Empowered patients
Sustainable healthcare

Ensuring fairness and inclusion in AI-guided patient screening

Q&A with **Sara Reis** –
PhD Medical Imaging & Computing,
Data Scientist, HN

hn-company.co.uk

Here, researcher and HN Data Scientist, Sara Reis considers how machine-learning approaches and data can be used to identify those most in need of care but often under-served.

The current pandemic has laid bare the inequalities in access to services and the impact this has on health and social outcomes for marginalised segments of society.

This is not a new issue[1]. What is new, is the richness of the data to provide insights into the inequalities and their underlying causes, at least in countries where this data is collected to a high-level of accuracy such as in the UK and within the NHS and social care.

Using large, connected data sources, applying AI technology now allows services to actively identify and reduce social inequalities in both access and outcomes.

At HN, AI is used to screen, at point of care, patients at risk of poorly controlled disease and unplanned care. Identified patients are contacted and offered immediate clinical remote support and coaching.

Can AI help address inequalities, or does it contain biases that will exacerbate these challenges?

"If identifying and mitigating internal bias is widely known and defined and many tools have been developed in order to tackle it,[2] the same cannot be said on external bias."

Yes, AI can both exacerbate and reduce gaps in inequalities. HN's approach to using AI to support health and care systems, conscious of biases, is focused on surfacing and supporting unmet needs. Let me explain how.

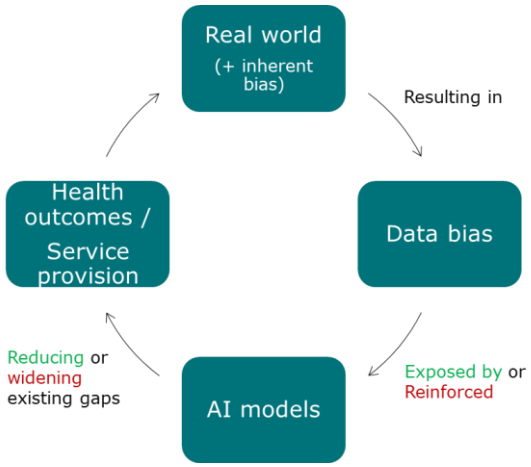
AI models learn from historically collected data to perform specific tasks, such as detecting cancer in medical images or identifying at-risk patients based on their healthcare consumption patterns. The collected data is a proxy of the 'real world', thus reflecting the inherent societal and historical biases and inequalities for certain populations – usually called protected groups[3].

[1] The Economist, "Covid-19 has shone a light on racial disparities in health", https://www.economist.com/international/2020/11/21/covid-19-has-shone-a-light-on-racial-disparities-in-health. Last accessed in Dec 2020

[2] Jakub Wiśniewski and Przemysław Biecek, "Flexible tool for bias detection, visualization, and mitigation.", (2020), GitHub repository, https://https://github.com/ModelOriented/fairmodels/. Last accessed in Dec 2020

[3] Protected groups are identified in the Equality Act 2010 as group of persons defined by reference to a characteristic against which is it illegal to discriminate.

The diagram below shows how AI models have the risk of reinforcing historical and societal bias and exacerbate health inequalities.



AI models trained with these biased historical datasets may result in flawed predictions, therefore reproducing and amplifying historical patterns of health inequality and discrimination.[4] Moreover, key biases in the design, data, and deployment of an AI model may perpetuate or exacerbate health care disparities if left unchecked.[5]

We have undergone an extensive literature review on how to ensure fairness when designing, training and deploying an AI model in healthcare. Various proposed checklists have been published, the majority focusing on internal bias.[6]

Internal bias focuses on the inner workings of the data and model design. Checklists include answering the following:

“Was the best design used?”

“Is the training data less representative for a protected group?”

“Do labels mean the same for all groups?”

Taking this into account, HN follows a three-step fairness framework approach (full internal bias checklist can be found at the end of the article):

- **Awareness** – Apply best practices of fairness-aware design and implementation. HN models are trained and tested on properly representative, relevant, accurate, and generalisable datasets. Model architectures do not include target variables or

[4] Tat, Emily et al. "Addressing bias: artificial intelligence in cardiovascular medicine”, *The Lancet Digital Health* vol. 2,12 (2020): e635-e636. https://doi.org/10.1016/S2589-7500(20)30249-1
[5] Rajkomar, Alvin et al. “Ensuring Fairness in Machine Learning to Advance Health Equity.” *Annals of internal medicine* vol. 169,12 (2018): 866-872. doi:10.7326/M18-1990
[6] A. Madaio, Michael et al. "Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI." *In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. *Association for Computing Machinery, New York, NY, USA*, 1–14. (2020) https://doi.org/10.1145/3313831.3376445
Leslie, D. “Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector.” *The Alan Turing Institute.* (2019) https://doi.org/10.5281/zenodo.3240529
Lee, Nicol Turner et al. "Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms." *Center for Technology Innovation, Brookings*. Tillgänglig online: https://www. brookings. edu/research/algorithmic-bias-detection-and-mitigation-bestpractices-and-policies-to-reduce-consumer-harms/# footnote-7 (2019-10-01) (2019)

features which are unreasonable, morally objectionable, or unjustifiable towards protected groups.

- **Identification** – Bias detection techniques are implemented during the design, training, validation and deployment of our AI models.
  Our AI case-finding algorithm behaves fairly and equally independently of gender, deprivation index (IMD) or ethnicity.

- **Action** – Implement bias mitigation techniques, improve data collection/processing, include relevant features.
  We ensure a training dataset includes the index of multiple deprivation (IMD).

If identifying and mitigating internal bias is widely known and defined and many tools have been developed in order to tackle it,[7] the same cannot be said on external bias.

External bias in healthcare is directly linked to unmet needs. Groups who are not receiving appropriate level of services, either due to lack of supply (insufficient services available, or delivered in an appropriate manner), or simple lack of demand (service available, but not being used), can lead to poorer outcomes and exacerbate health inequality.[8]

At HN, we define external bias in our AI algorithms as any unwarranted variation that might arise from having insufficient or incomplete data about the process or outcome the AI is trying to model.

Patients might experience unwarranted variation in the health services that they are currently offered, prompting them to continuously seek care elsewhere in the health system. This phenomenon can go unnoticed only if a single data source, for example primary care data, was used.

Our experience from recruiting 1,800 patients in our randomised controlled trial (RCT) in tackling avoidable urgent and emergency care, has shown that a large proportion of patients use their local A&E department instead of the GP. Had we used only primary care data to identify these patients, we'd have most likely missed them as they wouldn't have had enough contact points with their GP.

While the above scenario assumes that patient requirements for health services are apparent, they are not being measured, resulting in no data to model. There is another side to the problem. There is a widespread concern that the patients who are under-utilising healthcare resources, relative to their need, will suddenly become high-intensity users. These patients will also be missed or ignored by the AI model, as they may not be represented in the datasets to train our model.

[7] Jakub Wiśniewski and Przemysław Biecek, "Flexible tool for bias detection, visualization, and mitigation.", (2020), GitHub repository, https://https://github.com/ModelOriented/fairmodels/
[8] Aragon Aragon, MJM et al. "Defining and measuring unmet need to guide healthcare funding: identifying and filling the gaps." *CHE Research Paper*, no. 141, Centre for Health Economics, University of York, York, (2017) pp. 1-46

Why can't we design health services based on a more holistic view of patient's care utilisation across sectors and using that information to focus on their needs and understanding their health levels? We believe that the system is slowly moving in that direction.

So where do you think is the largest potential bias?

"The main concern on our current AI model is under-utilisers of secondary care"

Due to the nature and well-defined understanding of internal bias, it's safe to assume that current AI models being developed or deployed will have gone through some type of internal bias checklist.

As mentioned, we believe tackling external bias, or unmet needs, is the largest and most challenging problem, as it requires more data from different care sectors, infrastructure, appropriate security standards and the information governance in place.

The main concern on our current AI model is under-utilisers of secondary care. Due to the lack of data points, the prediction model will not identify these group of patients ahead of them becoming high-intensity users.

Under-utilisers of health and care, either due to supply (lack of services available), or demand (services are available, but they are not accessed), generate less data and will be missed by the AI. This will in turn increase health inequalities and potentially widen existing levels of unmet needs.

You mentioned at the very beginning that we can use AI to address inequalities and reduce gaps – can you elaborate?

"Our approach pays attention to patients' characteristics, medical histories and care utilisation profiles across secondary care and where the data allows, primary care."

Health and care systems have limited resources and a constantly growing care demand which exceeds supply.

By using AI models, we can surface patients who might be at risk of spiralling care demands and adverse health outcomes which has often been caused by chronic inequalities and gaps in service provision and access.

While these AI algorithms are often only based on datasets from one care sector, wider determinants of health such as deprivation, should always be used alongside age, gender

and ethnicity as an absolute minimum. This ensures an effective, but by no means, definitive way to identify the patient characteristics and consumption patterns that are associated with potentially growing levels of unmet needs.

The topic of unmet needs has often been mentioned in relation to programmes targeting high-intensity users of secondary care. These programmes identify patients based on a fixed threshold of A&E attendances in a given period (E.g. the top 50 high-intensity users in the past year).

Current health high-intensity programmes might take weeks to target suitable patients and often focus on delivering care to those who already have established patterns of high-intensity usage, rather than those who need it most.

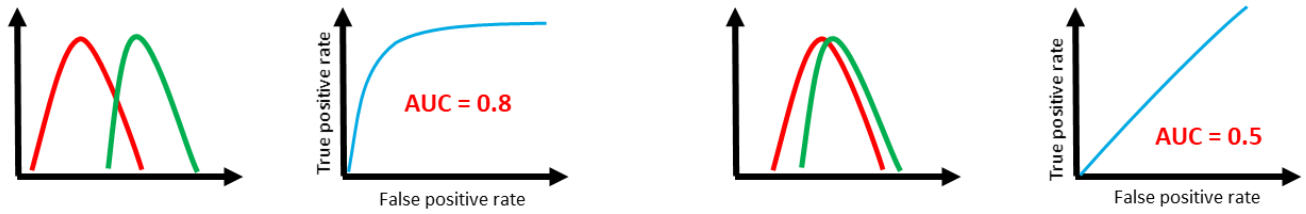
This is in stark contrast with HN – we use data from the A&E, inpatients and outpatients to identify patients predicted to be at the highest risk of becoming high-intensity users.

Our approach pays attention to patients' characteristics, medical histories and care utilisation profiles across secondary care and where the data allows, primary care.

We examine the patient's patterns in presenting at the care provider at a given time point (E.g., the last 12 months) and how these patterns change over time. Our deployed models achieve a minimum c-statistic (area under the curve – AUC) of 0.8, which is a measure of the quality of prediction. This means that for every 100 patients we identify, there is an 80% likelihood that the model correctly distinguishes which patients are predicted to be high-utilisers and those who are not.

The following example demonstrates an AI model with an AUC of 0.5, is as good as a toss of a coin.

A very quick visualisation of two AI models



A fairly useful model **A model as good as a coin toss**

Developing good quality AI models that are deployed daily and tailored to the frontline service, allows for much earlier intervention and maximises the positive benefits for the patient and health system.

We see AI as a great way to capitalise on the vast streams of NHS data and make a significant difference to frontline services that are being developed. Large trials and high

intensity programmes that are not data or AI driven, have shown limited results[9]. AI algorithms, however, should be used wisely - it certainly isn't a panacea and has its limitations (please see the appendix for further information).

What are the future plans to implement the Fairness framework within HN? What's been your progress and how far do you believe you can get? How much can this impact the system as a whole?

"HN is in a good place to advance this work by partnering with NHS Digital, York University and Lloyd's Register Foundation, to develop guidance and frameworks on AI Assurance, Safety and Fairness which is also relatively new research"

We have developed the internal fairness checklist to ensure all our models are unbiased when it comes to the data they use.

AI is a clear function of the data that it is built on and as discussed earlier, the key challenge is to undress external biases or unmet needs. We don't yet have the security, technological and information governance infrastructure on a national scale to achieve this yet.

I have been fortunate to work at HN where we have been progressing the development of a solution hosted in the private cloud that allows scaling and rapid deployment of AI models while adhering to industry-leading security and information governance standards.

Most AI companies focus on alternative measures of raw performance metrics of the model (classification accuracy, mean absolute error, areas under the curve and so on). Those measures are important, but our focus is different.

Our aim is to identify patients who are deteriorating but health demand isn't being surfaced (either at the GP or hospital level), this won't be picked up by our current models. This will be achieved by combining multiple datasets, such as primary and secondary care and comparing consumption patterns.

By combining multiple data sources of consumption – primary, secondary, ambulance, local authority – we will have sufficient data points to surface unmet needs earlier. For example, patients with low primary care activity patterns, but some ambulance and secondary care activity might reveal rising risks.

AI models built on these rich datasets will benefit those unserved by the system – those on the wrong side of the inverse care law. We see AI as a way of surfacing and addressing health inequalities baked into standard approaches.

[9] Steventon, Adam et al. "Effect of telephone health coaching (Birmingham OwnHealth) on hospital use and associated costs: cohort study with matched controls" *BMJ* (2013); 347: f4585 (https://www.bmj.com/content/347/bmj.f4585)

HN is in a good path to advance this work by partnering with NHS Digital, York University and Lloyd's Register Foundation, to develop guidance and frameworks on AI Assurance, Safety and Fairness which is also relatively new research.[10] HN is also actively working with Safe Havens across the UK that have access to linked datasets to further test its AI models and Fairness framework.

Does fairness only stop with the data that you are modelling and highlighting any biases that might exist within it? Does it not leave too big a gap for the health system to close?

"HN's aim is to create an integrated approach, where technology and frontline clinical staff can effectively collaborate to deliver the best care for patients."

Imagine, a retired asthmatic with a recent stroke living in a deprived area with little access to health services other than their local A&E department.

They are recovering well, but the panic attacks have recently started again which they don't know whether to attribute to their asthma or recent stroke. They are not aware of any other services to rely upon for responsive and trustworthy advice and treatment except the A&E department and so commence the first of a sequence of calls to 999 over the next few weeks.

As the patient makes his fourth attendance (illustrative example, see footnote)[11] to the A&E department, the AI model identifies them as at risk of becoming a high-cost high need patient.

So now what? Are we going to alert the increasingly busy staff at the hospital? Or is it going to be a letter sent to the GP that the patient visited more than a year ago?

For us at HN, using AI as identification of any biases or gaps in care is an important step, but there needs to be an effector of the AI to fundamentally change the culture, behaviours and incentives of patients[12] and improve the efficiency of the health and care system.

HN's aim is to create an integrated approach, where technology and frontline clinical staff can effectively collaborate to deliver the best care for patients.

We believe by doing so we bring two complementary factors; years of clinical practice working with patients and data-driven caseloads ensuring maximum impact from a finite

[10] https://www.hn-company.co.uk/ethics-and-equality/
[11] Illustrative example – In reality, the number of attendances which will trigger the patient's identification vary from one patient to the other. It will depend on factors such as patient's medical history, past attendances and personal characteristics.
[12] Emanuel, Ezekiel J, and Robert M Wachter. "Artificial Intelligence in Health Care: Will the Value Match the Hype?" *JAMA* vol. 321,23 (2019): 2281-2282. doi:10.1001/jama.2019.4914

workforce by using the scalability, consistency and quality insights that are a by-product of good technology.

This is especially relevant as the 1% of patients who consume 53% of all non-elective (NEL) bed days are a very dynamic and transient population, where only 80% of the group's membership will change the following year. Following that approach HN has achieved a significant decrease in non-elective admissions by 24%, number of A&E attendances by 34% and elective admissions by 24% as well as increased patient-reported outcomes (PROMS).

It is well evidenced that coaching without AI doesn't work[8] and that AI without dedicated resources doesn't necessarily achieve the intended improvements in system efficiencies[13].

What enablers would you need to achieve these plans?

"We strongly see patient groups as a key stakeholder to consult with on how their data is used"

Interoperability remains elusive and core digital tools, especially electronic health records are vastly underutilised.

The unfortunate fact is that despite the vast advancements that have been made, healthcare data is mostly collected, stored and moved digitally without much capacity or long-term thinking to structure, analyse and model it in a meaningful way.

The relatively slow adoption of GDPR principles into system-wide frameworks has led to misinterpretation and a sluggishness in developing technology that has privacy and security embedded in its design.

This has ultimately led to a very diverse landscape, where each player has vastly different digital, security and information governance maturity. Projects are usually put to a complete halt by outdated technologies and standards. The system as a whole needs to elevate its digitisation. We'd welcome guidance from national bodies to set the direction and strategies.

This flows well into another point of interest – businesses need to actively consult with regulators and foster an open discussion. There should be a three-way relationship, where the patient is placed at the centre.

Regulators and national bodies have a vital role in setting minimum standards, policies, frameworks, especially in such a novel field as AI Fairness, Ethics and Safety. It is, however, ultimately the patient who will be most impacted by any decisions no matter how far up the chain they have been made.

[13] Snooks, Helen, et al. "Predictive risk stratification model: a randomised stepped-wedge trial in primary care (PRISMATIC)." *NIHR Journals Library*, January 2018. doi:10.3310/hsdr06010

We strongly see patient groups as a key stakeholder to consult with on how their data is used by AI, what AI products will they like to benefit from and how they define topics such as AI Fairness and Ethics.

We have covered quite a bit of material so far, how does this all come together for HN in an NHS context?

"Our main aim for the future is to uncover patients who are deteriorating but their health demand isn't being surfaced"

High-intensity users are transient, meaning this group changes its membership significantly over time. They have multimorbidity and show increasing levels of demand for services. AI can identify these patients at the right stage – early enough to intervene and make a difference. This is an integrated approach, linking the AI to the clinical assessment by nurses and the actual patient facing coaching intervention.

Building AI models responsibly and diligently is an important step that ensures algorithms are an accurate representation of the system that our data is trying to describe and approximate.

This is how we address internal bias in AI models and often is the only bias that the data scientist can completely remove with the data they are given. However, this sets a strong foundation to highlight any wider societal biases that might be embedded in the systems we've just modelled with our AI.

Our main aim for the future is to uncover patients who are deteriorating but their health demand isn't being surfaced (either at the GP or hospital level). This won't be picked up by our current models.

This will be achieved by combining multiple datasets, interoperable and scalable technology, and actively working with partners across the UK to test our modelling approach and fairness framework.

We strongly believe that a service that combines an effective AI with the patient facing experience of frontline clinical staff can become part of NHS standard routine care. This will improve patient outcomes, create more sustainable care and foster an NHS system that offers fair access and better outcomes for its patients. The NHS is one of the most heralded public institutions and we shouldn't forget this very principle that brought it about 72 years ago.

Finally, the patient will always have the final say. While AI models can be effective at identifying at-risk patients, a "new bias" can appear as patients refuse to use services and therefore, are at heightened risk of increased bias and health inequality.

This can be minimised by continuous engagement with citizen and patient groups, particularly focusing on the underserved who achieve the most unequal outcomes.

Ultimately, we are designing services to suit the population needs and it's our responsibility to be aware, help identify and take decisive actions to close the inequality gap.

About HN

HN is a healthcare company that delivers AI guided case-finding, remote monitoring, clinical coaching and virtual ward solutions to the NHS.

Since our UK launch in 2015, we have developed into an award-winning NHS partner and, from March 2020, have been supported by the NHS Innovation Accelerator to scale nationally.

We provide practical applications of population health management, going beyond just identifying high-cost, high-need patients and actually intervening to support them to improve their health outcomes and reduce their care consumption. hn-company.co.uk

About Sara Reis

Sara has a PhD in Medical Imaging and Computing from the University College London. Originally from Portugal, Sara studied BSc and MSc Biomedical Engineering and Biophysics, at the University of Lisbon.
Prior to joining HN, Sara worked at The Royal Marsden NHS Foundation Trust and Parliamentary Digital Service.

Contact

For further information, please contact Angela Bambridge.

Follow us on LinkedIn and Twitter.

Appendix:
Internal AI fairness checklist

- ❑ Determine the goal of the AI model, review it with stakeholders, including protected groups.
- ❑ Ensure the model is related to the desired patient outcome and can be integrated into clinical workflows.
- ❑ Decide what groups to classify as protected and protective attributes to use.

- ❑ Ensure that protected group can be identified (e.g. there is data on ethnicity, gender, IMD, age).
- ❑ Assess whether the protected group is represented adequately in terms of numbers and data quality.

Ensure that the model is equally accurate for patients in the protected and nonprotected groups but evaluating metrics such as:

- ❑ Equal sensitivity
- ❑ Equal sensitivity and specificity
- ❑ Equal positive predictive value

- ❑ Ensure that fair model performance translates into deployment and there is no unwarranted discrepancy in service provision/receipt