PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Feasibility of a hyper-acute stroke unit model of care across England. A modelling analysis.
AUTHORS	Allen, Michael; Pearn, Kerry; Villeneuve, Emma; Monks, T; Stein, Ken; James, Martin

VERSION 1 – REVIEW

REVIEWER	charles wolfe KCL London UK
REVIEW RETURNED	27-Jun-2017

GENERAL COMMENTS	This is a paper regarding reconfiguration of services in England using a heuristic genetic algorithm. Such modelling is potentially useful to scenario set for discussions on specific changes that can be made at for example a regional level. 1. The authors appear to take routine HES data and small area statistics codes with travel times to model potential numbers of stroke ? hyper acute units required. The modelling is based on guidance on travel times and potential to benefit from hyper acute treatments. As such it reads like a modelling exercise but without any context of patient preference, definition of units and step down units, or training/staffing feasibility. These issues need more description in the discussion at least but preferably in the modelling and background. 2. The authors need to reflect on changes in the UK that have based reconfigurations on regional and small national (Scotland) numbers. Unless these data are used for example for a national strategy they would not be of relevance to a regional/STP model of modelling really. 3. The HES data does not include stroke mimics that could inflate the numbers by a third in the acute phase and this requires modelling. 4.Throughout it is unclear what definitions of unit are being advocated. In London there are Hyperacute units for a limited number of days-3ish with step down to stroke units. This needs to be clearer in text and in the modelling the number of step down units accounted for and lengths of stay factored in. 5.The estimates of size of units seems devoid of reality really. no account of stroke mimics, range from 600 to 2500 actual strokes.

	The largest in the country presently is around 1200 and this has huge workforce implications. a unit of 2500-2800 with mimics would require two teams etc and this is arguably not a useful model to consider. 6. Have the authors considered the rise in need for beds over next twenty to thirty years of around a third at least due to the ageing population?

REVIEWER	Nathan Proudlove Alliance Manchester Business School University of Manchester UK
REVIEW RETURNED	25-Jul-2017

GENERAL COMMENTS	This is an interesting, potentially practically valuable, and entertaining paper. However the logic is hard to follow in parts, so some clarifications are needed – in the main text and appendix. 1. justify and clarify the overall approach Data preparation: It's not clear what the Signals from Noise tool is for. The website (and name) seems to suggest dashboards and SPC rather than tools to e.g. map patients to LOSA, extract LOSA map coordinates The problem type: You might mention the size of the search space, i.e. a max of 127^2 = 10^38 possible solutions (configuration of units). And so why you need to use a fairly unguided/'brute force' heuristic search approach? (see 2) The approach: Ultimately, I think, you are producing a Pareto front to give you a set of 1,000 to 5,000 non-dominated solutions. From these you are picking out the best and solutions for each of having a total of N=1 to 127 stroke units to examine trade-offs. The key output, then is the Pareto set. It is not computationally feasible to determine which of the 10^38 possible solutions are in the Pareto set – way too many to search through by complete enumeration even with very fast software?! And/or it would (with 13 dimensions) still result in a massive number of solutions? Some non-dominated solutions are still clearly rubbish/unacceptable (but non-dominated) – e.g. having only a few centres with maximum travel distance around 500 minutes. Does this suggest that adding some 'hard' constraints to cull solutions (or forbid creation of solutions in parts of the solution space) might speed things up? Is it justified not to add these constraints because it would slow down the algorithm materially and you want to give the search free reign to generate the data for a full trade-off curve (1-127 centres)? (Fuller comparison with other approaches – set covering? Constraint satisfaction (mathematical programming)? – may perhaps be beyond the scope of what's expected for this journal, more important for an operational research-type journal.)

	2. clarify how the guts of your algorithm works It's not clear how the "seed population" (of 10,000?) is constructed? (sect 2.1). Running the genetic algorithm (2.2) with fewer than the 13 objectives, starting from randomised starting populations of 10,000 and combining (how?) the resulting sets of 1,000 – 5,000 final populations from each run, and culling to (exactly) 10,000 (how?)? After the construction of the seed population (from multiple runs?), is there just one 'run' with all 13 objectives? P.5 l.34 & l.42suggest not all the variables are used? WHICH are used to generate your results? Do the results vary much if starting from a different seed population? Or if run again from the same seed population (with different random numbers driving e.g. mutation)? Roughly how long did a run take?! The performance metric of the multi-objective algorithm is just whether or not a solution is on the Pareto front? (p.20 Step 5) 3. general It's not clear what you mean by predicted vs. actual admissions (p.5 l.19+ & Figure 1). Is this from a pre-genetic algorithm stage, with all N=127 units 'open' to see how well the allocation of whole LSOA populations' stroke cases to the nearest stroke unit (to the geographical centroid) matches with how many people actually turned up at each? But then it says "across the modelled configurations"? Is there a point to/conclusion from this section? The confusion is a bit compounded by going straight on to the results from the genetic algorithm search. It seems a bit counterintuitive to have non-monotonic curves (e.g. fig 2). Is this because there are limited numbers of Pareto front solutions with low numbers of units, giving 'random' kinks at ~N=2 and ~N=16? It would be useful to indicate how many Pareto front solutions there are across various N= levels. It would be useful to say a bit more (in the appendix?) about the tools (and so expertise) used: the LSOA matching of the patient-level data, the Maptitude and Milecharter data, then these providing data files for the main algorithm which was coded (in Python?) and called the genetic algorithm from a downloaded library? Typos/improvement • P.4 l.23: cause loss of disability-adjusted life years?! • P.5 .19: "modelld" • P.6 l.43: 2% would be > 60 mins; how many currently? • P.6 l.43: 2% > 60 minutes vs. p.8 l.6 & p.9: 99% < 60 mins vs p.3 at least 98%...

VERSION 1 – AUTHOR RESPONSE

REVIEWER 1

1. The authors appear to take routine HES data and small area statistics codes with travel times to model potential numbers of stroke ? hyper acute units required. The modelling is based on guidance on travel times and potential to benefit from hyper acute treatments. As such it reads like a modelling exercise but without any context of patient preference, definition of units and step down units, or training/staffing feasibility. These issues need more description in the discussion at least but preferably in the modelling and background.

ANSWER: Discussion, para 1 has been expanded to include more detail regarding definitions of a HASU including reference to national specifications for staffing and training. The methods and discussion clarify that this modelling exercise is focussed on hyper-acute care, rather than step-down care. We appreciate the importance of step-down care, but that phase of work was beyond the scope of this work. We end the conclusion with the point that while our paper focuses on the feasibility of the HASU aspect of care, thought must be given to the follow-on or step-down care.

2. The authors need to reflect on changes in the UK that have based reconfigurations on regional and small national (Scotland) numbers. Unless these data are used for example for a national strategy they would not be of relevance to a regional/STP model of modelling really.

ANSWER: On the recommendation of Reviewer 1, we have added further reflection on the metropolitan/ regional reconfigurations in London and Greater Manchester to paragraph 4 of the Discussion; the generalisability of these urban reconfigurations (in which travel time is much less of a discriminatory factor between centres) to more geographically dispersed provision elsewhere is debated, although we emphasize in the Discussion that modelling solutions can illuminate but cannot dictate regional planning. We agree with Prof Wolfe that planning at the level of an STP cannot occur in isolation, not least because of significant boundary zone issues, and that national level modelling such as ours identifies the need for a larger unit of planning – we make this point in paragraph 4.

3. The HES data does not include stroke mimics that could inflate the numbers by a third in the acute phase and this requires modelling.

ANSWER: We have confined our modelling to the '600 strokes/year' threshold in NHS England planning guidance, but we acknowledge – and address in paragraph 3 of the Discussion – the inflating effect of stroke mimics (Reviewer 1 refers to a figure for mimics derived from our ref 24). In any system conveying FAST-positive patients to a HASU, this proportion is likely to remain constant, so is not itself a decisive or confounding factor in the modelling. However we make specific reference in the Discussion (para 3) to the issue that, once FAST-positive mimics are also included, an upper size limit of 2,500 strokes/year would represent a substantial challenge to the infrastructure and workforce of large HASUs.

4.Throughout it is unclear what definitions of unit are being advocated. In London there are Hyperacute units for a limited number of days-3ish with step down to stroke units. This needs to be clearer in text and in the modelling the number of step down units accounted for and lengths of stay factored in.

ANSWER: Rather as in our response to point 1, we have expanded para 2 of the Introduction and para 1 of the Discussion to include more description of the characteristics and relation between HASUs and step-down (sometimes called acute or post-acute) units. The number of step-down units does not itself affect the size or travel distance to a smaller number of HASUs delivering the initial 72 hours of care, although we accept that it adds to the complexity of systems. Given that our study is restricted to optimising hyperacute provision, we have not included the potential configuration of step-down units in the model.

5.The estimates of size of units seems devoid of reality really. no account of stroke mimics, range from 600 to 2500 actual strokes. the largest in the country presently is around 1200 and this has huge workforce implications. a unit of 2500-2800 with mimics would require two teams etc and this is arguably not a useful model to consider.

ANSWER: We agree with Reviewer 1 that the creation of very large HASUs, larger than the 2,000 confirmed strokes/year currently admitted to Salford Royal following the most recent Greater Manchester reconfiguration, could be an unintended consequence of rationalisation to fewer hyperacute centres, and the disbenefits of such large units (if any) are not well understood. We are grateful to Prof Wolfe for prompting us to explore the issue of the 'upper limit' of size in greater detail, particularly as there is much less observational data about the benefits and disadvantages of very large units in contrast to the data around the desirable minimum size of any unit. We have amended analysis and results to focus on a maximum unit size of 2,000 admissions per year (down from 2,500 in the original maunscript), as that better aligns with the current largest unit size. We have added data to the Results section (last paragraph) that explore the impact on HASU numbers. We have added more data, analysis and a new figure describing the distribution of HASU size to the Results (last paragraph of results and new Figure 5), which demonstrate that in a centralised model containing 75-85 HASUs, there would be about 10% of centres that were larger than confirmed stroke 1500 admissions per year, but still lower than our amended maximum unit size of 2,000 admissions per year. We have also specifically addressed the issue of very large units in an expanded para 2 of the Discussion.

The maximum feasible size of any single HASU will be an even greater consideration with the progressive introduction of mechanical thrombectomy, which will provide further impetus towards the creation of much larger regional hyperacute stroke services under a 'mothership' model – something that we address in the final paragraph of the Discussion.

6. Have the authors considered the rise in need for beds over next twenty to thirty years of around a third at least due to the ageing population?

ANSWER: The projected rise in stroke incidence over the next 20-30 years due to demographic change will be significant in some regions of England, although there is imprecision in these estimates and they will be counterbalanced to an unknown extent by improvements in stroke prevention (Rothwell PM et al, Change in stroke incidence, mortality, case-fatality, severity, and risk factors in Oxfordshire, UK from 1981 to 2004. Lancet, 2004; 363: 1925-1933). Such projections will militate against enforcing the lower recommended limit for admissions too strictly, and may incline planners to err towards a lower 'upper limit' at, say 2,000 stroke admissions/year to allow for such growth. We have included an expanded treatment of this issue in the Discussion, para 3.

REVIEWER 2

Comment: This is an interesting, potentially practically valuable, and entertaining paper.
However the logic is hard to follow in parts, so some clarifications are needed – in the main text and appendix.

ANSWER: Thank you for the comments. As BMJ Open has a general readership we have responded to the more technical questions primarily in the methods appendix. We have tried to keep the method section relatively 'light' for the paper, but with significant further detail added to the appendix.
We have substantial bolstered the appendix following the useful comments here. We would be happy to move more methodological detail to the main paper if that is preferable.

Comment:1a. justify and clarify the overall approach Data preparation: It's not clear what the Signals from Noise tool is for. The website (and name) seems to suggest dashboards and SPC rather than tools to e.g. map patients to LOSA, extract LOSA map coordinates

ANSWER: Yes, I hope we have now clarified. The Signals for Noise tool is indeed usually used for SPC. However Lightfoot also have access to the national HES data set, and we have access to that data through them. But we appreciate mention of the specific tool used is probably unhelpful at best and misleading at worse. So we have amended just to say that we have accessed HES data managed through Lightfoot. We use none of the SPC functionality of the Signals from Noise tool; it is simply used as the front end to be being able to query the national HES data set.

Comment:1b. The problem type: You might mention the size of the search space, i.e. a max of 127^2 = 10^38 possible solutions (configuration of units). And so why you need to use a fairly unguided/'brute force' heuristic search approach? (see 2)

ANSWER: This is now included in section 1 of the method appendix

Comment: 1c. The approach: Ultimately, I think, you are producing a Pareto front to give you a set of 1,000 to 5,000 non-dominated solutions. From these you are picking out the best and solutions for each of having a total of N=1 to 127 stroke units to examine trade-offs.
The key output, then is the Pareto set. It is not computationally feasible to determine which of the 10^38 possible solutions are in the Pareto set – way too many to search through by complete enumeration even with very fast software?! And/or it would (with 13 dimensions) still result in a massive number of solutions?

ANSWER: The primary problem is the number of combinations of hospitals (10^38) is far too large to compute. A secondary problem is that the Pareto front itself can also become unmanageable with large numbers of objectives, but this is managed in the algorithm by selecting from the Pareto front (either randomly or by biasing the set towards those with with a greater crowding distance from their nearest neighbour solution).

Comment: 1d. Some non-dominated solutions are still clearly rubbish/unacceptable (but non-dominated) – e.g. having only a few centres with maximum travel distance around 500 minutes. Does this suggest that adding some 'hard' constraints to cull solutions (or forbid creation of solutions in parts of the solution space) might speed things up? Is it justified not to add these constraints because it would slow down the algorithm materially and you want to give the search free reign to generate the data for a full trade-off curve (1-127 centres)?

ANSWER: We have avoided use of hard constraints. This is in keeping with the biological inspiration of genetic algorithms – that these apparently 'poorer' solutions may be links to better solutions in future generations. An interesting observation in genetic algorithms is that being too strictly 'elitist' can hamper better solutions in the long run. Paralleling biology, genetic diversity is as important as genetic elitism, and all genetic algorithms have ways of maintaining diversity at the expense of elitism. Strict elitism in genetic algorithms leads to faster local optima at the expense of slower better solutions. But we have made a note in the methods section that the algorithms can be speeded up by targeting only a specific number of hospitals, in which case hard cut-offs can be applied to eliminate solutions outside of an acceptable range (though we have not applied that method for the results described here).

Comment: 1e. (Fuller comparison with other approaches – set covering? Constraint satisfaction (mathematical programming)? – may perhaps be beyond the scope of what's expected for this journal, more important for an operational research-type journal.)

ANSWER: We did not do extensive testing against other methods (as that was not our aim). The scale and complexity of the problem (including a look-up from 5 million road travel times for each patient of 32,000 locations to each of up to 126 hospitals) makes it 'non-trivial' for mathematical approaches. We did look to see if a commercial mathematical optimisation engine, 'Gurobi', could identify solutions, but it always ran out of memory on our problem. We did also test a greedy algorithm and hill-climbing methods; they provided good solutions but only worked with weighted objectives which we wanted too avoid. We did test other GA methodologies, particular SPEA-2 and MOED. All performed quite similarly.

Comment:2a. clarify how the guts of your algorithm works It's not clear how the "seed population" (of 10,000?) is constructed? (sect 2.1). Running the genetic algorithm (2.2) with fewer than the 13 objectives, starting from randomised starting populations of 10,000 and combining (how?) the resulting sets of 1,000 – 5,000 final populations from each run, and culling to (exactly) 10,000 (how?)? After the construction of the seed population (from multiple runs?), is there just one 'run' with all 13 objectives? P.5 l.34 & l.42suggest not all the variables are used? WHICH are used to generate your results? Do the results vary much if starting from a different seed population? Or if run again from the same seed population (with different random numbers driving e.g. mutation)?

ANSWER: We've added more detail to the appendix including the mix of reduced objectives used in the initial runs. Details of underlying Genetic Algorithm methodology are given in the on-line appendix in sections 4.1 to 4.7, with specific details of our implementation in section 4.8). We have not quantified variation between runs, but anecdotally as part of each run we visually monitor a key objective (usually the proportion of patients attending a unit that has at least 600 admissions per year that is within 30 minutes of the patients home location). We generally see convergence between runs of less than 0.5 percentage points in this result, suggesting all runs find solutions close to the best solution identified by other runs. If this is of more interest we are happy to do a more quantitative study.

Comment: 2b. Roughly how long did a run take?!

ANSWER: More detail is added to the appendix section 4.9 (close to the end). Generally runs take 2-7 days, but multiple runs can be run in parallel. It is very possible that optimised Genetic Algorithm libraries would run faster; we wrote bespoke code rather than rely on general GA libraries. The Pareto front identification is the main bottleneck, taking about 15 minutes for a population of 10,000 solutions on a single core of a 2GHz processor.

For this work the time spent coding is longer than the runs, so while we try to reasonably optimise the code (e.g. using as many matrix calculations as possible) we don't approach in the same way someone would approach an algorithm that has to be run very frequently.

Comment:2c. The performance metric of the multi-objective algorithm is just whether or not a solution is on the Pareto front? (p.20 Step 5)

ANSWER: Yes. In earlier work (not reported here) we used more weighting of objectives to order Pareto fronts. Experience has taught us that different stakeholders can weight objectives significantly differently. Rather than try to weight objectives during 'optimisation' we now prefer to identify populations of Pareto front solutions that may then be discussed by stakeholders- alternative 'weights' may be discussed applied later. For example, we have found that stakeholders vary in their view of how acceptable it is to improve overall net population benefit at the cost of a minority who may be disadvantaged by the change. These are moral/ethical questions that we prefer to leave open in our modelling. We investigated MOED as an alternative GA methodology that works with weighted scores, but with multiple populations each with a different vector of weights; performance was similar to NSGA-II (the basis of our GA) but we considered NSGA-II to be ideologically simpler and easier to communicate to stakeholders.

Comment:
3. general
3a. It's not clear what you mean by predicted vs. actual admissions (p.5 l.19+ & Figure 1). Is this from a pre-genetic algorithm stage, with all N=127 units 'open' to see how well the allocation of whole LSOA populations' stroke cases to the nearest stroke unit (to the geographical centroid) matches with how many people actually turned up at each?
But then it says "across the modelled configurations"? Is there a point to/conclusion from this section? The confusion is a bit compounded by going straight on to the results from the genetic algorithm search.

ANSWER: Yes, this is testing the assumption that we can predict admissions by assuming that patients attend their closest hospital. We've amended the first paragraph of the results to hopefully make the comparison clearer. The phrase 'across the modelled configurations' was unnecessary at best and confusing at worst. It has been removed. The key point here is whether it is reasonable to assume that patients will be taken to their closest appropriate hospital. The general answer from our analysis is 'yes', but with an acceptance that where hospitals are close together (which occurs in cities, especially London) destination is more flexible. We have not dwelled on this in the discussion because it potentially distracts from the key messages of the paper.

Comment:3b. It seems a bit counterintuitive to have non-monotonic curves (e.g. fig 2). Is this because there are limited numbers of Pareto front solutions with low numbers of units, giving 'random' kinks at ~N=2 and ~N=16? It would be useful to indicate how many Pareto front solutions there are across various N= levels.

ANSWER: There is no expectation that the Pareto front will always be smooth – maximum distance especially changes in steps, often driven by the most remote sparse populations (we find that Cornwall, North Devon, Cumbria, Northumberland and Lincolnshire are commonly the source of the maximum travel time). Non-smooth changes in the Pareto front are most likely to reflect these step-changes. But there is also always the possibility that kinks in the fronts may be sde to the Genetic Algorithm not fully elucidating the full Pareto Front. If ther appears to be a 'reverse' in the front at any particular point, then that is more likely to reflect a point of the front that could be improved with further exploration.

Comment:3c. It would be useful to say a bit more (in the appendix?) about the tools (and so expertise) used: the LSOA matching of the patient-level data, the Maptitude and Milecharter data, then these providing data files for the main algorithm which was coded (in Python?) and called the genetic algorithm from a downloaded library?

ANSWER: Added to relevant sections. I hope with adequate satisfaction, but we're happy to add more detail. A link to the code used and the underlying data (CSVs for admissions by LSOA and travel distances from all LSOAs to all acute hospitals) is provided in the appendix. This is made available under a permissive Open Source license.

Comment:The code is in Python/NumPy – we wrote bespoke genetic algorithm (GA) code rather than using any downloaded GA libraries. This gave us full control/understanding of the code, possibly at the expense of computational efficiency.

The Lightfoot HES data outputs number of patients (given a list of specific diagnostic codes) for each LSOA. We then matched LSOA to a postcode that the GIS could use for estimating travel distances to hospitals. ONS provide the population weighted centroid (Eastings and Northings) for each LSOA. The post office provide the Eastings and Northings for each postcode in the UK; we identified the postcode that was closest (Pythagorean) to each LSOA population-weighted centroid. That postcode was used as the source of the patient for travel time calculation.

Maptitude is a propitiatory Geographical Information System that provides estimated travel times for routes given any two postcodes (MileCharter is an add-in that allows this to be done on a large matrix of start and end points). Our aim is to switch to, and modify as necessary, an Open Source alternative for greater transparency and reproducibility, but further work is needed to validate Open Source alternatives (such as PyRoute working with Open Street Map data).

3c: Typos/improvement
• P.4 l.23: cause loss of disability-adjusted life years?!
• P.5 .19: "modelld"
• P.6 l.43: 2% would be > 60 mins; how many currently?
• P.6 l.43: 2% > 60 minutes vs. p.8 l.6 & p.9: 99% < 60 mins vs p.3 at least 98%...

ANSWER: All corrected. Current population >60 min away is ~0.3%. On the last point ….Thank you for spotting We've made that consistent. With an admissions target range of 600-2000 we can reach 98.6% within 60 minutes with a large group of solutions clustered just either side of 98.5% so I think we've rounded differently in different places. We've now used 98% within 60 minutes (when we're rounding to the nearest percentage point) or the more precise 1.5% greater than 60 minutes when focussing on that figure.

VERSION 2 – REVIEW

REVIEWER	Nathan Proudlove MBS University of Manchester UK No Competing Interest
REVIEW RETURNED	25-Sep-2017

GENERAL COMMENTS	Substantive issues all addressed. Appendix much clearer. 3 very minor corrections (2 typos, and a suggestion), then Accept: p.7, l.13 ”fwer” p.21 l.11 (appendix p.1): “6,000” should be 600 p.21 l.46 I suggest something like “the code contains a bespoke implementation of a genetic algorithm, based on NSGA-II[8], written in …” (I can't see a Response to the Reviewers, but not really necessary here. The role of Lightfoot Solutions is a LITTLE clearer, but would still have been interesting to know more about this - how they provided/acted as the portal top HES and on what terms - presumably commercial? An additional observation, but I didn't raise it last time so need not be reacted to now!, is that the key genetic algorithm references are relatively old - is this still state of the art?!)

VERSION 2 – AUTHOR RESPONSE

From the reviews comments

Suggested changes accepted and manuscripts amended:

3 very minor corrections (2 typos, and a suggestion), then Accept:
p.7, l.13 ”fwer” (correct)
p.21 l.11 (appendix p.1): “6,000” should be 600 (correct)
p.21 l.46 I suggest something like “the code contains a bespoke implementation of a genetic algorithm, based on NSGA-II[8], written in …” (wording change accepted)

(Other comments ask for no response, but hopefully this may be useful).

The Peninsular CLAHRC have agreed access to HES through Lighfoot Solutions for research use of HES. This is paid-for, but at lower than commercial rates. Access is provided by Web Portal (all HES data stays on Lightfoot Solutions servers, and we access only aggregated data).

NSGA-II is still a very popular algorithm. It was beyond the scope of this paper, but we did look at other GAs (such as SPEA-2, and MOED) before selecting NSGA-II.