefront of this technology.

Assure and validate models via demonstration in MDC labs

[18] *https://ncats.nih.gov/tissuechip*

The calls for MDC action were for independent validation and assurance of models *via* demonstration by assessing responses to known molecules. This should be as open as possible - as much of the current work is not published - and it should support the presales process by demonstrating third-party use. The choice of model is important. CROs believe that MDC should focus on areas that are higher-risk projects, or in partnership with companies to jointly reduce the risk. SMEs can identify niche models that currently are not well-served by the UK CRO community. The sector asked MDC to pick a couple of long-term areas in which to develop expertise and capability. Suggestions included oncology and immuno-oncology (needing first to identify specific opportunities), neuroinflammation, neurodegeneration and fibrosis.

Independent validation and assurance of models *via* demonstration by assessing reponses to known molecules

While SMEs asked if MDC could aggregate or make validation sets of compounds for reviewing models, pharma companies said they were willing to share compounds directly for CCM makers to test on their CCMs. Collaborative guidance of the standard controls by tissue type is being published from a group of 18 pharma companies, so coordination with such initiatives will be important.

Pharma companies said they were willing to share compounds directly for CCM makers to test on their CCMs

Most respondents noted that the pace of change within the field made it important for MDC to help create a network of suppliers and users across the country to share information and techniques quickly with smaller companies. This network would include all the related disciplines across academia, pharma and suppliers. In particular, user group meetings that brought lab staff together would be very valuable. Finally, MDC could support network members as an honest broker between model users and suppliers, as well as supplying expert advice to support and critique the experimental approaches and data from CCM companies.

Create a network of suppliers and users across the country to share information and techniques quickly with smaller companies

Broader government recommendations

The request for broader government support for CCMs was focussed on stimulus funding (Figure 13). Over 75% of respondents want specific funding for early-stage companies, academia and translation. The combination of high market demand and lack of funding suggests that CCMs are an area of current market failure, with a key need for validation and funding to translate excellent basic research into industrial adoption.

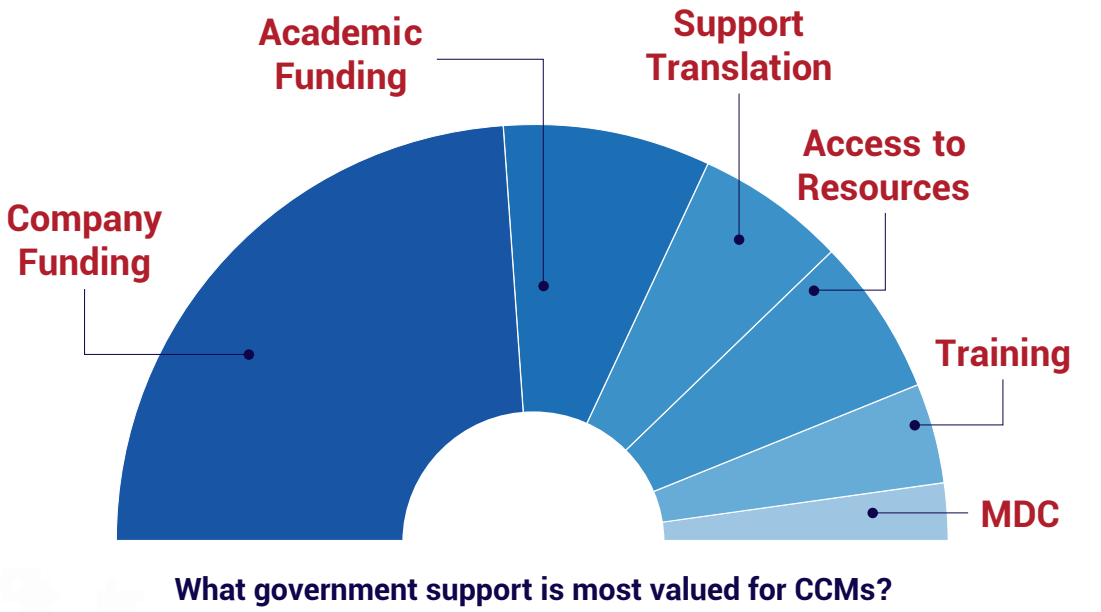


Figure 13

Acknowledgments

Thank you for reading the State of the Discovery Nation 2019. Thanks also to the experts across the UK SME community for their contributions to it. The state of the drug discovery nation is improving and can benefit from continued support from the MDC, BIA, and broader government.

We will continue to conduct research to ensure that we continue listening to the sector and shape MDC's strategy in response to evidence.

Please contact MDC if there is additional evidence that we should be considering. Our ambition is to go beyond incremental change, to transform the industry, and we look forward to working with you to achieve it.

APPENDIX — MARKET REVIEW METHODOLOGIES

Methodology to create the company database

The development of the database of MDC relevant SMEs was performed in four phases: developing the initial list of companies; classifying by sector and segment; supplementary data gathering; and filtering and validation.

The initial list of companies was compiled using data from the OLS 2017 companies list, which was published in May 2018. This list of companies was expanded using data from commercial sources, industry bodies, and Innovate UK, coupled with internal data from the MDC staff and CRM systems.

Once a long list of companies had been produced, an exercise was performed to remove duplicates and to classify new companies based on their primary sector and segment using the same descriptors as OLS. A focussed data-gathering exercise was then performed using Companies House accounts information (e.g. to determine the number of employees, and the last reported status of the company). The company list was then filtered to identify the MDC in-scope companies using the following criteria.

Main Criteria	Examples of Exclusions
Company is an SME based on the number of staff (i.e. fewer than 250 staff)	Large pharma companies
Company is involved in MDC relevant activities	Service and Supply companies involved in logistics, legal etc. Core companies involved in Advanced-therapy medicines or the manufacture of medicines
Active UK registered company (based on data from Companies House)	Dormant, in liquidation, not a going concern, non-UK registered companies.
Company has sufficient (online) data freely available to make an assessment	Newly incepted companies / in stealth mode

A validation review was then performed within the Catapult of final list of companies, and once complete, the list was locked down for a detailed analysis.

In performing this study, it is recognised that this group of companies changes quickly, as commercial and research programmes succeed and fail. Therefore, the results should be treated as a snapshot of the market at a point in time.

Methodology to estimate the number of employees

The number of employees can be used as a proxy for the size of companies where sufficient data exists to make estimates for this.

Companies do not always disclose the number of employees that they have. OLS put employment in ranges due to the proprietary nature of their source data. To develop our estimate of the number of employees across the sector, MDC gathered data from the Annual Report and Accounts for a sample of 12% of companies in order to estimate an average number of employees in each band. These data are presented below and suggest that companies tend to have staff below the midpoint of the band.

Employment Band	No. of companies with known no. of employees	Average no. of employees for that band	No of companies with an estimate no. of employees
0-4	65	2	794
5-9	36	6	181
10-19	24	13	136
20-49	24	29	117
50-99	20	69	46
100-249	15	158	33

Survey details

In 2018, MDC and the BIA repeated their annual joint online survey of the community to understand the opportunities and challenges felt by companies in the sector, and what support they were asking for. The online survey contained detailed sections on two areas of particular interest: AI in drug discovery and CCMs, as well as a repeat of the 2017 questions on perspectives of government support mechanisms. The survey was sent out in collaboration with the BIA and filled in by 60 companies (representing a similar number of companies to the previous report, as last year's survey also included academic and government responses).

The survey population was representative of senior commercial leadership: a quarter were Chief Executive Officers, half were senior R&D, and the remainder mainly Chief Officers or board. One-third of respondents worked in companies active in small molecule drug discovery, and a quarter worked in CROs. In terms of disease focus, one-third were cross-disease, and one-third were spread across oncology, anti-infectives and immunology.

md.catapult.org.uk
info@md.catapult.org.uk
@meddisccat
linkedin.com/company/meddisccat
01625 238734

bioindustry.org
info@bioindustry.org
@BIA_UK

CATAPULT Medicines Discovery

BIA UK BioIndustry Association