

Harald Witte\*, Tobias U. Blatter, Priyanka Nagabhushana, David Schär, James Ackermann, Janne Cadamuro and Alexander B. Leichtle

# Statistical learning and big data applications

<https://doi.org/10.1515/labmed-2023-0037>

Received March 28, 2023; accepted May 9, 2023;

published online June 2, 2023

**Abstract:** The amount of data generated in the field of laboratory medicine has grown to an extent that conventional laboratory information systems (LISs) are struggling to manage and analyze this complex, entangled information (“Big Data”). Statistical learning, a generalized framework from machine learning (ML) and artificial intelligence (AI) is predestined for processing “Big Data” and holds the potential to revolutionize the field of laboratory medicine. Personalized medicine may in particular benefit from AI-based systems, especially when coupled with readily available wearables and smartphones which can collect health data from individual patients and offer new, cost-effective access routes to healthcare for patients worldwide. The amount of personal data collected, however, also raises concerns about patient-privacy and calls for clear ethical guidelines for “Big Data” research, including rigorous quality checks of data and algorithms to eliminate underlying bias and enable transparency. Likewise, novel federated privacy-preserving data processing approaches may reduce the need for centralized data storage. Generative AI-systems including large language models such

as ChatGPT currently enter the stage to reshape clinical research, clinical decision-support systems, and healthcare delivery. In our opinion, AI-based systems have a tremendous potential to transform laboratory medicine, however, their opportunities should be weighed against the risks carefully. Despite all enthusiasm, we advocate for stringent added-value assessments, just as for any new drug or treatment. Human experts should carefully validate AI-based systems, including patient-privacy protection, to ensure quality, transparency, and public acceptance. In this opinion paper, data prerequisites, recent developments, chances, and limitations of statistical learning approaches are highlighted.

**Keywords:** ChatGPT; clinical decision-support systems; laboratory medicine; machine learning; personalized medicine; wearables.

## Introduction

In recent years, new technologies have led to a steep increase of the amount of data generated in the field of laboratory medicine, including clinical chemistry, hematology, microbiology, genetics, and various “omics”-approaches while today’s laboratory information systems (LISs) can hardly cope with this plethora of data. This rich pool of “Big Data” needs innovative approaches and technical improvements to render it accessible for clinical data science [1, 2]. Advanced analysis methods which were formerly only manageable by a handful of experts, are now widely utilized and enter clinical routine. This includes statistical learning, a framework of machine learning (ML) that allows to build predictive models based on underlying data. The combination of “Big Data” and statistical learning disrupts the conventional field of laboratory medicine, allowing improved diagnostics and the development of accurate prognostic models using data not only from analytical instruments, but also from the large pool of data available from electronic health records (EHRs). We will briefly highlight chances and challenges of recent developments for the application of statistical learning in clinical laboratory sciences.

## Overview

Statistical learning is an ML framework to infer distributional properties of available data with the goal to allow

---

Harald Witte and Tobias U. Blatter contributed equally as first authors. Janne Cadamuro and Alexander B. Leichtle contributed equally as senior authors.

**\*Corresponding author: Harald Witte**, Department of Clinical Chemistry, Inselspital—University Hospital Bern, 3010 Bern, Switzerland; and Department of Biomedical Research (DBMR), University of Bern, 3010 Bern, Switzerland, E-mail: harald.witte@unibe.ch. <https://orcid.org/0000-0002-4421-3580>

**Tobias U. Blatter**, Department of Clinical Chemistry, Inselspital—University Hospital Bern, Bern, Switzerland; and Graduate School for Health Sciences (GHS) – University of Bern, Bern, Switzerland. <https://orcid.org/0000-0002-0298-8177>

**Priyanka Nagabhushana, David Schär and James Ackermann**, Department of Clinical Chemistry, Inselspital—University Hospital Bern, Bern, Switzerland. <https://orcid.org/0009-0005-2025-108X> (P. Nagabhushana). <https://orcid.org/0003-7306-0354> (D. Schär). <https://orcid.org/0000-0002-4215-1232> (J. Ackermann)

**Janne Cadamuro**, Department of Laboratory Medicine, Paracelsus Medical University Salzburg, Salzburg, Austria. <https://orcid.org/0000-0002-6200-9831>

**Alexander B. Leichtle**, Department of Clinical Chemistry, Inselspital—University Hospital Bern, Bern, Switzerland; and Center of Artificial Intelligence in Medicine (CAIM), University of Bern, Bern, Switzerland. <https://orcid.org/0000-0002-6528-9904>

predictions, i.e., find a predictive function for the data. This approach follows the scientific method: First, a hypothesis is formulated for a phenomenon of interest. Based on detailed observations (i.e., collected data), a model of the phenomenon is set up and continuously refined. The ultimate aim of the model is to predict the phenomenon from similar new data. In the setting of statistical learning and artificial intelligence, this process is “just” automated. This allows statistical learning procedures to process large amounts of data and detect inherent patterns which would be hard or even impossible to infer otherwise. The multiple approaches to statistical learning include supervised, unsupervised and reinforcement learning [3]. The now omnipresent “artificial intelligence” (AI) also makes use of these data-intense ML-frameworks.

Laboratory medicine and clinical science generate “Big Data” of manifold types, including electronic health records (EHRs), imaging data or data of various “omics”-approaches including metabolomics, genomics and proteomics. Such data is – in addition to its sheer amount – often highly complex and intrinsically entangled. Here, the ongoing process of automatization in laboratory medicine offers new possibilities to incorporate advanced statistics into fully autonomous routines with the welcome side effect of lower costs [4]. Hence, statistical learning and clinical data are a predestined match to the benefit of patients and research (Figure 1).

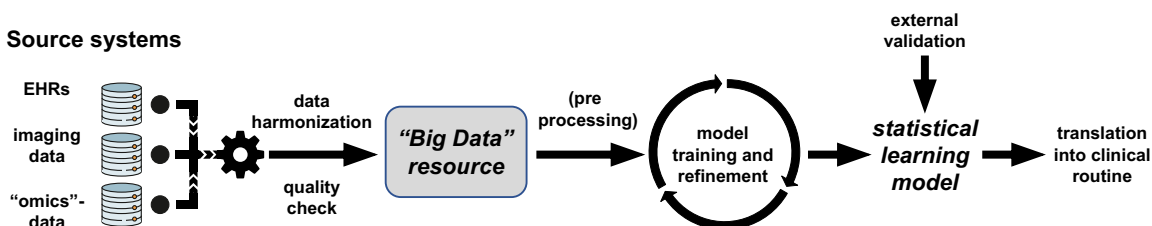
## Data prerequisites

To be most useful for clinical and research purposes, data should ultimately conform to the FAIR data principles, i.e., the data should be findable, accessible, interoperable, and reusable, which poses a substantial challenge to current LISs [2]. Frequently, clinical data is stored in clinical data warehouses (CDWs) or data lakes, with CDWs being common

for structured data like EHRs and laboratory results and data lakes being more suitable for larger data volumes like images or “omics”-data in their native, raw formats. Non-relational databases (e.g., NoSQL-databases, referring to the difference from traditional structured query language (SQL)-databases) are an alternative approach without a fixed schema, which may offer enhanced flexibility, speed and scalability as compared to CDWs, but may be more complex to set up. Federated analysis approaches in turn constitute a secure, privacy-preserving alternative to centralized data storage [5].

LISs tend to be historically grown, therefore the data in the source systems often do not comply with current standardized quality measures and may lack correct formatting or suffer from incompleteness. Consequently, the extract, transform, load (ETL)-processes to prepare data for analysis and clinical research (as a “Big Data”-resource) are cumbersome. In the worst case they are not addressed at all, e.g., due to a lack of resources, thereby precluding research and use of available data. The aim should therefore be to harmonize data and supplement it with metainformation at the source to optimize usefulness and interoperability of data over time and between centers by using controlled vocabulary as well as international standards and classification systems [2] – this is admittedly a significant investment but will pay off down the road by streamlining internal processes and facilitating national and international collaborations. Individual “lighthouse projects” can break down the enormous challenge into smaller parts by first implementing certain aspects such as quality control or introduction of specific standards to start with.

Common standards and systems include the Unified Code for Units of Measure (UCUM) for units, the Anatomical Therapeutic Chemical (ATC) classification system for drugs, the International Classification of Diseases (ICD) for diseases and maladies, the Logical Observation Identifiers Names and Codes (LOINC) for analyses, or the Systematized Nomenclature of



**Figure 1:** Simplified workflow diagram for statistical learning approaches. Clinical data often stem from multiple source systems, e.g., for different types of data like laboratory and imaging data, or for different internal origins, e.g., results from central laboratories vs. point-of-care-testing (POCT). Data needs to be harmonized, quality-checked, and combined to an interoperable “Big Data”-resource, respecting patient-privacy. Data may be preprocessed and used for training, validation and refinement of a statistical model. The final aim after external validation is the translation into clinical routine. Notably, novel federated analysis approaches allow the analysis of multi-cohort data without the need to store data in a centralized manner [5]. The need for data to be qualified at the point of care nevertheless remains of course.

Medicine-Clinical Terms (SNOMED-CT) for medical terms. Suitable standardized formats for data exchange, e.g., the Fast Healthcare Interoperability Resources (FHIR) standard or graph-based data representations such as the Resource Description Framework (RDF) will promote data interoperability further. Information on the type of method applied as well as on manufacturer and type of instruments or reagents used should also be provided along with laboratory results. Here, standardized unique identifiers are most useful, e.g., from the Global Unique Device Identifier Database (GUDID) or the European Database on Medical Devices (EUDAMED), as well as medical device nomenclatures such as the Global Medical Device Nomenclature (GMDN) or the European Medical Device Nomenclature (EMDN). Notably, for some analyses, e.g., immunoassays, such additional meta-information is particularly important and indispensable to make data fully comparable between laboratories (or avoid comparison of “apples and pears”).

## Recent developments

In this section we will briefly present a few recent publications highlighting different perspectives of the immense potential of statistical learning models – due to space constraints, this list is a small selection far from complete. For a broader overview, we refer the interested reader to recent reviews [3, 6, 7].

One of the core fields for statistical learning models in a medical context is diagnostics. Comparably simple approaches like logistic regression models as well as more complex deep learning (DL) models can forecast emergence or severity of a disease, for example from laboratory results of COVID-19 patients [8, 9]. Similarly, an approach based on Bayesian Model Averaging in combination with orthogonal data augmentation used routine hospital data for the prediction of myocardial ischemia [10].

Ideally, AI-based diagnostics will eventually reduce mortality, e.g., by frequent causes like sepsis or cardiac death from arrhythmia. A recent DL approach using cardiac magnetic resonance images and clinical covariates predicts individualized survival curves for patients with ischemic heart disease [11]. Prediction (and consequently avoidance) of sepsis is another important focus of research, be it from a set of biomarkers [12] or EHRs [13]. These accurate and generalizable predictions offer a non-invasive and relatively cheap next-level support for clinical diagnostics and decision-making.

Reliable diagnoses are also key to ensure cost-effectiveness by avoiding unnecessary repeated testing. Algorithms can support the interpretation of test outcomes, e.g., of rapid HIV

tests in rural areas of developing countries to reduce the number of false negatives and false positives [14]. Likewise, ML-based models can improve the precision of diagnoses, for example kidney diseases, to choose the optimal therapeutic approach [15]. While the assembly of “Big Data”-sets can be cumbersome and expensive, statistical models – once trained – are easy to apply and can be cost-saving, in particular when incorporated in readily available devices like smartphones. Their worldwide spread has paved the ground for a wide variety of health-related applications. This includes apps like fitness trackers, often in combination with wearable devices, but also medical applications including non-invasive image-based detection of anemia [16] or skin cancer [17], and detection and management of hypertension [18]. Notably, several applications that offer smartphone-based support for wound care and urine analysis have already been CE-marked and FDA-cleared [19]. Such systems open up improved access to novel diagnostic approaches for rural areas or developing countries and can in addition reduce the workload of clinical staff in overwhelmed health systems.

Another advantage of “Big Data” research is the potential to confirm or rectify previous underpowered studies. A recent study for example cautions against popular claims that the gut microbiome is a driving factor in autism spectrum disorder. The results instead endorse a model in which the reduced microbial taxonomic diversity found in autistic patients is consequence rather than cause of autism [20].

Curiously enough, artificial intelligence (AI) may by now also offer support in areas supposed to be genuinely human. A recent study trained a multi-task RoBERTa-based bi-encoder model to characterize empathy in conversations [21]. Human peer supporters of online mental health support platforms receiving real-time feedback from this AI were enabled to provide more empathic responses to support seekers – a good example of a fruitful human-AI collaboration. Along those lines, the assessment of human sentiment by AI-models using social media data is a growing field of interest. While holding a lot of commercial potential (e.g., revealing customer preferences), it can be of clinical interest as well, for example for the detection of depression [22].

## Challenges and limitations

Personalized medicine is one of the fields envisioned to benefit most from AI-models. The combination of multifaceted “Big Data” pools containing ever more details about an individual and the growing possibilities to analyze and interconnect this information, notably also across multiple data pools, offers many chances to generate predictions and insight. Personalization of healthcare, however, may require

the collection of sensitive patient data which is now being accumulated at an unprecedented scale [23]. This includes “classical” medical data or health data from wearable devices, but also non-medical data like personal habits, movement profiles, personal web traffic, or even directly identifying information like social security numbers or mobile phone numbers. This data is collected by or shared with multiple private and public stakeholders. Whether this always happens on the basis of well-informed consent, where “data donors” are fully aware of all implications of data sharing, may at least be doubted [24]. This opens the door for violations of patient privacy and confidentiality with the risk of data being mishandled or potentially misused [25]. It is therefore crucial to ensure that data is collected on the basis of well-informed consent, but also stored as well as analyzed in a secure and responsible manner. Importantly, well-intentioned measures such as “de-identification” may not suffice – full anonymization of detailed sensitive data (including health data) can be difficult to attain [26]. Some voices therefore advocate to accept a (small) risk of re-identification of patients for the sake of medical progress [27]. It would be desirable that data consumers (e.g., researchers or “end users”) act in an ethical way by their own effort, but in absence of clear guidelines (“Do’s and Don’ts”) it may vary widely what individuals find “acceptable”. Therefore, there is definitely a need for clear ethical guidelines for “Big Data” research, as using “Big Data” comes along with great responsibility [28, 29]. An elegant solution for this dilemma may be offered by federated data processing approaches which keep data exclusively on-site of a trusted data provider (e.g., a hospital) and only reveal aggregate data to a researcher (“no copy, no move”-principle) [5].

A sometimes neglected limitation of AI-based models originates in the data itself: “Big Data” does not inevitably warrant “high-quality data”. Data harmonization and rigorous checks to remove bias and ensure plausibility and quality of data need to be the rule before data is made available to statistical learning models (Figure 1), both for training and the actual analysis [2]. Depending on the application, such measures during the ETL-process include the removal of data where crucial information is missing (e.g., diagnosis codes) or values which have been entered wrongly (for example, implausible birthdates, body weights or heights), removal of biasing information where applicable (e.g., gender, information on ethnic background, or income), removal of obviously wrong measurements (for example, laboratory values incompatible with living patients), the addition of meta-information on analysis methods (e.g., to avoid the comparison of incompatible assays by providing information on manufacturer and version of a test

kit), or the unification of results of compatible assays under a common standardized code (e.g., mapping of multiple internal analysis codes to one suitable LOINC) [30].

Statistical learning models are to a large extent as good as the data they are trained with – trained with biased data, their output will be biased, too [31]. A recent study for instance highlights that current stroke-prediction models perform worse for Black as compared to White adults [32] – further effort is urgently called for to eliminate such bias. This also applies in a broader medical context, as a model trained on specific patient populations to predict medical outcomes inherently reflects patients’ access to health care, stigma around seeking medical care, and cultural health norms [33]. Another issue with statistical learning algorithms is their lack of transparency. Explainable artificial intelligence (XAI) which reveals the “reasoning” behind its decisions (“white-box” models) may be the way forward here [34, 35]. A different sort of bias is neither rooted in the data nor the statistical learning models: Will all patients benefit alike from elaborated new applications, or will it primarily be those anyway privileged already?

Future advanced models will ideally help to identify therapeutic targets on the level of the individual patient. A recent high-profile genomics publication, however, suggests that the currently available tools for clinical interpretation need further refinement to achieve this milestone – large amounts of data alone are not sufficient [36]. In part this may be due to erroneous model setup. While novel tools have drastically facilitated the generation of ML-based models they do not automatically circumvent methodological pitfalls, including data leakage, leading to reproducibility failures [37].

## Conclusions and outlook

Statistical learning and “Big Data” applications offer immense chances to the field of laboratory medicine and beyond, including improved diagnostic and prognostic models as well as potential decision support. The diverse and overarching “Big Data” their success is built on, however, comes along with major challenges which call for a considerate, responsible implementation of future applications. International efforts of the scientific community, society as a whole, and regulatory bodies will be required to balance progress and patient privacy.

Currently, we are also witnessing the debut of generative AI to a broad audience of both expert and lay users, with large language models (LLMs) using natural language processing (NLP) like ChatGPT being the most celebrated ones for now. This is very likely more than a brief hype but

rather the beginning of an exciting era where LLMs may support experts with the analysis of the growing amount of literature, facilitating the generation of new clinical hypotheses or simply staying up to date with recent developments, resulting in better-informed decisions of clinical staff. LLMs may also improve communication between experts of different fields or doctors and patients – it is not necessarily a core strength of every specialist to convey expertise in a straightforward manner. If “digital literacy” is fostered at early stages of education, AI-based systems may also enable personalized training of students and healthcare professionals. Overall, recent developments offer numerous opportunities to diagnostics and “Big Data” applications and may bring education, research, and healthcare delivery to the next level [38, 39].

While it is tempting to see new promising approaches through rose-colored glasses, they should undergo the same added-value assessment as any other new methodology to avoid “ChatGPT hallucinations” (plausible-sounding yet incorrect statements of AI-systems) [40, 41]. Ideally, protection and security of patients and their data should be a natural “side effect” for responsible scientists and developers, rather than a primary target. Full transparency of decisions which algorithms are influencing is the way to go to ensure public acceptance.

Eventually, it is not guaranteed that statistical learning algorithms will actually grasp biological or medical significance, i.e., draw the right conclusions from detected patterns. We therefore advocate a future where human experts and AI work hand in hand to combine the best of both worlds – in laboratory medicine and beyond.

**Acknowledgments:** Icons used in Figure 1 have been made by Freepik from flaticon.com or by icons8.com.

**Research funding:** This work has been funded by grants from the Bern Centre for Precision Medicine (PGX-link PGM), the Swiss National Science Foundation (2021–01104), and the Swiss Personalized Health Network (2018DEV22).

**Author contributions:** All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

**Competing interests:** Authors state no conflict of interest.

**Informed consent:** Not applicable.

**Ethical approval:** Not applicable.

## References

1. Haymond S, McCudden C. Rise of the machines: artificial intelligence and the clinical laboratory. *J Appl Lab Med* 2021;6:1640–54.
2. Blatter TU, Witte H, Nakas CT, Leichtle AB. Big data in laboratory medicine-FAIR quality for AI? *Diagnostics* 2022;12:1923.
3. Habehh H, Gohel S. Machine learning in healthcare. *Curr Genomics* 2021;22:291–300.
4. Naugler C, Church DL. Automation and artificial intelligence in the clinical laboratory. *Crit Rev Clin Lab Sci* 2019;56:98–110.
5. Froelicher D, Troncoso-Pastoriza JR, Raisaro JL, Cuendet MA, Sousa JS, Cho H, et al. Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption. *Nat Commun* 2021;12:5910.
6. Sebastian AM, Peter D. Artificial intelligence in cancer research: trends, challenges and future directions. *Life* 2022;12. <https://doi.org/10.3390/life12121991>.
7. Bunch DR, Durant TJ, Rudolf JW. Artificial intelligence applications in clinical chemistry. *Clin Lab Med* 2023;43:47–69.
8. Singh V, Kamaleswaran R, Chalfin D, Buño-Soto A, San Roman J, Rojas-Kenney E, et al. A deep learning approach for predicting severity of COVID-19 patients using a parsimonious set of laboratory markers. *iScience* 2021;24:103523.
9. Schmidt W, Józwiak B, Czabajska Z, Pawlak-Buś K, Leszczynski P. On-admission laboratory predictors for developing critical COVID-19 during hospitalization - a multivariable logistic regression model. *Ann Agric Environ Med* 2022;29:274–80.
10. Liniger Z, Ellenberger B, Leichtle AB. Computational evidence for laboratory diagnostic pathways: extracting predictive analytes for myocardial ischemia from routine hospital data. *Diagnostics* 2022;12. <https://doi.org/10.3390/diagnostics12123148>.
11. Popescu DM, Shade JK, Lai C, Aronis KN, Ouyang D, Moorthy MV, et al. Arrhythmic sudden death survival prediction using deep learning analysis of scarring in the heart. *Nat Cardiovasc Res* 2022;1:334–43.
12. Su M, Guo J, Chen H, Huang J. Developing a machine learning prediction algorithm for early differentiation of urosepsis from urinary tract infection. *Clin Chem Lab Med* 2023;61:521–9.
13. Colborn KL, Zhuang Y, Dyas AR, Henderson WG, Madsen HJ, Bronsert MR, et al. Development and validation of models for detection of postoperative infections using structured electronic health records data and machine learning. *Surgery* 2023;173:464–71.
14. Turbé V, Herbst C, Mngomezulu T, Meshkinfamfard S, Dlamini N, Mhlongo T, et al. Deep learning of HIV field-based rapid tests. *Nat Med* 2021;27:1165–70.
15. Triep K, Leichtle AB, Meister M, Fiedler GM, Endrich O. Real-world health data and precision for the diagnosis of acute kidney injury, acute-on-chronic kidney disease, and chronic kidney disease: observational study. *JMIR Med Inform* 2022;10:e31356.
16. Mannino RG, Myers DR, Tyburski EA, Caruso C, Boudreaux J, Leong T, et al. Smartphone app for non-invasive detection of anemia using only patient-sourced photos. *Nat Commun* 2018;9:4924.
17. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017;542:115–8.
18. Visco V, Ferruzzi GJ, Nicastro F, Virtuoso N, Carrizzo A, Galasso G, et al. Artificial intelligence as a business partner in cardiovascular precision medicine: an emerging approach for disease detection and treatment optimization. *Curr Med Chem* 2021;28:6569–90.
19. Healthy.Io. <https://healthy.io/> [Accessed 27 Mar 2023].
20. Yap CX, Henders AK, Alvares GA, Wood DLA, Krause L, Tyson GW, et al. Autism-related dietary preferences mediate autism-gut microbiome associations. *Cell* 2021;184:5916–31.e17.
21. Sharma A, Lin IW, Miner AS, Atkins DC, Althoff T. Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nat Mach Intell* 2023;5:46–57.

22. Babu NV, Kanaga EGM. Sentiment analysis in social media data for depression detection using artificial intelligence: a review. *SN Comput Sci* 2022;3:74.
23. Taylor P. Total data volume worldwide 2010–2025. Statista. <https://www.statista.com/statistics/871513/worldwide-data-created/> [Accessed 27 Mar 2023].
24. Grady C. Enduring and emerging challenges of informed consent. *N Engl J Med* 2015;372:855–62.
25. Khanijahani A, Iezadi S, Agoglia S, Barber S, Cox C, Olivo N. Factors associated with information breach in healthcare facilities: a systematic literature review. *J Med Syst* 2022;46:90.
26. Vokinger KN, Stekhoven DJ, Krauthammer M. Lost in anonymization - a data anonymization reference classification merging legal and technical considerations. *J Law Med Ethics* 2020;48:228–31.
27. Seastedt KP, Schwab P, O'Brien Z, Wakida E, Herrera K, Marcelo PGF, et al. Global healthcare fairness: we should be sharing more, not less, data. *PLoS Digit Health* 2022;1: e0000102.
28. Ferretti A, Ienca M, Velarde MR, Hurst S, Vayena E. The challenges of big data for research ethics committees: a qualitative Swiss study. *J Empir Res Hum Res Ethics* 2022;17:129–43.
29. Pennestrì F, Banfi G. Artificial intelligence in laboratory medicine: fundamental ethical issues and normative key-points. *Clin Chem Lab Med* 2022;60:1867–74.
30. Witte H, Nakas C, Bally L, Leichtle AB. Machine learning prediction of hypoglycemia and hyperglycemia from electronic health records: algorithm development and validation. *JMIR Form Res* 2022;6:e36176.
31. Vokinger KN, Feuerriegel S, Kesselheim AS. Mitigating bias in machine learning for medicine. *Commun Med* 2021;1:25.
32. Hong C, Pencina MJ, Wojdyla DM, Hall JL, Judd SE, Cary M, et al. Predictive accuracy of stroke risk prediction models across Black and white race, sex, and age groups. *JAMA* 2023;329:306–17.
33. Ntoutsis E, Fafalios P, Gadiraju U, Iosifidis V, Nejdli W, Vidal ME, et al. Bias in data-driven artificial intelligence systems—an introductory survey. *Wiley Interdiscip Rev Data Min Knowl Discov* 2020;10:e1356.
34. Bernal J, Mazo C. Transparency of artificial intelligence in healthcare: insights from professionals in computing and healthcare worldwide. *NATO Adv Sci Inst Ser E Appl Sci* 2022;12:10228.
35. Amann J, Vetter D, Blomberg SN, Christensen HC, Coffee M, Gerke S, et al. To explain or not to explain?—Artificial intelligence explainability in clinical decision support systems. *PLoS Digit Health* 2022;1: e0000016.
36. Andre F, Filleron T, Kamal M, Mosele F, Arnedos M, Dalenc F, et al. Genomics to select treatment for patients with metastatic breast cancer. *Nature* 2022;610:343–8.
37. Kapoor S, Narayanan A. Leakage and the reproducibility crisis in ML-based science. *arXiv [cs.LG]*; 2022. <https://doi.org/10.48550/arXiv.2207.07048>.
38. Will ChatGPT transform healthcare? *Nat Med* 2023;29:505–6.
39. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit Health* 2023;2: e0000198.
40. Shen Y, Heacock L, Elias J, Hentel KD, Reig B, Shih G, et al. ChatGPT and other large language models are double-edged swords. *Radiology* 2023;307:e230163.
41. Cadamuro J, Cabitza F, Debeljak Z, De Bruyne S, Frans G, Perez SM, et al. Potentials and pitfalls of ChatGPT and natural-language artificial intelligence models for the understanding of laboratory medicine test results. An assessment by the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Working Group on Artificial Intelligence (WG-AI). *Clin Chem Lab Med* 2023;61: 1158–66.