Case study – Health Inequalities on the Elective Waiting List and factors leading to Emergency Admission

Future of analysis in the brave new world of digital

Simon Wellesley-Miller

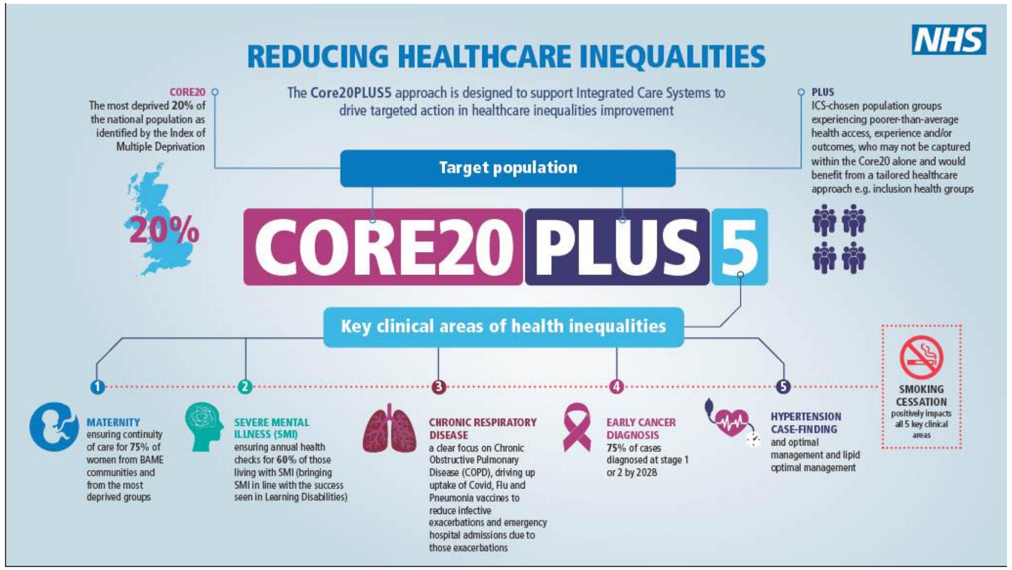
Senior Analytical Manager
Performance Analysis Team -South West
Elective Recovery Insights Team – National
NHS England
MSc Graduate
Healthcare Data Science
University of Exeter

South West Lead
Association of Professional Healthcare Analysts
Fellow
NHS-R Community
SPC Analytical Champion
NHSE Making Data Count

NHS 75 England

Project objectives

- To understand where there has been significant impacts on health inequalities in the South West, with respect to accessing acute services during Covid 19.
- Outputs of the project will support future areas of focus to address the gap in ensuring accessing services is equitable across the South West.
- Analysis to support this project should be automated and available to enable continuation post MSc and become part of business as usual.
- Look at national schemes and how this work can support national planning



Definitions

Health inequalities are systematic differences in the health status of different population groups. These inequities have significant social and economic costs both to individuals and societies. (1)

Waiting list - A list of patients waiting to receive a consultative, assessment, diagnosis, care or treatment activity from an organisation.
The list is maintained for an identified care professional or service within an organisation. (2)

Emergency admission - Patients admitted to hospital when admission is unpredictable and at short notice because of clinical need. (2)

Watchdog warns NHS on safety for patients stuck on waiting lists

Tories push PM to act faster on waiting lists

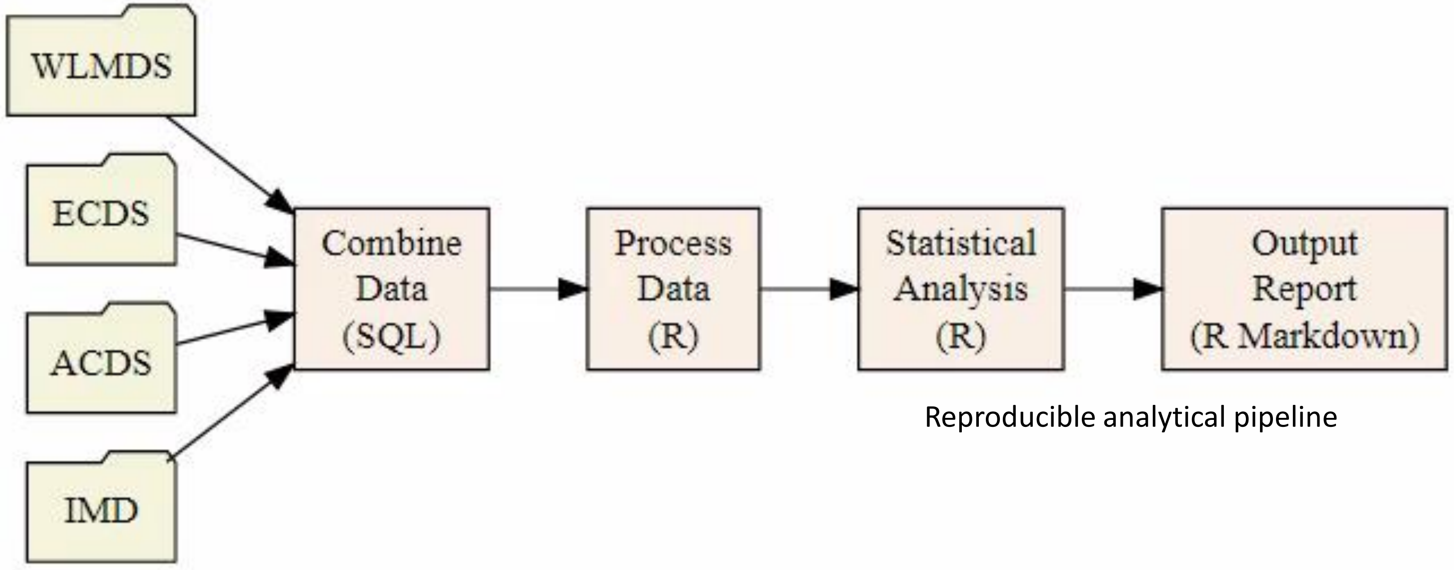
Almost 200,000 patients now waiting at least a year for NHS operations

Obtaining data

- Main dataset taken from waiting list minimum dataset (WLMDS)
- Emergency care data set (ECDS)
- Admissions data set (ACDS)
- Public Health data for Index of multiple deprivation (IMD) [4]
- Public Health data for costal community identification
- Requires a patient level dataset with patient identifiable information
- Appropriate information governance

Data wrangling

- SQL to pull WLMDS with required features
- Linking WLMDS with EDCS and ADCS in SQL across treatment functions
- Pulling public health datasets into analysis
- Calculating local IMD deciles to be able to compare systems
- Collating data into R for analysis and development of pipeline
- Data quality, missing data, assumptions
- Output into automated report



Reproducible analytical pipeline

What does a waiting list look like?

- There is not a single waiting list – is by treatment function and by organisation
- Looking at South West Region, over 50+ organisations, each its own independent lists across 150+ treatment functions
- Waiting lists are far from normally distributed
- There is also an issue that patients that are waiting may become an emergency admission.
- Linking the waiting list dataset to ED attendances and Emergency Admission data, train a model to identify features that may lead to emergency admission

There are 0 patient outliers over 150 weeks, and 0 patients waiting over 300 weeks.
The longest wait is 97 weeks.
Outliers over 200 days have been removed from the graph to maintain scale.

What is a health inequality within a waiting list?

- These approaches have flaws as they perhaps simply describe the patients
- Calculate local IMD for comparison
- IMD contains an element of health outcome – risk of self correlation
- Quantitative or descriptive techniques
- Comparison to rates by expected prevalence

	Hospital X (N=2361)	Hospital X (N=1096)	Overall (N=3457)
Stated gender			
Male	1280 (54.2%)	520 (47.4%)	1800 (52.1%)
Female	1081 (45.8%)	576 (52.6%)	1657 (47.9%)
Other	0 (0%)	0 (0%)	0 (0%)
Not stated	0 (0%)	0 (0%)	0 (0%)
Not known	0 (0%)	0 (0%)	0 (0%)
Ethnicity code			
any_asian	7 (0.3%)	3 (0.3%)	10 (0.3%)
any_black	3 (0.1%)	1 (0.1%)	4 (0.1%)
any_mixed	2 (0.1%)	1 (0.1%)	3 (0.1%)
any_other	9 (0.4%)	0 (0%)	9 (0.3%)
any_white	1788 (75.7%)	583 (53.2%)	2371 (68.6%)
not_known	552 (23.4%)	508 (46.4%)	1060 (30.7%)
Age			
0 to 19	32 (1.4%)	26 (2.4%)	58 (1.7%)
20 to 39	196 (8.3%)	93 (8.5%)	289 (8.4%)
40 to 59	524 (22.2%)	201 (18.3%)	725 (21.0%)
60 to 79	1246 (52.8%)	540 (49.3%)	1786 (51.7%)
80+	363 (15.4%)	236 (21.5%)	599 (17.3%)
IMD quintile			
1-2	521 (22.1%)	130 (11.9%)	651 (18.8%)
3-4	504 (21.3%)	218 (19.9%)	722 (20.9%)
5-6	494 (20.9%)	265 (24.2%)	759 (22.0%)
7-8	366 (15.5%)	294 (26.8%)	660 (19.1%)
9-10	471 (19.9%)	184 (16.8%)	655 (18.9%)
Missing	5 (0.2%)	5 (0.5%)	10 (0.3%)
Waiting list type			
Inpatient	212 (9.0%)	25 (2.3%)	237 (6.9%)
Outpatient	2149 (91.0%)	1071 (97.7%)	3220 (93.1%)
Weeks waiting			
0-18 weeks	1458 (61.8%)	718 (65.5%)	2176 (62.9%)
18-26 weeks	401 (17.0%)	346 (31.6%)	747 (21.6%)
26-40 weeks	355 (15.0%)	32 (2.9%)	387 (11.2%)
40-52 weeks	94 (4.0%)	0 (0%)	94 (2.7%)
52-78 weeks	46 (1.9%)	0 (0%)	46 (1.3%)
78-104 weeks	7 (0.3%)	0 (0%)	7 (0.2%)
104+ weeks	0 (0%)	0 (0%)	0 (0%)

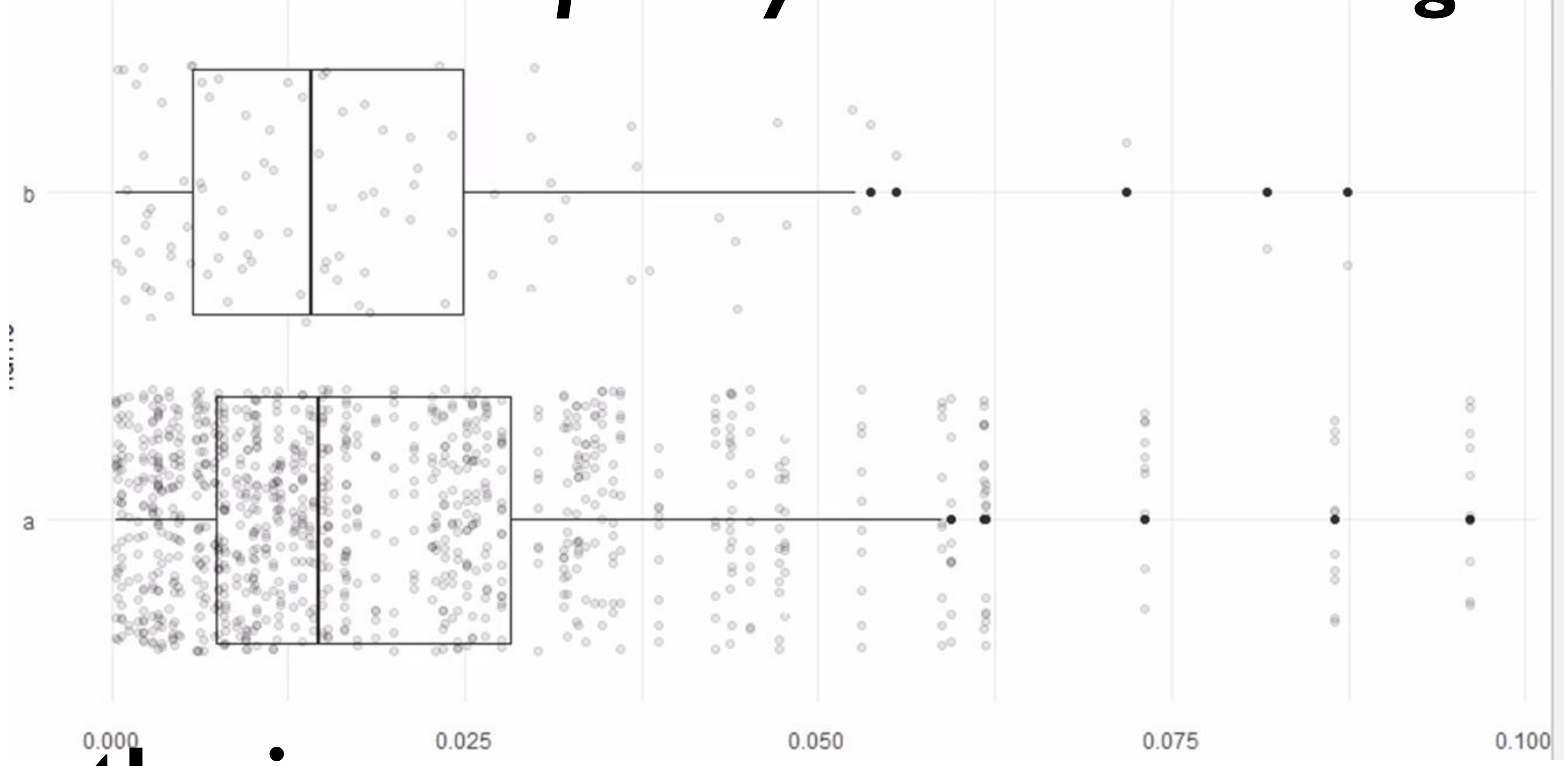
What is a health inequality within a waiting list?

- What if there was a 95% to 5% percent split by gender on a waiting list?
- Would we consider that to be a health inequality?
- Conditions and disease do have a demographic prevalence
- Literature review identified some previous work to look at distribution of waits, this mainly looked at the difference in mean and median, or via a survival analysis, many others tried to link to prevalence rates [5] [6]

My hypothesis

If a waiting list is equitable, then the distribution of patient cohorts within the overall distribution would be the same

What is a health inequality within a waiting list?



My hypothesis

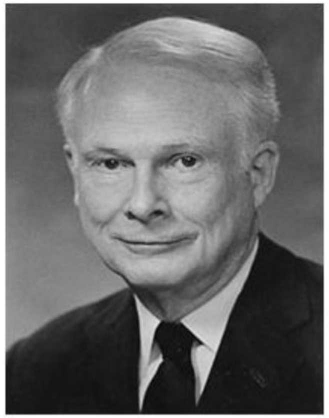
If a waiting list is equitable, then the distribution of patient cohorts within the overall distribution would be the same

What is a health inequality within a waiting list?

- Identified a statistical test to compare distributions of cohorts within a waiting list
- Utilises a rank sum approach to compare sample significance across non parametric distributions [7]
- Created analytical pipeline to examine waiting lists by treatment code, by organisation and by health inequality demographic
- Runs at scale and only return those that reach statistical significance with a p value < 0.05
- Where statistical significance identified run further pairwise comparisons to identify significance between pairs within cohorts
- Dynamically create exception report for analysis and discussion
- Can be run for any treatment function (with slight tweaks could be run for other regions)

William Kruskal

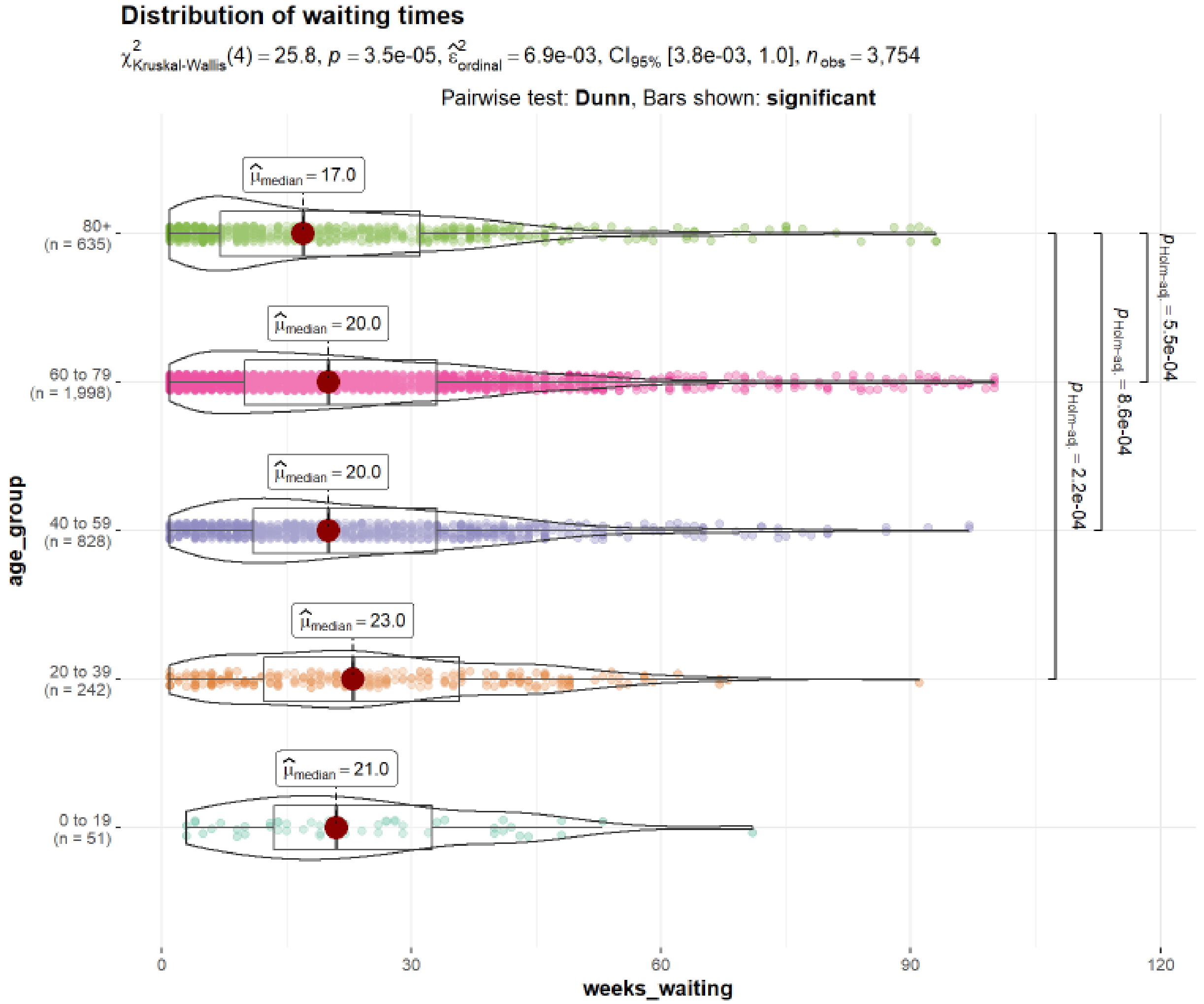
Wilson Allen Wallis



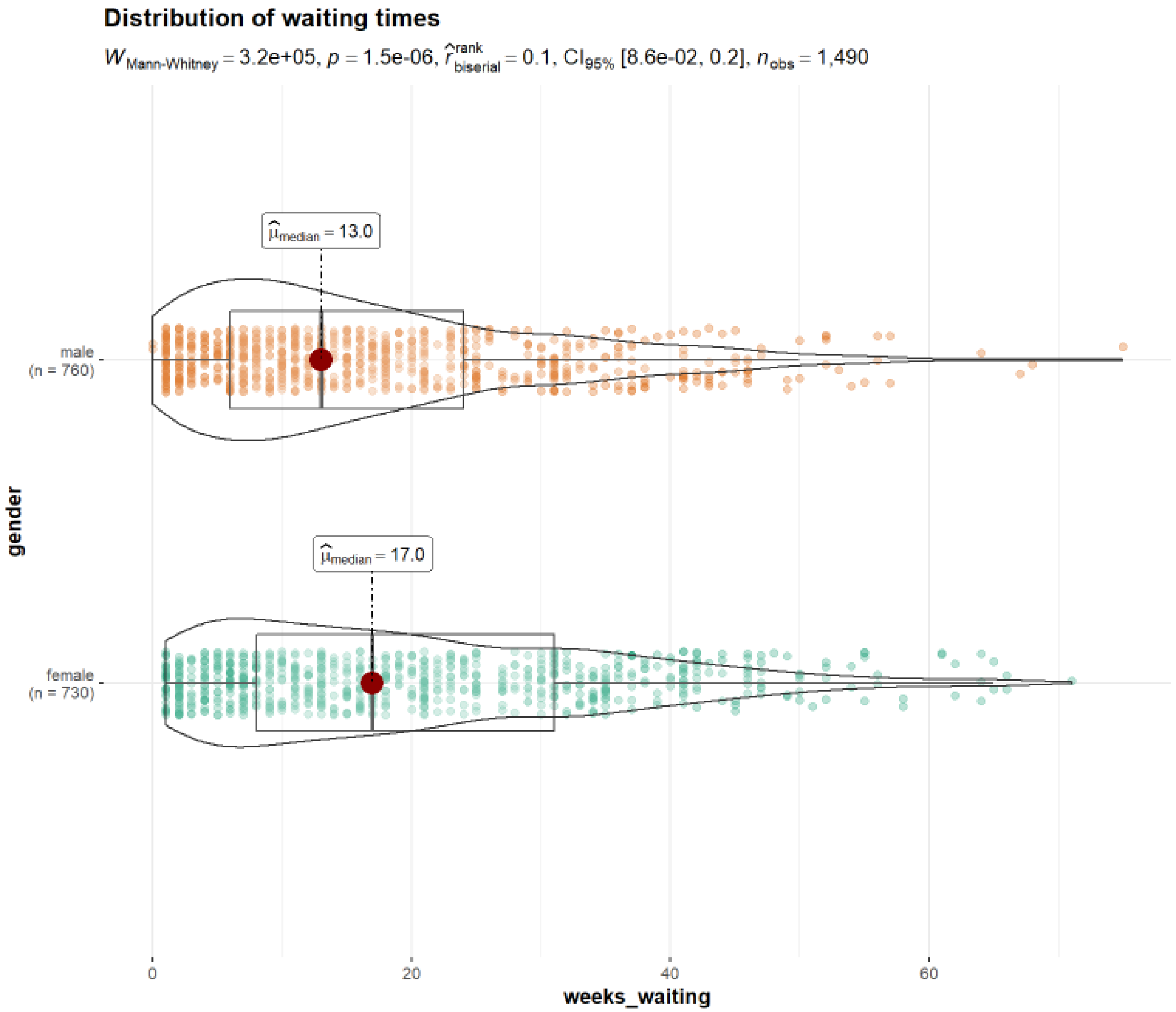
$$H = \frac{12}{N(N+1)}\left(\sum \frac{R_i^2}{n_i}\right) - 3(N+1) \text{ where,}$$

n_i = the total number of points in the i^{th} sample
R_i = the rank sums the i^{th} sample
N = the total number of sample points

What is a health inequality within a waiting list?



There are 3,754 patients waiting at Royal Cornwall (Treliske).
The longest median wait by group is 23 weeks and the shortest is 17 weeks.
A difference of 6 weeks.Overall, the median wait is 20 weeks.
Waiting lists are not normally distributed and so this is only a crude indication of central tendancy.
The most significant difference between groups is between age group 20 to 39 and age group 80+



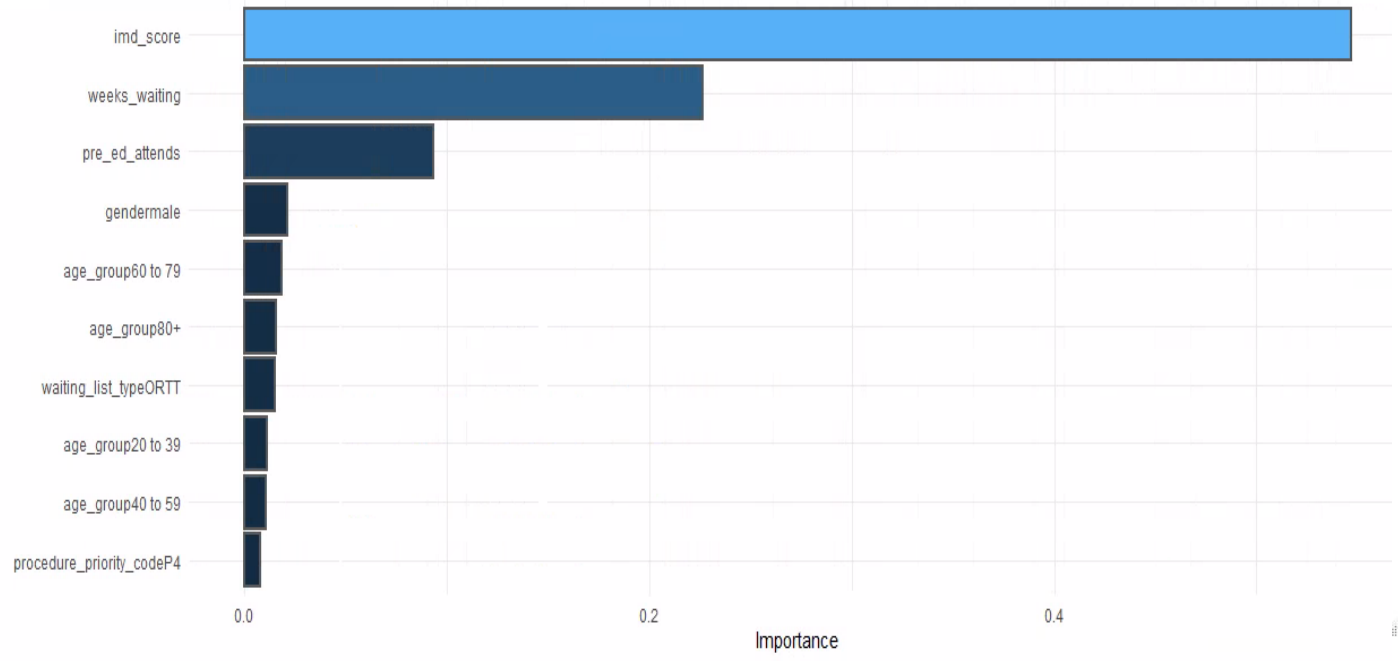
There are 1,490 patients waiting at North Devon District.
The longest median wait by group is 17 weeks and the shortest is 13 weeks.
A difference of 4 weeks.Overall, the median wait is 15 weeks.
Waiting lists are not normally distributed and so this is only a crude indication of central tendancy.
There are only two groups within this analysis and so only a pairwise comparison has been carried out.

What are the risks leading to emergency admission?

- Chose to train a machine learning XGBoost [8] model
- Flexible
- In built capacity for missing values
- Good performance
- Interpretability of results, can give explainable feature importance
- Rather than specific prediction, determining feature importance is the goal
- Useful for risk stratification of patients and potential support of clinical triage

XG Boost Model

- No need to scale or centre data
- Creates a decision tree
- Utilises a space filling parameter grid to try various hyper parameters in the model itself and selects best AOC
- Uses 10 fold cross-validation using stratification for tuning the model
- Returns interpretable feature importance



Results

- Reproducible analytical pipeline – code to be published
- Feedback from operational leads, clinicians and analysts
- Does identify elements of inequality
- Risk factors – hardly breaking new ground but good to support intuition with data

Discussion

- Not all inequalities picked up are necessarily an issue
- Element of clinical decision making and ongoing risk stratification
- Not all disparities are clinically meaningful – there may clinical need to prioritise treatment
- Data quality is poor – especially around ethnicity
- The pipeline is built so that additional features can easily be added to the analysis, things such as serious mental illness (SMI), learning disability (LD), long term conditions (LTC) etc
- IMD is very crude and covers an area of up to 3,000 people (better measures being developed)

Conclusion

- So much time wrangling the data
- Report is just a snapshot
- Additional features to be added, serious mental illness, learning disability, long term conditions, coastal status etc
- More features, better analysis, better model!
- Code written to be a simple adaption to add new features – have transferred to UDAL to be one workflow
  - Choose a site and one click to produce report
  - Can be run with local WLMDS
- Utilised at local level to stratify risk and support clinical decision making
- Starting to be used at a regional level to support strategic decision making and inequalities oversight

End of my MSc Project

Continuing work

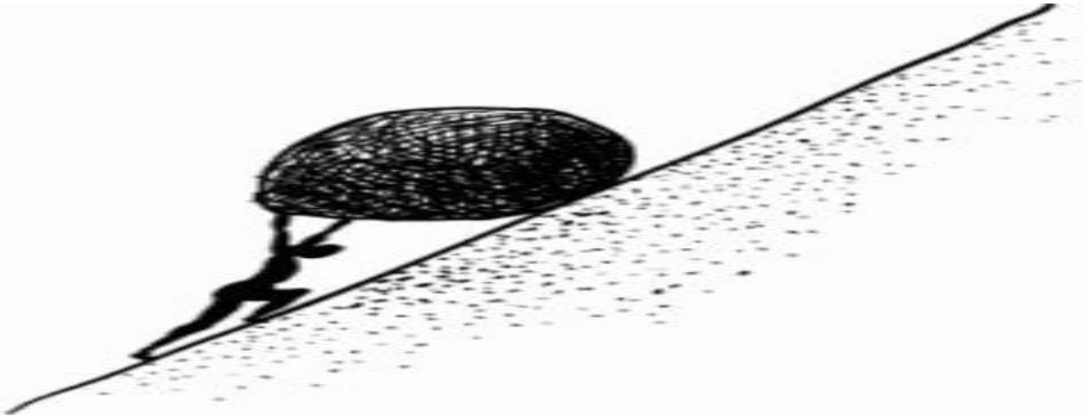
- Really fortunate to have opportunity to work with ERIC and to continue work with wider team of data engineers and scientists.

- Current work is linking to master patient index to improve data quality - clearer picture of who is waiting

- Additional features from master patient index and stratification datasets – additional inequalities to add such as MH, LD, co morbidities, care home etc etc

- Developed robust statistical analysis of those waiting longest. Are certain demographics being adversely affected?

- Work is conducted at massive scale across treatment functions by ICBs and potentially provider level.



Continuing work

- Conversion of raw numbers into population rates built from master patient index at LSOA level and building up. Will allow standardised comparison of rates by population.

- This will enable comparisons across systems to identity genuine statistical outliers.

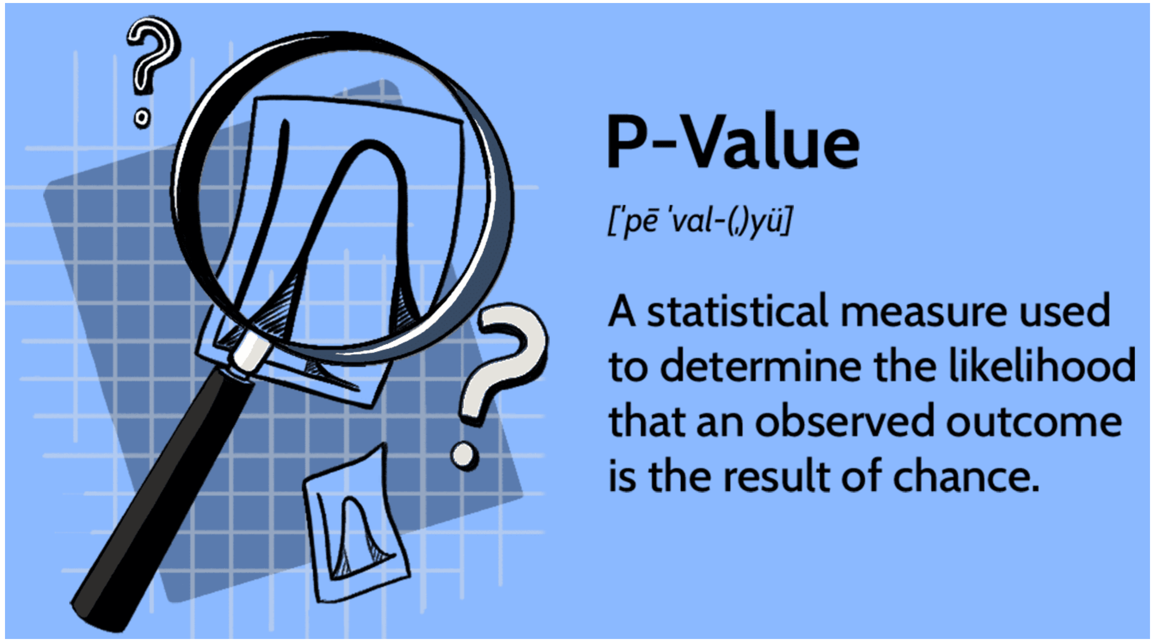
- Looking to compare more wider public health measures to compare prevalence on lists

- Machine Learning cluster analysis of long waits – find out truly what are the most common profiles for patients waiting – to build insight and design targeted interventions

- Time series analysis – looking at statistical change over time

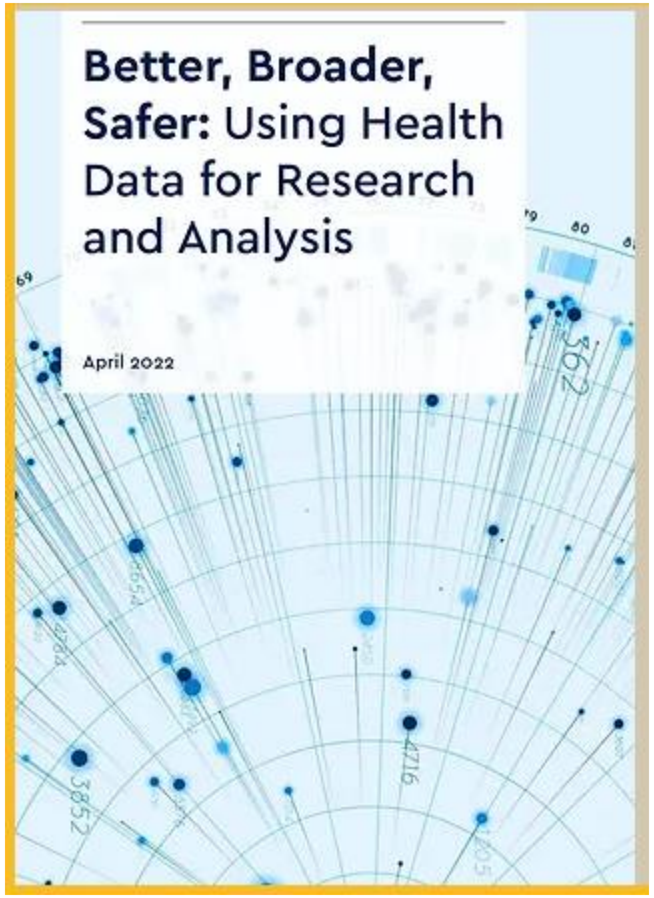
Continuing work

- Process mining – extract patient pathways from data to identify bottlenecks
- Demand and capacity modelling to predict future waiting list
- Identify key areas of concern and apply targeted interventions
- Evaluate targeted interventions



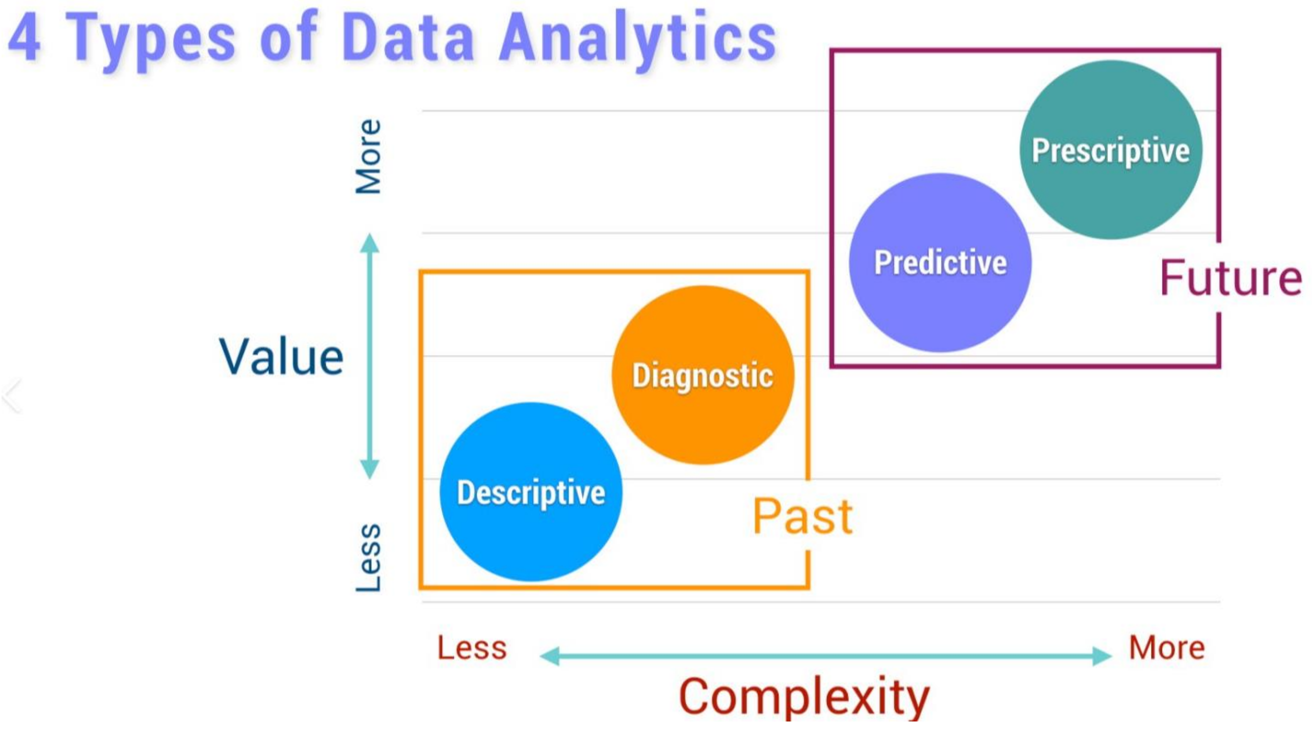
Future of analytics

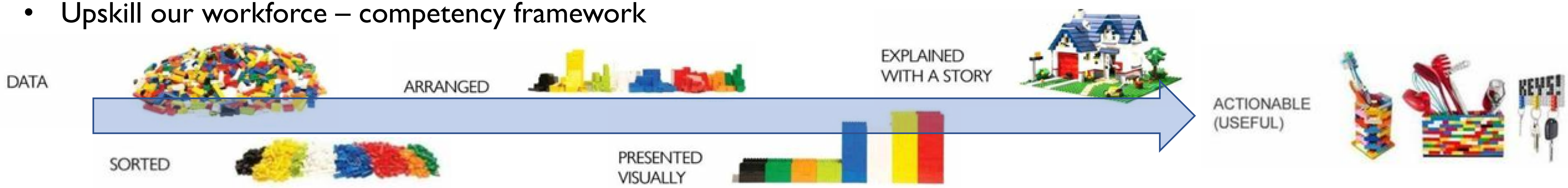
- Really exciting
- Blue widget counting – ***NO!***
- Change the analytical questions and conversations
- Stop analysists producing data and get them producing analysis
- New ways of working – new tools
- Accept Goldacre review
- Default is to share code



How digital can help

- Allow analysts tools and infrastructure to do this cool stuff
- We need data linkages and environments
- Change the analytical questions and conversations
- Accept Goldacre review
- Default is to share code
- Stop analysists producing data and get them producing analysis
- Upskill our workforce – competency framework





Any questions?

Thank you!

References

[1] https://www.who.int/news-room/facts-in-pictures/detail/health-inequities-and-their-causes
[2] NHS Data Model and dictionary - https://www.datadictionary.nhs.uk/classes/waiting_list.html
[3] https://www.england.nhs.uk/about/equality/equality-hub/national-healthcare-inequalities-improvement-programme/core20plus5/
[4] https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019
[5] Dimakou, S., Dimakou, O. & Basso, H.S. Waiting time distribution in public health care: empirics and theory. *Health Econ Rev* **5,** 25 (2015).
[6] Dimakou, S., Parkin, D., Devlin, N. et al. Identifying the impact of government targets on waiting times in the NHS. Health Care Manag Sci 12, 1 (2009)
[7] Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. Journal of the American Statistical Association, 47, 583-621.
[8] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). New York, NY, USA: ACM.

Thanks to

Lucy Wellesley-Miller, Vicki Cruize, Tobin Savage, Jo Mcsweeny, Martin Huscroft, Julia Gonzalez-Esquerre, Fiona Leat, Prof. Martin Pitt, Prof. William Henley and many others for support with the project