Please cite this article in press as: Jiang et al., Artificial intelligence and automation to power the future of chemistry, Cell Reports Physical Science (2024), https://doi.org/10.1016/j.xcrp.2024.102049

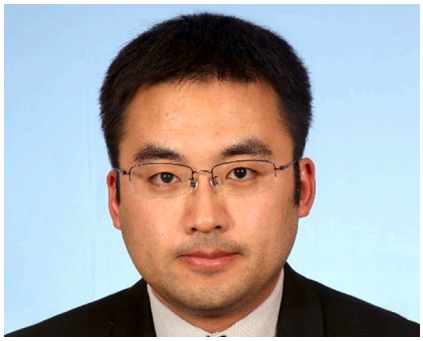
Cell Reports
Physical Science

CellPress
OPEN ACCESS

Voices

Artificial intelligence and automation to power the future of chemistry

In our traditional impression of chemical laboratories, researchers wear white coats and safety goggles to conduct experiments. However, many recent developments in the field make use of autonomous synthesis robots with integrated artificial intelligence (AI)-driven machine-learning units. These benchtop devices might outperform human chemists in terms of speed and accuracy, which could accelerate the discovery of molecules and materials for various applications. In this Voices piece, we ask a panel of experts from institutes in China: How are AI and automation shaping the future of chemistry?



Xuefeng Jiang
Shanghai Key Laboratory of Green Chemistry and Chemical Process, Institute of Eco-Chongming, School of Chemistry and Molecular Engineering, East China Normal University
State Key Laboratory of Organometallic Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences
State Key Laboratory of Elemento-organic Chemistry, Nankai University

Unlocking chemistry's future: Artificial intelligence-driven instrumentation revolutionizes discovery

Artificial intelligence (AI) is set to revolutionize the field of chemistry, offering unprecedented opportunities for innovation and efficiency. The importance of instrumentation and hardware in AI-driven chemistry is profound, because they serve as the critical link between theoretical models and real-world applications. Advanced instrumentation, such as high-throughput (HT) screening devices, spectrometers, and chromatography systems, enables the collection of vast datasets that are essential for training and validating AI algorithms. Moreover, the integration of robotics and automation hardware in chemical laboratories allows for the execution of complex experiments with precision and speed, further enhancing the capabilities of AI in predicting chemical reactions, designing novel materials, and optimizing synthesis pathways. Consequently, the synergy between state-of-the-art instrumentation and AI is driving revolutionary advancements in chemistry, leading to faster discovery cycles, reduced costs, and a deeper understanding of chemical phenomena.

In the pursuit of advancing AI-powered chemistry, my research has focused on the design and optimization of cutting-edge scientific instruments and hardware. These instruments play a critical role in enabling the seamless integration of AI algorithms into chemical research workflows, facilitating rapid experimentation and big data collection/analysis. For instance, our work has led to the development of AI-driven robotic synthesis platforms and HT screening systems. These platforms automate chemical synthesis processes and enable the screening of vast chemical libraries with unprecedented speed and efficiency. By coupling AI algorithms with state-of-the-art instrumentation, we can accelerate the discovery of novel compounds and materials with tailored properties for various applications, ranging from drug discovery to materials science.

Moreover, AI-powered instrumentation enhances the accuracy and reliability of chemical analysis. By leveraging AI algorithms to analyze complex spectroscopic data generated by instruments such as mass spectrometers and NMR spectrometers, we can extract valuable insights into molecular structures and chemical reactions. This not only accelerates the pace of scientific discovery but also enables researchers to gain a deeper understanding of chemical systems and phenomena. Furthermore, AI-driven data analysis tools facilitate the interpretation of experimental results and guide future research directions. By automating the analysis of large datasets and identifying correlations and patterns, AI algorithms enable researchers to make informed decisions and optimize experimental parameters more effectively.

As we look toward the future, the synergy between AI and advanced scientific instrumentation will continue to drive innovation in chemistry. By further advancing

Please cite this article in press as: Jiang et al., Artificial intelligence and automation to power the future of chemistry, Cell Reports Physical Science (2024), https://doi.org/10.1016/j.xcrp.2024.102049

the development of AI-powered instruments and hardware, we can unlock new frontiers in chemical research and address some of the most pressing challenges facing society, such as drug discovery, renewable energy, and environmental sustainability.

In conclusion, the integration of AI with advanced scientific instrumentation and hardware is poised to power the future of chemistry. By harnessing the collective power of AI and cutting-edge instrumentation, we can accelerate scientific discovery, drive innovation, and pave the way for a cleaner, healthier, and more sustainable world.

Sanzhong Luo
Center of Basic Molecular Sciences (CBMS), Department of Chemistry, Tsinghua University

Advancements and challenges in AI for synthetic chemistry

AI is rapidly advancing in the field of synthetic chemistry, with applications in automatic data extraction, property and reaction prediction, retrosynthetic analysis, and autonomous synthesis. The integration of AI into synthetic chemistry has the potential to revolutionize the way research is conducted in the coming decades. One promising aspect of AI in synthetic chemistry is the concept of "synthetic copilot." These intelligent copilots can collect and summarize the latest publications tailored to researchers' interests. Furthermore, they can provide researchers with suggestions for new research directions based on the most recent findings. They can also assist chemists in exploring new ideas, such as designing novel molecules and synthetic pathways. In the laboratory, the embodied assistants can perform experiments while online analysis instruments automatically detect and analyze the results. Undoubtedly, all the processes will be meticulously supervised and guided by the chemist's profound expertise, creative ideation, imaginative thinking, and chemical intuition.

However, despite these advancements, several challenges must be addressed to realize the full potential of AI in synthetic chemistry. One primary challenge is the availability of high-quality synthetic databases for training AI models. Although data-extraction algorithms exist, they have not yet achieved the level of complexity required for parsing intricate chemical semantics. Leveraging large language models (LLMs) to collect chemical information and transfer it into a structured database is a highly valuable goal. Additionally, developing multi-modal models capable of extracting illustrations and molecular structures is essential for building a diverse and informative database. High-throughput experimentation (HTE) also offers valuable data for model training.

Chemical feature engineering is another crucial aspect of AI technology in synthetic chemistry. It involves mapping molecules and their correlations to abstract mathematical structures known as chemical representations. However, achieving a loss-free mapping is challenging. Researchers focus on specific properties and computable parameters of molecules to develop descriptors such as molecule fingerprints, chemical languages (e.g., SMILES and InChI), molecule graphs, and point clouds. Although significant progress has been made, challenges remain in developing descriptors for chiral molecules and accurately representing complex chemical reactions and their intricate correlations among reactants, catalysts, solvents, and reaction conditions. Collaboration between computer scientists and chemists is essential for designing new chemical representations and reaction networks that address these challenges.

The synergy among AI models, reaction data, and experimental practices is transforming synthetic chemistry. Diverse AI models empowered reaction optimization, catalyst discovery, and compound synthesis, enhancing efficiency and accuracy. However, challenges persist in terms of data scarcity, algorithm interpretability, robust experimental validation, accurate detection and analysis of reaction results,

Please cite this article in press as: Jiang et al., Artificial intelligence and automation to power the future of chemistry, Cell Reports Physical Science (2024), https://doi.org/10.1016/j.xcrp.2024.102049

and the capture of the complexity of chemical systems. Integrating AI with robotics and automation could revolutionize experimental workflows. Innovative strategies such as active learning and Bayesian optimization can significantly accelerate the cycle of new reaction discovery and optimization, paving the way for the autonomous laboratory of the future.

Artificial general intelligence (AGI), particularly LLMs, has made substantial contributions to AI in chemistry. AGI can comprehend chemical knowledge, extract and analyze chemical texts, and predict chemical properties or reactivity. Additionally, it can assist researchers in planning and conducting experiments, reducing time and costs. AI chemistry assistants powered by LLMs can reason about complex chemistry problems by connecting to various chemical tools and databases. However, challenges related to reliability and professional competence need to be addressed for wider application in chemical research. As AGI technology progresses, it holds the potential to become more effective in complex chemical tasks, transcending its role as a "bridge between humans and machines."

In conclusion, AI has the potential to significantly impact synthetic chemistry by making various aspects of the research process more intelligent. Although there are challenges to overcome, collaborative efforts among interdisciplinary disciplines can lead to innovative solutions. The integration of high-quality data, advanced molecular feature engineering, autonomous laboratory technologies, and the application of AGI throughout the entire research pipeline will bring in a "Cambrian period" for synthetic chemistry.

Kuangbiao Liao
Guangzhou Laboratory

Harnessing HTE to empower AI-driven synthetic chemistry

In 1828, the German chemist Wöhler achieved a milestone with the synthesis of urea, marking the inception of synthetic chemistry. Over the ensuing two centuries, synthetic chemistry has been a cornerstone of scientific progress, significantly shaping human society. The outcomes of synthesis are typically influenced by a multitude of variables, such as reaction conditions and substrate structures. Chemists aim to elucidate the complex relationship between these variables and outcomes, yet the vast reaction space poses a significant challenge. Although human instinct and intuition have driven the development of synthetic chemistry, it is essentially an experiment-based science. Synthetic chemistry has long been associated with inefficiencies and laborious processes. For instance, the optimization of reactions, discovery of new reactions, and the synthesis of complex natural products often demand months or even years of meticulous work from chemists.

Fast forward to Synthesis 4.0, propelled by the fourth industrial revolution, which integrates cutting-edge technologies such as automation, data science, and AI into chemistry. Embracing this paradigm shift, the chemistry community has begun developing AI chemistry to address pertinent chemical challenges. AI is ubiquitous today, powering everything from personalized recommendations to autonomous driving and drug discovery. It has emerged as an effective tool to identify hidden patterns in data. Synthetic chemistry, to some extent, can entail pattern recognition for the construction of target molecules. To enhance understanding and broaden synthetic reaction applications, there has been a surge in interest among chemists in developing AI-based models for reaction prediction. The reaction dataset, comprising substrate structures, reaction conditions, and outcomes (yield or selectivity), forms the cornerstone for developing AI-based models. Although vast amounts of data have been accumulated, issues such as bias toward positive data, data inconsistency, and lack of annotation plague both public and proprietary databases. Consequently, several AI-based models for reaction prediction have been

Please cite this article in press as: Jiang et al., Artificial intelligence and automation to power the future of chemistry, Cell Reports Physical Science (2024), https://doi.org/10.1016/j.xcrp.2024.102049

reported, yet the task of building accurate prediction models remains challenging and urgently needed.

The issue of data collection could potentially be addressed through HTE, a cutting-edge technology that harnesses robotics to conduct numerous reactions in parallel, offering a pathway to standardized and reliable datasets. By virtue of its capacity to minimize bias toward data-point selection, HTE holds the potential to markedly enhance the quality and comprehensiveness of collected data. Through rapid and parallel experimentation, HTE facilitates the generation of robust datasets, thus enabling more accurate and reliable analyses. Moreover, the standardized nature of HTE protocols ensures consistency across experiments, further bolstering the reliability of the resulting datasets. As a result, researchers can glean deeper insights into reaction mechanisms, identify trends, and develop predictive models with greater confidence. Ultimately, the integration of HTE into research workflows promises to revolutionize data collection practices in chemistry and propel advancements in fields ranging from drug discovery to materials science.

Integrating automation, data science, and AI with traditional synthetic chemistry has revolutionized our approach to exploring chemical space and addressing synthetic challenges. Central to our endeavors is the development of effective, affordable, and user-friendly HTE instruments. Our journey into HTE-empowered AI chemistry represents a paradigm shift in chemical research, offering unparalleled opportunities for exploration and discovery.

Shan Jiang
School of Physical Science and Technology, ShanghaiTech University

Data-driven material discovery: Integrating AI with automation

Discovering new materials is crucial because it drives innovation across various industries and applications, significantly benefiting our society. However, the vast material space, filled with enormous possibilities of potential compositions and structures, presents a huge challenge to the rapid identification of promising candidates for specific applications. Traditional methods of material discovery rely heavily on labor-intensive manual experiments. Researchers conduct experiments to synthesize and characterize materials, and the process is both time-consuming and resource intensive.

To overcome these obstacles, innovative approaches are necessary to effectively explore the vast chemical space. Data-driven material discovery represents a transformative paradigm in materials science, revolutionizing the approach to material design. In this approach, data are recognized as valuable resources for deriving knowledge from material datasets. By using advanced tools such as material databases and AI, this paradigm aims to unlock new possibilities for material research and has the potential to comprehensively explore the material space.

One of the most important applications of computation in material discovery to date is property prediction, which is often faster than experimental characterization. HT computation is an effective method for evaluating material performance and discovering new materials. As the number of structures to be screened increases, computational calculations become impractical due to the high cost. AI techniques, such as machine learning (ML), can learn patterns from known datasets and make predictions about unknown data in the field of materials. These techniques can be trained on datasets of known materials and their properties and then used to predict the properties of new materials, significantly reducing the cost and time required for material discovery and design. Furthermore, the ideal scenario involves successfully engaging in "inverse design," when a molecule is specifically designed to fulfill a set of predefined criteria for a functional material using AI. Inverse design leverages AI to reverse the traditional material discovery process. Instead of synthesizing materials and then testing their

Please cite this article in press as: Jiang et al., Artificial intelligence and automation to power the future of chemistry, Cell Reports Physical Science (2024), https://doi.org/10.1016/j.xcrp.2024.102049

properties, this approach starts with the desired properties and uses AI algorithms to design molecules that are predicted to exhibit these traits.

Given the immense scope of synthesizing and testing a vast array of materials, integrating automation and robotics into the process can dramatically increase the volume of materials that can be synthesized and evaluated—potentially by several orders of magnitude. Automation technologies, such as automated synthesis robots and HT screening systems, can handle repetitive and labor-intensive tasks with precision and speed, freeing researchers to focus on analysis and innovation. Robotics can perform complex synthesis protocols, prepare samples, and conduct preliminary evaluations of material properties all while ensuring high consistency and repeatability. This enhancement not only accelerates the pace of discovery but also expands the range of materials that can be practically explored and developed.

In the future, automation and AI are poised to significantly accelerate both experimental and computational research programs, particularly when these technologies are integrated to complement each other. This synergy between automated experimental processes and AI-driven computational analysis creates a feedback loop that can rapidly iterate and refine hypotheses and experiments, thereby speeding up the discovery and development of new materials and technologies. As AI and automation become increasingly sophisticated, their integration is expected to lead to more efficient, innovative, and productive scientific research.

Jing Ma
School of Chemistry and Chemical Engineering, Nanjing University

Open black box for material design: Insights learned from cross-scale and closed-cycle automation workflow

AI has made considerable advances in material design, e.g., ML-assisted molecular properties prediction, force-field generation at atomic or molecular scale, HT screening of material candidates, function-directed inverse design strategy, and automatic robotics synthesis. Researchers dream of developing efficient and high-performance AI tools with traits such as high accuracy, high speed, high dimensional material space, HT, high transferability, high accessibility, and high consistency to develop novel functional materials. It is really a dilemma: the higher accuracy the ML model achieves, the more hyperparameters are involved and, thus, the less opaque the model is. Some researchers also have doubts about the ability of insight learning and creative thinking from data and ML, in comparison with human wisdom in using the knowledge and imagination for materials discovery.

Many attempts have been devoted to increasing explainability in the materials design workflow of "structural/functional units design—materials synthesis—materials characterization/spectral assignment—performance optimization/enhancement" with the embedding of material knowledge or the learned features. However, those explainable ML models and AI techniques have been built in separate stages with different spatial and time scales. One could conceive that joint learning or sequential learning in two or more different stages could probably lead to insight through increasing time and length scales to cross scales from atom, molecule, aggregate, and then phase domain to device. Cross-scale machine learning is still rare due to the huge gaps in features and models between different scales. The lack of alignment of computational results at the microscopic scale and in real-world experiment feedback hinders closing the loop of automatic "design & prediction—synthesis & assembly—characterization—performance optimization."

In my opinion, graphs and graph AI are good standing points for bridging the different scales to allow a closed-cycle automatic material exploration in near future. The nodes and edges in graph data could efficiently convey the relative information between different nodes, allowing better learning of relationships in many successful

Please cite this article in press as: Jiang et al., Artificial intelligence and automation to power the future of chemistry, Cell Reports Physical Science (2024), https://doi.org/10.1016/j.xcrp.2024.102049

applications of AI in our daily life and scientific research. With a long history, graphs have indeed been used in molecular systems to simplify the Hamiltonian quantum mechanics of the π-conjugated hydrocarbons or metal complexes by using the adjacency matrix and connectivity. At a molecular scale, it is quite natural to set each atom as a node and each bond as an edge without or with the periodic boundary condition. Graph data structure can also be applicable to mesoscopic property prediction with the coarse-grained model by taking the grain beads as nodes and intergrain interaction as edges. The conversion of various experimental spectra (such as XRD, IR, and XAS) and images (STM/STEM/SPM) into graph structural data yields good performance in material-specific applications.

Although an LLM has boosted the transformation of the multi-modal input to chemically or physically meaningful symbols, knowledge is usually lacking in LLMs. Some novel and user-friendly algorithms are desired for automatic knowledge graph (KG) construction and reinforcement learning from human feedback. The relation between the reaction substrates and material synthesis conditions (e.g., solvent and temperature selections) could be explored in KGs. The dynamic integration of LLMs with KGs would provide numerous opportunities for realizing multi-model and multi-scale synergism, closing the loop of computation-experiment, and automatically optimizing multi-tasks. Such powerful material agents are expected to open a black box in prediction models and provide new insight into material innovation.

Jun Jiang
Key Laboratory of Precision and Intelligent Chemistry, University of Science and Technology of China

Powering the future of chemical and material discoveries with AI-driven autonomous chemistry

Chemistry and materials science are undergoing a profound transformation driven by advancements in AI and robotics. This transformative shift toward autonomous chemistry is characterized by integrating sophisticated AI algorithms and automated systems into the fabric of daily laboratory research activities, fundamentally changing how scientific investigations are conducted. Recent advancements in AI, particularly in LLMs, have greatly enhanced autonomous chemistry. AI models and agents have become integral to automating chemistry, proving instrumental in closing the predict-make-measure discovery loop and interpreting scientific data. AI-driven autonomous systems utilize their intelligent capabilities to plan experiments, interact with robotics, and manage data, significantly improving experimental efficiency and precision. Alongside AI, robotics has also made significant strides in laboratory automation, offering diverse systems tailored to meet the complex demands of autonomous chemistry. These systems include automated HT platforms, precision robotic arms, mobile robots, collaborative robots, and others.

Autonomous chemistry has been progressing through three major phases, each marked by significant advancements in the field's approach and capabilities. Phase one involves the establishment of individual autonomous labs designed to tackle specific challenges. These labs typically operate in isolation, focusing on localized problems without much inter-lab communication or data sharing. Currently, most autonomous labs globally are in this phase. We are rapidly transitioning into phase two: adopting coordinated strategies such as cloud-based systems for delocalized and asynchronous research. This phase involves distributing tasks in an experimental workflow and orchestrating them via AI, facilitating seamless data and resource integration across labs, overcoming geographical and temporal constraints to expand discovery potential.

On the horizon, phase three envisions advanced nationwide or global networks of intelligent scientist systems that we have conceptualized and proposed. We envision these integrated networks of intelligent systems conducting end-to-end autonomous research, showing high degrees of cognitive and operational integration by

Please cite this article in press as: Jiang et al., Artificial intelligence and automation to power the future of chemistry, Cell Reports Physical Science (2024), https://doi.org/10.1016/j.xcrp.2024.102049

merging AI models with robotic processes. Notably, this phase aims to develop human-machine collaborative systems, where cognitive intelligence supports the proposal and analysis of scientific problems through knowledge fusion. This process drives robotic experimental systems and computer simulations, producing high-quality data that feed back into the AI models to optimize and refine them, thereby solving complex scientific challenges. This integration transforms autonomous labs into highly connected, efficient entities that transcend traditional research boundaries, enabling scientific breakthroughs on a nationwide or global scale. The implementation of intelligent scientist systems, as we envision, will involve creating centralized platforms that gather and analyze data, develop intelligent models, and refine scientific methods and technologies. These platforms, acting as scientific "brains," will guide distributed innovation facilities that support users in achieving specific scientific breakthroughs. This integrated approach will foster a new form of scientific research organization: centralized, resource-intensive development and deployment of scientific intelligence will drive distributed, localized experimental operations to catalyze innovations. This structure ultimately lowers the barriers to interdisciplinary and cross-domain research, enabling scientists and researchers in both academia and industry at all levels to engage in highly specialized experimentation and personalized scientific inquiry.

As the landscape of chemical and materials science research continues to evolve, the potential of AI-driven autonomous chemistry is becoming increasingly apparent. The journey from individual autonomous labs to extensive networks of intelligent systems will catalyze a revolutionary shift in how we approach scientific inquiries and challenges. This transition maximizes the efficiency and efficacy of autonomous chemistry research and democratizes the ability to innovate across academic disciplines and various industries. As we move forward, the continued and evolving fusion of state-of-the-art AI and robotic technologies with chemistry and materials science promises to accelerate the discovery of superior chemicals and materials with desired functions, delivering significant benefits to society at large.

Zhigang Shuai
School of Science and Engineering, The Chinese University of Hong Kong.

Time for AI to meet organic light-emitting diodes

The molecular design of organic light-emitting diode (OLED) materials requires a data-driven technique for three essential parameters: the emission wavelength, the peak width for better color purity, and quantum efficiency for brightness.

OLEDs have become the focus of research in both academia and industry since the discovery of room-temperature low-voltage thin-film organic electroluminescence. Now, OLEDs have been widely used in mobile phone and television displays. OLEDs devices do not need backlight, allowing energy-efficient display with wide viewing angles, high-contrast color, fast response, flexibility, and even transparency. According to IDTechEx, by 2030 the global market value for OLEDs will reach over 60 billion USD.

Optical emission stems from the lowest molecular excited state according to Kasha's rule. The emission wavelength is determined by the optical gap and the radiative decay rate is proportional to the square of transition dipole moment (oscillator strength). For fluorescence, it is the lowest singlet excited state (S_1) and for phosphorescence the lowest triplet (T_1), whereas for TADF, it is the nature of S_1 and the S_1–T_1 gap (relevant for reverse intersystem crossing. It is a formidable challenge to determine the positions of S_1 and T_1 as well as the transition dipole moment including spin-orbit coupling through quantum chemistry. It seems the only choice for calculating the excited state for a typical OLEDs molecule with a

Please cite this article in press as: Jiang et al., Artificial intelligence and automation to power the future of chemistry, Cell Reports Physical Science (2024), https://doi.org/10.1016/j.xcrp.2024.102049

few tens of atoms is the time-dependent density functional theory. And after more than 20 years of efforts on the assessments of adjusting the exchange-correlation functionals, a generally precise and reliable computational approach is still elusive, especially for TADF for which both S_1 and T_1 should be determined at the same time.

Narrower full width at half maximum (FWHM) for the emission spectrum is demanded for color purity. In general, for highly efficient OLED molecules, both the emission peak position and FWHM in thin film and solution phases should be close to each other, indicating weak intermolecular interaction in amorphous film, so that prediction for molecular emission FWHM could serve as good indicator for color purity. In this respect, the recently proposed multi-resonant TADF molecules demonstrated narrow emission. Intrinsically, it is the electronic excited state vibronic coupling that determines the FWHM. It has been shown that the thermal vibration correlation formalism (TVCF) as implemented in a computational package MOMAP can present systematic prediction on the emission line-shape including FWHM, if not for the absolution value within 10%–20% error.

Nevertheless, as far as the quantum efficiency is concerned, the situation becomes much more complicated. The total quantum efficiency is expressed as a product of three factors: (1) that of carrier recombination rate (not all the electro-pumped carries could form an electron-hole pair because there is always a current in the device); (2) that of portion for emissive species limited by spin statistics; (3) that of molecular luminescence quantum efficiency $\eta = \frac{k_r}{k_r+k_{nr}}$, where according to Einstein's spontaneous radiation theory, the spontaneous radiative decay rate is $k_r = \frac{8\pi^2\nu_{fi}^3}{3\varepsilon_0\hbar c^3}\mu_{fi}^2 \approx \frac{f\nu_{fi}^2}{1.5}$ (f is the oscillator strength and ν is the emission wavenumber), and the essential challenge lies in determination of non-radiative decay k_{nr}. Although k_r is relatively easy to predict, k_{nr} is far from that. TVCF has been a good starting point to reveal the relationship between quantum efficiency and molecular structure for a number of OLED systems, which had been not only employed to OLED quantum efficiency but also applied to rationalize aggregation-induced emission and pure organic phosphorescence, as well as to molecular design of optical sensing/detection and photocatalysis. To meet the ever-growing OLED market, precise prediction for quantum efficiency as well as wavelength and FWHM is becoming imperative, and the data-driven AI approach seems to be the most appropriate choice.

DECLARATION OF INTERESTS

The authors declare no competing interests.