The Institution of Engineering and Technology

Artificial intelligence for drug discovery

Leveraging machine learning to accelerate development of new medicines for healthcare crises.

theiet.org/healthcare

Foreword

Dr Philippa Whitford MP
Chair of the All-Party Parliamentary Group on Vaccinations for All.

By sifting through large numbers of potential drugs, active molecules and cellular targets, artificial intelligence (AI) can drastically shorten the development pathway for the discovery of new drugs or the repurposing of existing medication for new applications.

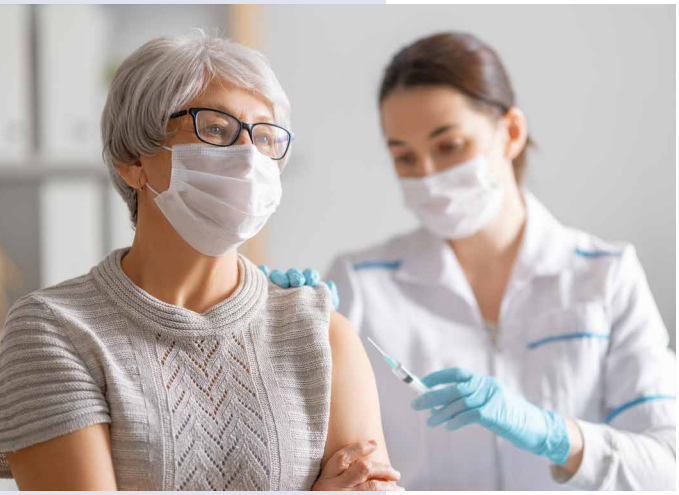
As well as getting new drugs or vaccines to patients within a shorter timescale, it makes research much more cost-effective by selecting therapeutic agents or targets with sufficient promise to justify the more expensive, laborious and time-consuming clinical trial phases of drug development.

This is of particular importance in the drive to find therapeutic options quickly, as has been the case during the Covid-19 pandemic, or when seeking remedies for rare diseases, in which traditional large-scale trials are difficult due to low patient numbers. In both cases, selecting agents with the greatest chance of efficacy is critical to success.

The pandemic has prompted extensive trials of repurposed drugs for use against Covid-19, with developers able to draw on new, accelerated selection methods that use mass screening of compounds enabled by machine-learning tools. The steroid dexamethasone and arthritis medicine tocilizumab are among those approved for emergency use to treat patients with Covid-19 in the UK.

Innovation has been central to the development of Covid-19 vaccines, with mRNA vaccines – also notable for their short production times – used for the first time against the disease. The Pfizer-BioNTech and Moderna jabs, which both have efficacy rates above 90%, are among Covid-19 vaccines to use the novel technique.

The advent of effective mRNA vaccines opens up huge potential to construct tailored vaccines for some of the most pernicious infectious diseases, such as TB and malaria, which cause large-scale morbidity and mortality in low and middle-income countries. Growing antimicrobial resistance (AMR) has helped to spread many such diseases.



The Covid-19 pandemic has set a higher bar for international sharing of scientific journal articles through initiatives such as the Covid-19 Open Research Dataset (CORD-19), and of research tools through portals such as the AI-powered Epitopes.world.

It has also galvanised international collaboration over drug and vaccine development and delivery. Global initiatives include a US$2 billion Covid-19 vaccine development fund led by the Coalition for Epidemic Preparedness Innovation; the World Health Organisation's Solidarity Trial for Covid-19 treatments, and the multi-agency COVID-19 Vaccines Global Access (COVAX) distribution drive. Governments, industry and academia should build on this momentum and deploy recent scientific advances and new collaborative platforms to tackle the ongoing global health threats posed by existing and emerging infectious diseases. The novel technologies highlighted in this report are likely to prove crucial to such efforts.

Introduction

Professor Peter Bannister CEng FIET
IET Healthcare Sector Executive Chair

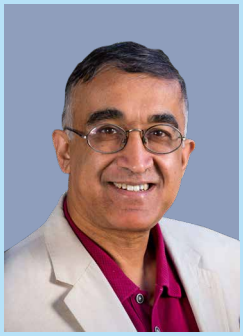
How can artificial intelligence help us to address the central challenges in drug discovery today?

In many ways, drug discovery and artificial intelligence (AI) are the perfect partners. From a data perspective, the requirement in pharma and biotech for the capability to search a vast number of candidate solutions for a match with the desired biological response lends itself naturally to machine learning approaches that can handle the size of the input datasets involved and the requirement for pattern recognition.

For both payer and patient, the advantages on offer are reduced cost and time to identify target molecules, which promise to bring new therapeutics to market in a more responsive manner. Numerous examples now exist of AI being leveraged in the discovery of new therapeutics including vaccines, with the latter addressing an immediate need brought on by the Covid-19 outbreak. While novel solutions are being sought, there are also many good examples of existing biotech and pharma solutions being repurposed, for example to address the challenge posed by rare diseases.

Despite rapid and positive growth in this space, there are still few – if any – drugs to date that have completed the full development pathway, after being initially identified using AI. Demonstrating the true value of these discovery approaches in terms of improved patient outcomes will require a coordinated approach between in-silico methods and established pre-clinical and clinical testing protocols, necessitating a multi-disciplinary effort. It will be exciting to track many of the innovations highlighted in this report over the next few years.

Professor Peter Bannister CEng FIET
IET Healthcare Sector Executive Chair



Dr Ronjon Nag
Interdisciplinary Distinguished Careers Institute Fellow, Stanford University, and Founder, R42 Institute

Which AI techniques have the greatest potential to accelerate drug discovery?

Knowledge graphs are the result of algorithms combing through millions of biomedical articles, analysing their content.

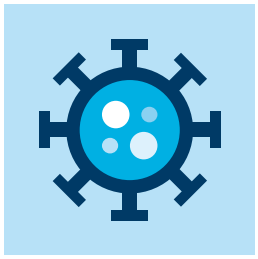
If molecule A is shown to affect molecule C in one article, and molecule B is shown to affect molecule C in another, then the knowledge graph may suggest a relationship between A and B. Other new information sources include patient group data, which indicates the effects of various treatments on diseases, and a plethora of genomic databases. By combining analyses of this data, a set of biomarkers can be derived, against which existing drugs or new molecular leads can be reviewed for potential effectiveness.

Dr Ronjon Nag
Interdisciplinary Distinguished Careers Institute Fellow, Stanford University, and Founder, R42 Institute

Artificial intelligence for drug discovery is published by the Institution of Engineering and Technology. Please note that the views expressed in this publication are not necessarily those of the IET. It is not intended to be a guidance note with a specified set of recommendations or actions but rather seeks to add understanding and debate around the topic.

Unblocking the drug pipeline

Data scientists are unleashing ultra-fast algorithms on vast chemical and biological databases to predict treatments for some of the world's most pressing health problems. Here are some key questions they are helping to answer.



How is AI being deployed for the rapid development of drugs to treat Covid-19?

The race to find a treatment for Covid-19 has highlighted the potential of machine learning for rapidly identifying new medical uses for existing drugs, known as drug repurposing. Hundreds of clinical trials of existing drugs to treat Covid-19 patients have been launched[1], with the Ebola anti-viral remdesivir[2] and steroid dexamethasone[3] the first to be approved for use in the UK. The application of AI techniques to the vast quantities of medical data now available has enabled candidates for repurposing to be rapidly predicted[4]. The approach benefits from the far shorter trials process for existing drugs.

How can we harness AI to step up production of new drug candidates?

Slow, expensive and high-risk, drug discovery brings pharmaceutical companies such low returns that development of some vital new medicines has virtually stalled[5]. However, advances in genomics have enabled a faster reverse pharmacology in which multiple drug candidates are screened against isolated biological targets planted in synthesised proteins, prior to live screening[6]. A new generation of machine learning tools that aim to predict potential new drugs through feature pre-processing of chemical and physiological databases accelerates the process further through mass screening in silico[7]. It could revolutionise the industry by slashing development time and costs.

Can AI accelerate the development of vaccines for new contagious diseases?

More than half the candidate Covid-19 vaccines listed by the World Health Organization are based on a DNA, RNA or protein-fragment platform[8]. These novel candidates can be rapidly developed using digital molecular libraries and machine learning tools such as generative adversarial networks[9]. For protein-fragment candidates, immunologists have collected nearly a million fragments of virus proteins that can be targeted by antibodies or T-cells[10]. Using such libraries, researchers have trained AI models to predict which fragments of the SARS-CoV-2 spike protein are likely to be detected by the immune system and thus serve as potential vaccines.

How can data science be mobilised to adapt clinical trials for pandemic conditions?

Measuring success in clinical trials relies on statistical analysis and repeated sampling, which is often a lengthy process. Yet in a pandemic, effective remedies are needed urgently. Researchers have turned to Bayesian statistics to design adaptive trials that aim to match pandemic conditions. The Bayesian approach allows remedies that are unsuccessful to be dropped from parallel clinical trials, while other arms can continue uninterrupted. Sample sizes and probability thresholds for an effective treatment can also be varied to match altered medical priorities[11]. A Bayesian adaptive framework for regulators has also been devised[12].

1 https://www.covid-nma.com/dataviz/
2 https://www.ft.com/content/387b544d-8354-46a9-9d1b-7005fc9714a2; https://www.labiotech.eu/medical/covid-19-remdesivir-europe/
3 https://www.gov.uk/government/news/world-first-coronavirus-treatment-approved-for-nhs-use-by-government
4 https://www.ft.com/content/877b8752-6847-11ea-a6ac-9122541af204
5 https://pubmed.ncbi.nlm.nih.gov/20130567/; https://dukespace.lib.duke.edu/dspace/bitstream/handle/10161/12742/DiMasi-Grabowski-Hansen-RnD-JHE-2016.pdf;sequence=1 ; https://www.scientificamerican.com/article/a-dearth-of-new-meds/
6 https://bjui-journals.onlinelibrary.wiley.com/doi/full/10.1111/j.1464-410X.2001.00112.x
7 https://www.cell.com/trends/pharmacological-sciences/fulltext/S0165-6147%2819%2930135-X
8 https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines
9 https://mc.ai/real-world-deep-neural-network-architectures-for-pharma-industry/
10 https://www.brookings.edu/techstream/can-artificial-intelligence-help-us-design-vaccines/
11 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3495977/
12 https://hdsr.mitpress.mit.edu/pub/non4rfk6/release/1

AI gives developers the edge in race for new remedies

Artificial intelligence is playing a transformative role in medicine and drug discovery. The sheer size of the molecule libraries used to screen for new drug candidates means it is now practically impossible for individual researchers to review everything themselves. It is therefore not surprising that researchers and developers are now looking to the unparalleled data-processing potential of AI systems to make drug discovery a faster and less costly process. Here are some examples.

AI pioneer targets rare diseases

Cambridge firm ramps up drug hunt through in silico screening

British scale-up Healx is pioneering an AI-based system for predicting new drug candidates to treat rare diseases.

There are some 7,000 rare diseases, which affect more than 400 million people worldwide, but only five percent of them have approved treatments. The conditions typically have too few patients for traditional drug development to be profitable.

Healx tackles the problem by using repurposed drugs – alone or in combination - to cut out three stages of the traditional eight-stage process. It aims to reduce the average development period from 12 years to five and the cost from around $2-3 billion to $70 million[13].

Its Healnet platform takes advantage of the explosion of medical data over recent decades, including millions of additional journal articles and genomes, and data from online patient groups, as well as advances in machine learning for biochemistry.

Healnet uses cutting-edge AI tools including a knowledge graph to conduct "massively parallel, hypothesis free" analysis of existing drugs against rare diseases to predict drug candidates for multiple diseases at once[14].

The predictions are then assessed by expert drug hunters and promising candidates are put forward for late-stage clinical trials. Healx aims to get 100 potential drugs into trials by 2025.

Oxford takes on "camouflage" bacteria

University spin-off aims to crack antibiotic resistance problem

The latest AI techniques are being mobilised by Oxford Drug Design (ODD) to tackle the global health threat of growing antibiotic resistance.

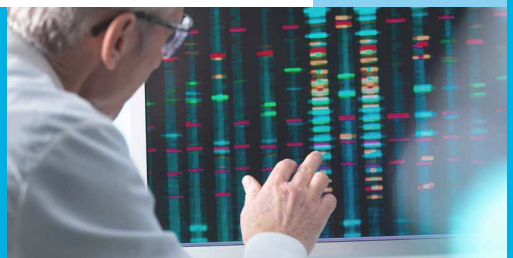
The UK medtech firm has come up with new drug candidates aimed at combating a range of Gram-negative bacteria that are deemed to pose a critical threat by the World Health Organization[15].

The organisms have a slime layer that acts as camouflage against detection by the human immune system and no new antibiotics have been produced to combat them for more than 50 years.

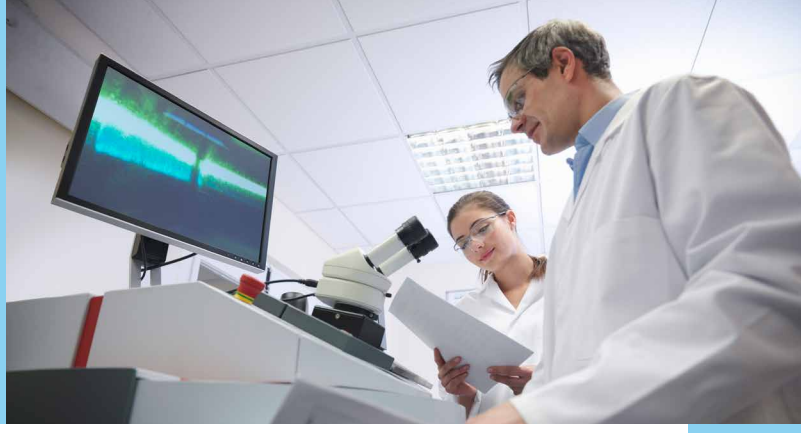
The aminoacyl tRNA Synthetase Inhibitors (aaRSIs) are a new chemical class with a low propensity for developing resistance and have been found to be active in vitro against Gram-negative bacteria including Enterobacteriaceae and Acinetobacter baumannii[16].

ODD's computer-aided drug discovery system builds on recent advances in the biochemistry of proteins. It uses machine learning and topological data analysis to predict the action of drug candidates on protein targets.

The in silico screening is performed on a proprietary database of interactions between 250 molecules and 16 proteins that was built over 15 years and began with a screen-saver grid computing project involving 3.5 million users.



[13] https://pubmed.ncbi.nlm.nih.gov/14693466/
[14] https://healx.io/healnet/
[15] https://www.who.int/medicines/publications/WHO-PPL-Short_Summary_25Feb-ET_NM_WHO.pdf
[16] https://oxforddrugdesign.com/pipeline/



Post-docs open door to faster vaccine search

AI-powered portal predicts Covid-19 vaccine targets

An innovative, AI-powered, open-source research portal has been launched to help accelerate worldwide vaccine development for Covid-19[17].

Epitopes.world uses an algorithm called CAMAP that can predict which parts of the SARS-CoV-2 virus are more likely to be epitopes, the part of an antigen that is presented by a virus at the surface of an infected cell.

Epitopes can be recognised by T-cells in the human immune system, which will mount a virus-specific response to them and proceed to destroy the infected cell.

The predictions can be used by researchers to generate a significantly shorter list of potential biological targets to test in the search for a vaccine, reducing a process that typically takes months or weeks to less than a day.

CAMAP is freely available on the platform for use by researchers anywhere in the world. The tool was developed by a multi-disciplinary team of post-doctoral researchers led by Dr Tariq Daouda of Harvard Medical School. The site's open-source technology also enables researchers to collaborate and share results.

More than 82,000 epitopes and related metadata, including 39,000 from SARS-CoV-2, 39,000 from SARS-CoV-1 and 104,000 from normal human sequences for comparison, are stored in the portal's back-end.

Speeding up the drug pipeline

Birmingham trials use agile design for faster results

Researchers at Birmingham University have adopted a Bayesian adaptive design to expedite early-stage clinical trials of drugs for treating patients with severe Covid-19 symptoms.

The Phase II trials are among four CATALYST clinical studies in progress across the UK that aim to identify promising candidates for RECOVERY, the national Phase III trials programme that assesses drugs for use against SARS-CoV-2.

The Birmingham trials are testing Namilumab, a human monoclonal antibody, and Infliximab, which is used to treat inflammatory conditions, against a control group and have seven more candidate drugs in the pipeline[18].

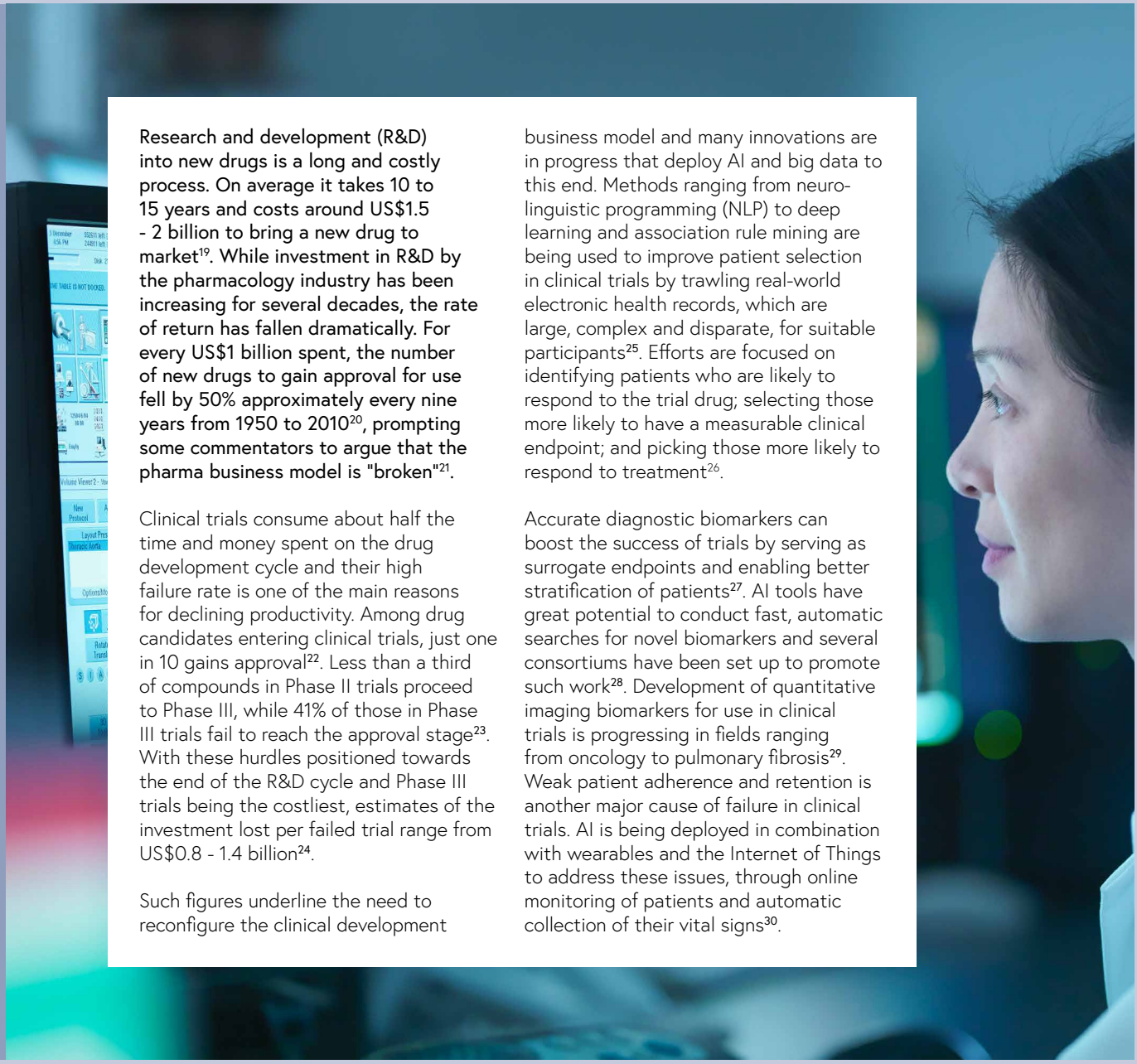
The three trial groups will each include between 40 and 60 hospitalised patients with symptoms of Covid-19 pneumonia, who are treated for 14 days and then tested using blood biomarkers.

The Bayesian design allows the research strategy to be modified as the trials progress using a priori information collected at each stage, enabling very effective drugs to be identified more rapidly.

Results in each group are assessed once 40 patients have been tested. If the drug shows a positive effect, it is recommended for inclusion in RECOVERY, while those with negative results are dropped and replaced.

[17] https://www.prweb.com/releases/team_of_phd_researchers_unveil_ai_powered_platform_to_open_source_covid_19_vaccine_development/prweb17097742.htm
[18] https://www.birmingham.ac.uk/news/latest/2020/06/catalyst-trial-drugs-announced.aspx

Addressing the wider productivity issue in pharma research and development



Research and development (R&D) into new drugs is a long and costly process. On average it takes 10 to 15 years and costs around US$1.5 - 2 billion to bring a new drug to market[19]. While investment in R&D by the pharmacology industry has been increasing for several decades, the rate of return has fallen dramatically. For every US$1 billion spent, the number of new drugs to gain approval for use fell by 50% approximately every nine years from 1950 to 2010[20], prompting some commentators to argue that the pharma business model is "broken"[21].

Clinical trials consume about half the time and money spent on the drug development cycle and their high failure rate is one of the main reasons for declining productivity. Among drug candidates entering clinical trials, just one in 10 gains approval[22]. Less than a third of compounds in Phase II trials proceed to Phase III, while 41% of those in Phase III trials fail to reach the approval stage[23]. With these hurdles positioned towards the end of the R&D cycle and Phase III trials being the costliest, estimates of the investment lost per failed trial range from US$0.8 - 1.4 billion[24].

Such figures underline the need to reconfigure the clinical development business model and many innovations are in progress that deploy AI and big data to this end. Methods ranging from neuro-linguistic programming (NLP) to deep learning and association rule mining are being used to improve patient selection in clinical trials by trawling real-world electronic health records, which are large, complex and disparate, for suitable participants[25]. Efforts are focused on identifying patients who are likely to respond to the trial drug; selecting those more likely to have a measurable clinical endpoint; and picking those more likely to respond to treatment[26].

Accurate diagnostic biomarkers can boost the success of trials by serving as surrogate endpoints and enabling better stratification of patients[27]. AI tools have great potential to conduct fast, automatic searches for novel biomarkers and several consortiums have been set up to promote such work[28]. Development of quantitative imaging biomarkers for use in clinical trials is progressing in fields ranging from oncology to pulmonary fibrosis[29]. Weak patient adherence and retention is another major cause of failure in clinical trials. AI is being deployed in combination with wearables and the Internet of Things to address these issues, through online monitoring of patients and automatic collection of their vital signs[30].

[19] https://www.sciencedirect.com/science/article/pii/S0165614719301300
[20] https://www.bio.org/press-release/bio-releases-largest-study-ever-clinical-development-success-rates
[21] https://endpts.com/pharmas-broken-business-model-an-industry-on-the-brink-of-terminal-decline/
[22] https://www.bio.org/sites/default/files/legacy/bioorg/docs/Clinical%20Development%20Success%20Rates%202006-2015%20-%20BIO,%20Biomedtracker,%20Amplion%202016.pdf
[23] https://academic.oup.com/biostatistics/article/20/2/273/4817524?login=true
[24] https://www.clinicalleader.com/doc/the-high-price-of-failed-clinical-trials-time-to-rethink-the-model-0001
[25] https://pubmed.ncbi.nlm.nih.gov/31218278/
[26] https://www.fda.gov/regulatory-information/search-fda-guidance-documents/enrichment-strategies-clinical-trials-support-approval-human-drugs-and-biological-products
[27] https://www.sciencedirect.com/science/article/pii/S0165614719301300
[28] https://fnih.org/our-programs/biomarkers-consortium; https://prevention.cancer.gov/major-programs/consortium-imaging-and-biomarkers
[29] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4717912/; https://pubmed.ncbi.nlm.nih.gov/29986154/
[30] https://www.ibm.com/blogs/research/2017/04/monitoring-parkinsons-disease/; https://aicure.com

The next steps

1. Data

- Extend or replicate open-access scientific literature databases on Covid-19 for future pharmaceutical research into other infectious diseases.
- Develop global standards for electronic health records and tackle the issue of incompatible data from different countries.
- Set up an open-access data portal for worldwide efforts to address the antimicrobial resistance problem.

2. Research

- Retain global collaborative networks for medical research into Covid-19 and extend them to drug discovery for bacterial superbugs and other infectious diseases.
- Introduce legislation allowing researchers to use anonymised health data from patient groups and wearables for drug discovery, with requirement to delete it after a year.
- Promote community grid computing initiatives for drug discovery.

3. Clinical trials

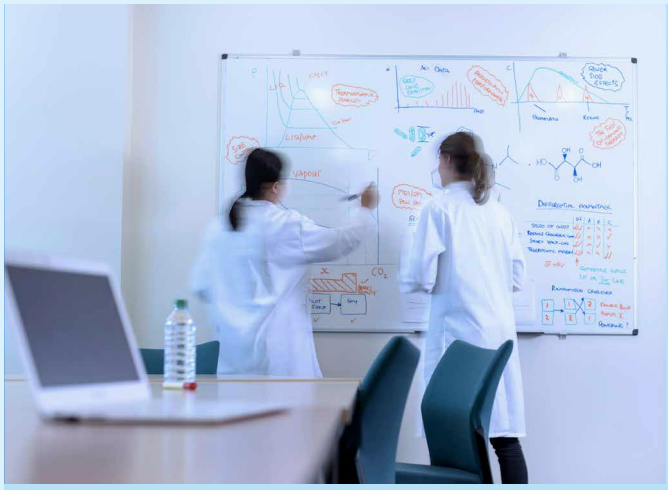
- Adopt parallel regulation of clinical trials used in the Covid-19 pandemic for all high-probability clinical trials.
- Review Bayesian adaptive trials of Covid-19 drugs to assess and refine design strategies for future use.
- Retain national clinical trials programmes set up for Covid-19 treatments and use them for developing drugs against other infectious diseases.

4. Business models

- Set up open collaborative initiatives between pharmaceuticals companies, entrepreneurs and universities to share publicly funded research with entrepreneurs on standard terms for faster development of new medicines against target diseases.
- Encourage health authorities to adopt Netflix-style, subscription-based payment contracts with drug companies to develop and supply medicines for problem diseases.

5. Finance

- Governments should support public-private-philanthropic accelerators as the new paradigm for financing drug and vaccine discovery for infectious diseases.
- The pharmaceuticals industry should jointly fund early-stage drug discovery by start-ups through initiatives like the AMR Action Fund.



We say...

The global healthcare burden is growing, due to new infectious diseases, rising antimicrobial resistance and ageing populations, while drug development for many diseases has become uneconomic. Urgent action is required to equip the pharmaceutical industry with the AI tools, business models and finance it needs to meet today's challenges. The following are our recommendations to support the acceleration of AI for drug discovery and the reduction of attrition rates, ultimately making more novel drugs available to citizens, in a faster and cheaper way.

1. **Set up a Global Data Warehouse to support drug development for infectious diseases.**

   The virtual storehouse for open-source biomedical data and electronic health records should be headed by a data czar responsible for setting data standards and guidelines for patient anonymity.

2. **Establish a National Institute for AI and Drug Discovery to lead on development of machine learning technologies for pharmaceuticals and dissemination of knowledge.**

   Multidisciplinary teams from data science, biochemistry, genomics and medicine could develop methods and protocols for in silico drug design and its applications across the industry.

3. **Launch high-profile competitions with large "cash for equity" prizes for discovery of new drugs for incurable diseases.**

   New awards inspired by the British government's £10 million Longitude Prize would incentivise start-ups to produce initially uneconomic remedies against diseases posing a global or regional health threat.

4. **Set up a volunteer National AI Corps of data scientists to tackle urgent drug development issues in future pandemics.**

   Expert teams drawn from across industry could be mobilised to address international and domestic health emergencies by deploying AI techniques for rapid drug and vaccine development.

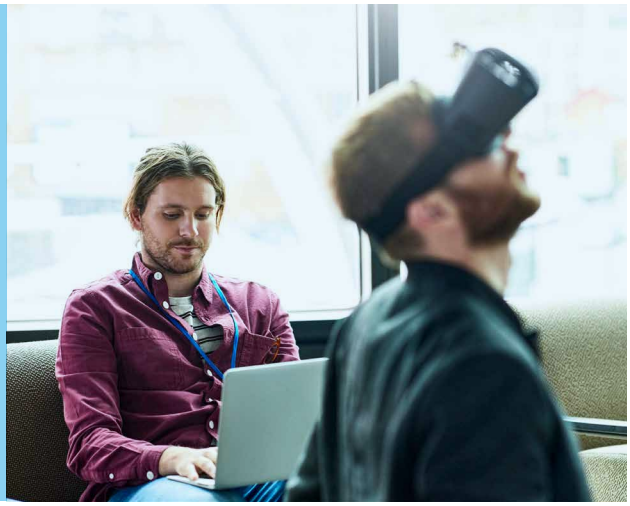
5. **Create a Pandemic World Bank to finance research into new antibiotics and other unprofitable drugs for infectious diseases.**

   The bank would provide regular, subscription-based funding and one-off incentives for pharmaceutical companies, global consortia and regional public-private partnerships to tackle global health threats and neglected diseases.

We are the Institution of Engineering and Technology (The IET), a charitable engineering institution.

Passionate about healthcare? We bring together expert practitioners from the healthcare industry, academia, the public and third sectors. Your specialist healthcare knowledge can inspire others and make a difference.

If you would like to get in touch, please contact us at **sep@theiet.org**



@TheIET theiet.org

E6D21002/Healthcare/0621