Evidence synthesis and decision modelling to support complex decisions: stockpiling neuraminidase inhibitors for pandemic influenza usage [version 2; peer review: 2 approved]

Samuel I. Watson¹, Yen-Fu Chen¹, Jonathan S. Nguyen-Van-Tam², Puja R. Myles², Sudhir Venkatesan², Maria Zambon³, Olalekan Uthman¹, Peter J. Chilton⁴, Richard J. Lilford¹

¹Warwick Medical School, University of Warwick, Coventry, CV4 7AL, UK
²School of Medicine, University of Nottingham, Nottingham, NG7 2UH, UK
³Public Health England, London, SE1 8UG, UK
⁴Warwick Business School, University of Warwick, Coventry, CV47AL, UK

Abstract

Objectives: The stockpiling of neuraminidase inhibitor (NAI) antivirals as a defence against pandemic influenza is a significant public health policy decision that must be made despite a lack of conclusive evidence from randomised controlled trials regarding the effectiveness of NAIs on important clinical end points such as mortality. The objective of this study was to determine whether NAIs should be stockpiled for treatment of pandemic influenza on the basis of current evidence.

Methods: A decision model for stockpiling was designed. Data on previous pandemic influenza epidemiology was combined with data on the effectiveness of NAIs in reducing mortality obtained from a recent individual participant meta-analysis using observational data. Evidence synthesis techniques and a bias modelling method for observational data were used to incorporate the evidence into the model. The stockpiling decision was modelled for adults (≥16 years old) and the United Kingdom was used as an example. The main outcome was the expected net benefits of stockpiling in monetary terms. Health benefits were estimated from deaths averted through stockpiling.

Results: After adjusting for biases in the estimated effectiveness of NAIs, the expected net benefit of stockpiling in the baseline analysis was £444 million, assuming a willingness to pay of £20,000/QALY ($31,000/QALY). The decision would therefore be to stockpile NAIs. There was a greater probability that the stockpile would not be utilised than utilised. However, the rare but catastrophic losses from a severe pandemic justified the decision to stockpile.

Conclusions: Taking into account the available epidemiological data and evidence of effectiveness of NAIs in reducing mortality, including potential biases, a decision maker should stockpile anti-influenza medication in keeping with the postulated decision rule.
Keywords
Pandemic influenza, evidence synthesis, bias modelling, neuraminidase inhibitors, stockpiling

This article is included in the Disease Outbreaks gateway.
Introduction

Like many other potentially catastrophic events for which governments need to prepare, influenza pandemics are rare. Although the risk is considered to be 3–4% per annum, the public health consequences are widely recognised to be potentially severe. The epidemiology of only a small number of influenza pandemics has been well studied and evidence for the effectiveness of remedinial influenza treatments in a pandemic scenario is scant. Yet, governments around the world still have to decide whether or not to stockpile anti-influenza medication like neuraminidase inhibitor (NAI) antivirals, such as oseltamivir (Tamiflu®) and zanamivir (Relenza®), as a defence against pandemic influenza.

The stockpiling of NAIs has been a controversial issue. Firstly, stockpiling may be seen to be a waste of large amounts of public money if the pandemic fails to materialise or if it is mild. In the United Kingdom, the previous Chief Medical Officer was criticised for spending £560 million on medicine that went largely unused in the 2009–10 pandemic. However, taking a default position of not stockpiling, or making the decision on the basis of intuition alone, is not justifiable given the rare but potentially catastrophic losses associated with pandemic influenza and the large cost of stockpiling.

Secondly, there has been a lack of conclusive evidence on the effectiveness of NAIs. Recent meta-analyses of randomised controlled trials (RCT) of seasonal influenza cases demonstrated reductions in rates of hospitalization, lower respiratory complications, and a decreased time to symptom alleviation but were unable to confirm or refute an effect of NAIs on more important clinical endpoints such as mortality. A caveat of these studies, which were required for licensure of drug in healthy adults, is that they were not powered to determine low frequency but critical endpoints such as mortality in a largely healthy adult population. A further meta-analysis of observational data from pandemic influenza did find evidence of a reduction in the risk of mortality when NAIs were given to patients hospitalised with influenza. Some authors have criticised it for being subject to a large degree of bias and rejected it as a suitable form of evidence with which to formulate policy decisions, though others argue that this evidence strongly supports the use of NAI treatment for influenza in hospitalised patients.

Evidence that has a bearing on death rates is not confined to measurement of mortality alone – there are other sources of relevant evidence. Clinical trials show that NAIs have beneficial effects on a number of outcomes as described above. The treatment has a plausible rationale and it works in vitro and in animal models for this zoonosis. An arguably extreme position is to assume that these observations contain no information regarding effectiveness in preventing the rarer, but more severe outcomes, such as death. People who take to heart Bradford Hill’s list of factors that should affect the interpretation of data (Box 1), would reject such a completely non-theoretical stance. But even within this framework conflicting conclusions may still be drawn, especially when inappropriately filtered through the lens of statistical significance. Estimation of potentially small effect sizes on rare endpoints is often characterised by uncertain and often conflicting evidence and any recent studies do conflict with those that support the effectiveness of NAIs.

Both an observed reduction and an increase in the risk of mortality are therefore potentially consistent with the aforementioned evidence. There is thus a compelling case for the synthesis of and extrapolation from various forms of evidence in order to examine the investment decision facing decision makers.

Previous studies have estimated how cost-effective NAI stockpiling would be under a range of different pandemic influenza scenarios. Stockpiling is generally estimated to be cost-effective. However, these studies took observational evidence of effectiveness, often from seasonal influenza studies, at face value and did not model potential biases that may have led to overestimation of benefits. Moreover, they only examined a limited number of specific future scenarios. The results of such cost-effectiveness models hinge on the available evidence of effectiveness and it may not be immediately clear to decision makers the implications of new evidence. We have therefore taken a different approach.

The calculation of the number of deaths from an influenza pandemic is simply calculated from a number of relevant variables such as the size of the population, the clinical attack rate, and the case fatality ratio. The effectiveness of NAIs in terms of relative risks can then be used to estimate the potential number of deaths averted through their use. A simple model can provide a useful framework to synthesise the available evidence while also remaining clear and transparent to decision makers. There is a large degree of uncertainty regarding the variables in the model, due to factors such as random mutations in the influenza virus, individual behaviour, and distribution of NAIs, nevertheless appropriate distributions can be specified for each variable and the uncertainty propagated through the model to estimate the distribution of possible numbers of deaths and resulting QALYs under the stockpiling and no stockpiling options. The model presented here exemplifies an approach to decision making under the types of uncertainty described above using a simple, transparent model to assist decision makers and to help inform the stockpiling decision.

Methods

Modelling approach

The methods used in this study are founded in normative decision theory, which considers what decisions we ought to take, and Bayesian statistics. We used a well-established technique based on expected utility theory to model the binary decision to stockpile or not to stockpile NAIs. Within this framework, the decision simplifies to a question of whether the expected net benefits of the stockpiling decision are positive.
The net benefit associated with stockpiling was set as the value of the deaths averted minus the costs of stockpiling. If the expected net benefit of stockpiling is positive then the decision would be to stockpile, and if it is negative, not to stockpile.

The value of the deaths averted was modelled as:

\[ \text{Pop} \times \text{Prob} \times \text{CAR} \times \text{CFR} \times \text{Hospital} \times \text{Treated} \times (1 - \theta) \times \text{QALY} \times \lambda \]

Firstly, the number of pandemic influenza deaths was calculated by multiplying the number of adults in the UK (Pop) by: the probability of there being a pandemic within the stockpile shelf-life (Prop), the clinical attack rate (CAR), and the case fatality ratio (CFR). We further multiply by the probability a pandemic influenza death occurred in hospital (Hospital), and the probability one of these patients receives NAIs (Treated). The number of deaths averted by NAI treatment in this population of NAI-treated adults was given by the relative risk reduction in mortality associated with NAI treatment \((1 - \theta)\). Finally, the value of these deaths averted was calculated by multiplying by the quality adjusted life years (QALY) associated with each pandemic influenza mortality (QALY), and the societal willingness to pay per QALY (\(\lambda\)). This model is further explicated in Figure 1.

We considered reductions in mortality among symptomatic adults resulting from stockpiling, but did not take into account possible additional effects on complications such as pneumonia or that community use might reduce complications, hospitalisation, or mortality. Only adults were considered on the grounds that NAI effectiveness is less certain in children and to determine if the decision to stockpile could be justified on the basis of any benefit among adults alone.

Decision modelling is founded in the Bayesian paradigm, which was used to evaluate the stockpiling decision for a future pandemic with unknown epidemiological variables and unknown effectiveness of NAI. A sub-model was specified for each epidemiological variable in the decision model. Data from previous pandemics were assumed to be observations from an underlying common distribution, the parameters of which were estimated using these data as described in the following section. The decision was then evaluated over posterior predictive distributions for the epidemiological parameters. We used a bias corrected effectiveness estimate for the effectiveness of NAIs as described below. The model was estimated using Markov Chain Monte Carlo (MCMC) with 10,000 iterations using R 3.2.3 and Stan 2.11.0. This method
obviates the need to conduct separate probabilistic sensitivity analyses since the posterior distribution of the net benefits represents the uncertainty about future influenza pandemics and NAI effectiveness. The expected net benefits represent the gains or losses from stockpiling, on average, given the different distributions for the different parameters. Convergence of the MCMC chains was assessed by visual inspection of autocorrelation, running mean, and trace plots in R.

Data and variables
The data and statistical code are provided with the paper.

Influenza pandemic epidemiology
The data used to estimate the parameters in the model were obtained from documents compiled to assess pandemic influenza and thus represent the decision maker’s prior knowledge. The shelf-life of oseltamivir, the principle drug comprising the vast majority of the NAI stockpile, is ten years.

The clinical attack rate and case fatality ratios from previous pandemics were assumed to be observations from beta distributions. Improper non-informative priors with a lower limit of zero were assigned to the parameters of these distributions, which were then updated with the data from the previous pandemics. We excluded the observation of a clinical attack rate of 60% in the 1889–92 Asiatic flu pandemic as the UK government’s worst case scenario is a clinical attack rate of 50%. The probability that a pandemic occurs in the shelf life of the stockpile was similarly estimated from the data with each decade between 1900 and 2010 as a binary observation equal to one if a pandemic occurred in that decade and zero otherwise. These binary observations were assumed to be observations from a Bernoulli distribution.

Effectiveness of neuraminidase inhibitors
No RCT evidence for the effectiveness of NAIs in reducing the risk of mortality in pandemic influenza was available. Too few deaths were observed in RCTs of seasonal influenza. We based our effectiveness estimate on a recently published pooled meta-analysis of observational, patient-level data from hospitalised pandemic influenza virus patients. We converted the odds ratios (OR) for mortality associated with NAIs (irrespective of time from onset) provided in the paper into relative risks (RR): $\text{RR} = \frac{\text{OR}}{1 - p + (p \times \text{OR})}$ where $p$ is the baseline (approximately 10%). The study was based on hospitalised patients, in order to apply the observed relative risk from hospitalised patients to the general population considered here, we made two conservative assumptions. First, we assumed that there would be no difference in the patients that would be hospitalised and those that would remain in the community in a no stockpile and stockpile scenarios. This is conservative because community treatment will be given earlier, on average, in the course of the disease if it can be administered in the community and there is evidence that the earlier the treatment is given, the better. Secondly, we assume that only deaths occurring in hospital in the non-stockpile scenario would be averted under the counterfactual stockpile scenario. A study of mortality in the A/H1N1 2009 pandemic in England, found that 92% of deaths (125 of 136 cases studied) occurred in hospital. Assuming that none of these 8% of deaths taking place in a non-stockpile scenario would be averted under the counterfactual is as conservative as it can be. The logic of our approach is laid out in Figure 1.

Bias modelling
In addition to these conservative assumptions regarding the application of in-hospital relative risk reductions to a community

---

**Figure 1.** Explanation of the model used to derive the value of the deaths averted due to neuraminidase inhibitor stockpiling.
Results

Summary of estimated parameters

Table 1 shows the posterior mean and 95% credible intervals for the parameters in the model. Using data from previous influenza pandemics, mean values (95% credible intervals) were as follows: clinical attack rate 23.8% (5.2%, 50.6%), case fatality ratio 0.7% (0.0%, 3.0%), and probability of experiencing a pandemic within a decade 38.5% (15.3%, 64.9%). The expected value for the mean QALY losses associated with influenza mortality was 15.2 (5.7, 20.9). The proportion of pandemic influenza deaths that occurred in hospitalised patients was 91.9% (86.9%, 95.8%).

The observed relative risk was 0.83 (95% confidence interval: 0.71, 0.94) and the bias corrected relative risk was estimated as 0.89 (0.71, 1.07). The principle sources of bias identified by the reviewers were selection bias, due to a lack of randomisation, the possibility that studies with a positive finding may have been more likely to volunteer their data for the meta-analysis, and attrition bias. Not all reviewers were in agreement about the overall effects of bias, but the median response was that there was an overestimation of treatment benefit.

Main results

Table 2 shows the results from various scenarios considered. The expected net benefit of stockpiling in the baseline analysis was

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean Value (95% Credible interval)</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costsa</td>
<td>£560,000,000</td>
<td>3</td>
</tr>
<tr>
<td>Willingness to pay per QALY</td>
<td>£20,000/QALY</td>
<td>29</td>
</tr>
<tr>
<td>Probability of pandemic in shelf life</td>
<td>38.7% (15.3%, 64.9%)</td>
<td>1,23</td>
</tr>
<tr>
<td>Adult Population</td>
<td>50.5 million</td>
<td>30</td>
</tr>
<tr>
<td>CARb</td>
<td>23.3% (5.2%, 50.6%)</td>
<td>1</td>
</tr>
<tr>
<td>CFRc</td>
<td>0.72% (0.01%, 2.97%)</td>
<td>1</td>
</tr>
<tr>
<td>QALY loss, mortality</td>
<td>15.2 (5.7, 21.0)</td>
<td>25,27,28,42,43</td>
</tr>
<tr>
<td>Proportion of pandemic influenza deaths in hospitalised patients</td>
<td>91.9% (86.9%, 95.8%)</td>
<td>25</td>
</tr>
<tr>
<td>Oseltamivir Effectiveness, mortality (relative risk)</td>
<td>0.83c (0.71, 0.94)</td>
<td>6</td>
</tr>
<tr>
<td>Bias corrected Oseltamivir Effectiveness (relative risk)</td>
<td>0.89 (0.71, 1.07)</td>
<td>6,26, Five independent assessors</td>
</tr>
</tbody>
</table>

CAR = clinical attack rate, CFR = case fatality ratio. Probabilities expressed as %.

aAssumed to be fixed.

bSee Appendix A for derivation.

cRelative risks converted from Odds ratios (0.81, 95% CI: 0.70, 0.93) using a baseline risk of mortality of 10%.
positive, which was caused by the very large number of deaths, many of which may be prevented by stockpiling, in the unlikely event of a severe pandemic. This can be seen in the long tail on the left of the distribution in Figure 2.

Figure 3 shows the decision under a range values for the effectiveness of NAIs, the percentage of hospitalised, symptomatic adults who would receive NAIs and willingness to pay per QALY threshold. If 100% of hospitalised, symptomatic adults with influenza received NAIs then the decision would be to stockpile as long as our threshold willingness to pay per QALY was greater than £11,116/QALY under our ‘bias corrected’ effectiveness estimate. When only 50% of hospitalised, symptomatic adults receive NAIs this threshold increases to £22,232/QALY, which would still be considered cost-effective in the range considered by NICE. The minimum percentage of hospitalised, symptomatic adults with influenza that would need to receive NAIs for the decision to be to stockpile at a threshold willingness to pay of £20,000 per QALY is 56%. Conversely, when the proportions of hospitalised, symptomatic adults who receive NAIs is 50%, 75%, or 100%, the minimum value for the relative risk of mortality associated with NAIs required for the intervention to be considered cost-effective at a £20,000 per QALY threshold is 0.88, 0.92, and 0.94, respectively.

Dataset 1. Raw data of ‘stockpiling neuraminidase inhibitors for pandemic influenza usage’
http://dx.doi.org/10.5256/f1000research.9414.d13265
This zip folder contains raw data behind the findings presented. The readme file provides a description for each data file.
Figure 3. Stockpiling decisions that would be made under a range of different values for the effectiveness of NAIs, the percentage of hospitalised, symptomatic adults that would receive NAIs, and thresholds for willingness to pay per QALY. The lines represent thresholds for decision makers. For any point inside the region bounded by a given line the decision maker should stockpile and conversely any point outside that region the decision maker should not stockpile.

Discussion
This study has found that the available evidence suggests that stockpiling NAIs for pandemic influenza is rational under a range of assumptions. Many of these assumptions are conservative, such as no reduction in adverse clinical outcomes other than mortality, no benefit in patients who would not have been hospitalised had there been no stockpile, and no effect in children. However, this decision required at least 56% of the influenza patients who would have died without a stockpile to receive NAIs if the threshold willingness to pay was £20,000/QALY. In the 2009 pandemic, 64% of hospitalised patients received NAIs, and in the United Kingdom specifically this proportion was 75%, suggesting that 56% is achievable, and that therefore, stockpiling is supported by the available evidence.

This paper is predicated on the purchase of a stockpile large enough to treat a large proportion of the population (80% in the UK) in the community and in hospital with NAIs. This may well be the correct strategy if new evidence emerges that community-based treatment reduces either complications, hospitalisations or mortality. Further research will be required; indeed, the Bayesian decision analysis used here can be extended to consider how much to stockpile rather than simply whether to stockpile. However, if the evidence base were to remain limited to mortality reductions in hospitalised patients, or if the societal willingness to pay per QALY was low, as it may be in many resource poor settings, a ‘hospital-treatment only’ policy might be considered. This would reduce the cost of the stockpile significantly. For example, in the 2009 pandemic only 0.5% of symptomatic cases were hospitalised, these patients would require far fewer doses than the 1.16 million courses (at a minimum) of NAIs dispensed in the 2009 pandemic. For a population of 50.5 million adults with a CAR of 25%, a hospitalisation probability of 0.5% would lead to only approximately 60,000 admissions. The evidence also suggests that more timely treatment of NAIs (within two days of symptom onset) is more effective than treatment at any point, which would suggest that the effectiveness of NAIs could be more favourable than modelled under the stockpiling policy. In all cases the decision would remain to stockpile NAIs.

Our conclusions are in line with the decision that would be made on the basis of cost-effectiveness evidence from previous studies. However, our study does not take observational evidence at face value, but ‘downgrades’ it, thereby yielding a reduced estimate of effectiveness and wider credible limits. We have calculated the distribution of possible deaths from pandemic influenza using a relatively simple mathematical model and then ‘averaged’ over the distributions of the variables rather than examining cost-effectiveness on a scenario-by-scenario basis. This approach is
intuitively simple and is aimed to provide correct inferences using a simple logical framework for the synthesis of the commonly available evidence in order to assist decision makers with a complex decision. The model allows the logical basis of the decision to be ‘reverse engineered’, allowing the decision to be critiqued within the framework established by the model. We note though that many people are highly sceptical about the benefits of NAIs. This study is not intended to be prescriptive. As Figure 3 illustrates, a decision maker with a highly sceptical belief about NAI effectiveness should not stockpile.

Obtaining an estimate for the bias in any particular study, or consolidated group of studies, is clearly an uncertain undertaking. There is an evidence base on bias arising from meta-regressions or other analyses comparing the results of imperfect studies to those of a ‘gold standard’. A recent Cochrane review comparing treatment effects reported in observational studies as compared to RCTs found that, “on average, there is little evidence for significant effect estimate differences between observational studies and RCTs...”34 It is not surprising, given the considerable uncertainties surrounding the meta-analysis cited here, that the differences between the reported effects and our bias corrected effect resembles the differences in empirical studies comparing observational studies and RCTs14,36.

We acknowledge weaknesses in our study. The only outcome considered in the analyses was mortality. Adverse events caused by NAIs may also generate increased costs and hence reduced benefit. For example, a review of clinical trial evidence of NAIs found an increased risk of nausea and vomiting associated with treatment1. The authors also reported a possible increase in the risk of psychiatric adverse events. However, this only reached statistical significance in exploratory analyses including a supra-licence dose and off-treatment periods. A more recent meta-analysis based on individual-level patient data of clinical trials focusing on licensed dose only found no such effects, but the number of events was small1. Neuraminidase inhibitors may also have protective effects against some adverse events such as cardiac events, and may reduce the risk of influenza-associated pneumonia and hospitalisation11–13. The benefit of treatment is unlikely to be grossly over-estimated and is likely to be under-estimated given our conservative assumptions. We have also not considered potential effects on children or from reductions in complications, hospitalisations or mortality that might be associated with community-based treatment, or any benefit arising from changing disease dynamics and reduced transmission; nor have we considered wider societal effects, such as productivity gains, reduced community transmission, and the value placed on a stockpile for a potentially risk averse population, all of which may increase the benefits of stockpiling.

We note that our analysis is focussed on the United Kingdom but that it may be of use to other countries. The model for the benefits of NAIs can be simply applied to new contexts. However, the determination of the costs of the stockpile remains difficult. The costs depend on the treatment strategy planned for a given country and any price negotiations between the manufacturer and the government. A useful tool in this context is the ‘headroom’ method that asks instead what the maximum amount a decision maker should be willing to pay for an intervention, given a willingness to pay per unit benefit. This is a useful direction for future research.

We have assumed independence between the clinical attack rate and case fatality ratio, as well as other variables, however there is some evidence to suggest that they could be correlated37. Nevertheless, the data are admittedly scant, and it is expected that this is a neutral assumption. Of course, if they are positively correlated then our conclusions become more conservative.

Our model examines the decision in the abstract and does not concern itself with externalities such as the possibility that availability of the drug will affect attitudes and hinder the effort to contain the spread of the disease, or that resistance to antivirals may develop. Nor have we considered the sensitivity of clinical diagnosis of influenza in identifying true positives or the costs and logistics of establishing a distribution process for the NAIs. The propensity to consult is also an important factor that may have affect the proportion of true positives, which in turn may have a bearing on the use of a stockpile if used on a “first come, first served” basis. Further research is required to optimize distribution and behaviour during a pandemic to ensure the cost-effectiveness of the stockpiling.

Conclusions
Taking into account the existing evidence on pandemic influenza and the effectiveness of NAIs the decision should be to stockpile, provided a utilitarian decision-making framework is used of minimising expected losses and hence maximising expected benefits.

Data availability
The data used to estimate the parameters in the model were obtained from documents compiled to assess pandemic influenza and thus represent the decision maker’s prior knowledge1.


Author contributions
SIW, RJL, and YFC conceived the study and developed the methodology; JSN-V-T, PRM, PJC, MZ, and SV contributed to the parameterisation of the model and provided background to pandemic influenza; SIW, RJL, YFC, OU, and PJC independently reviewed the observational evidence and adjusted effectiveness estimates for bias; SIW and RJL prepared the first draft of the paper; this and subsequent drafts were reviewed and revised by all authors.
Competing interests
JSN-V-T, PRM, and SV have co-authored the Muthuri et al. (2014) study which was supported by an unrestricted educational grant from F. Hoffman La Roche. PRM and JSN-V-T were MUGAS Review Board members that reviewed the oseltamivir data (both from randomised controlled trials and observational studies including data from the 2009/10 pandemic) and agreed on evidence gaps and a statistical analysis plan that would address these gaps. JSN-V-T is Chair of NERVTAG (New and Emerging Respiratory Virus Threat Advisory Group).

Grant information
SIW, RJL, YFC, and PJC are part-funded/supported by the National Institute for Health Research (NIHR) Collaborations for Leadership in Applied Health Research and Care West Midlands. This paper presents independent research and the views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements
This paper presents independent research funded by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care West Midlands. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. We would also like to thank two referees and commenters online for their feedback and discussion on the article.

Appendices
Click here to access the data.

References

24. Grant RL: Converting an odds ratio to a range of plausible relative risks for better communication of research findings. BMJ. 2014; 348: h4500. PubMed Abstract | Publisher Full Text


Open Peer Review

Current Referee Status: ✔️ ✔️

Version 2

Referee Report 17 March 2017

https://doi.org/10.5256/f1000research.12012.r21030

Joel Kelso

School of Computer Science and Software Engineering, The University of Western Australia, Crawley, WA, Australia

This revised version (2) of the paper addresses all the criticisms I raised in my report.

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Referee Report 27 January 2017

https://doi.org/10.5256/f1000research.10138.r19365

Joel Kelso

School of Computer Science and Software Engineering, The University of Western Australia, Crawley, WA, Australia

This article reports on a probabilistic cost effectiveness of stockpiling neuraminidase inhibitor antiviral drugs to mitigate against pandemic influenza deaths.

The study is methodologically sound. The decision-theoretic approach which selects the optimal course of action based on the utility of each outcome and the probability of each outcome conditional on each decision is appropriate. In this case the actions are whether or not to stockpile NAIs for use in a pandemic, and the outcome is the cost of the antiviral stockpile and the expected number of deaths during the shelf-life period of the antivirals.

The model of the expected number of pandemic deaths identifies is structurally sound and uses various appropriate data sources to quantify uncertainties present in all the parameters.

The conclusion that stockpiling NAIs is cost-effective for a sufficiently high willingness-to-pay cost per QALY follows from the model and the data used.
However, I think that more attention needs to be drawn to dependency of this result to the crucial antiviral effectiveness parameter. The methodology where all parameters are treated probabilistically in a uniform way is excellent; however additional one- or two-way sensitivity analyses are still valuable for providing insight into the effect of the most important parameters. The authors have done this for the proportion of hospitalised cases receiving antivirals; it seems appropriate to also do this for NAI effectiveness, given the ongoing debate on the subject.

I have divided further comments into three sections: major essential revisions, commentary with discretionary additions, and minor technical clarifications needed.

Essential Revision
------------------
1. As stated above, the reader would be well served with an additional figure similar to Figure 3, but plotting QALY threshold against mean NAI effectiveness. In my shallow experimentation with the author's model, it looks like at 20,000 GBP / QALY, NAIIs cease being cost effective at around 0.94 effectiveness (relative risk, compared to 0.89).

Commentary
----------

The suggestions made below I think might improve the article however I think the authors can best judge whether the additional effort and added complexity would be worthwhile or would be too diverting.

I commend the authors for including the model code in the Appendix, I managed to run this code with relatively little effort.

1. The methodology of using expert opinion to mitigate potential bias in the studies estimating NAI effectiveness is a practical measure that is probably worthwhile. Some additional detail on the process would be appreciated. For example: how were assessors selected? How much time did the reviewers take in their bias estimates?

2. In the discussion it could be noted that in a future pandemic with a large CAR or CFR, the proportion of severe cases receiving hospital care and the level of care are likely to be lower, simple due to hospitals being overwhelmed. The estimates of proportion of deaths occurring in hospital are from the 2009 pandemic which was very mild.

3. If NAIs have any effect in preventing further transmission, e.g if they shorten the period of viral shedding, then mass administration of antivirals may reduce the overall attack rate and consequent mortality even if NAIs are not effective for mortality reduction of severe cases. As the study's model does not capture this, this is another way in which the study is conservative.

Minor Technical Revisions
-------------------------
1. The CAR and CFR parameters used in the model are for a pandemic without NAI usage. Given that NAIs were used in the 2009 pandemic, should the CAR and CFR estimates for 2009 be included along side those of previous pandemics? If the 2009 CAR and CFR estimates are for example based on global data where NAI usage might be negligible that would be OK; but if they are based primarily on UK or US data they should possibly be excluded.
2. In the Appendix page 4 there is a citation [20] that isn’t given in a reference list.

3. In the last sentence of the 3rd paragraph, the RR derived based on the OR and 10% mortality is stated as 0.89. This is the same as the bias-corrected RR given in the next paragraph. Is this intentional? Or should it be the RR value based on the OR and 10% mortality (but without bias correction), in which case it should be 0.825 (from the formula).

4. The R / BUGS code in the Appendix worked almost without alteration. I found that I had to:

   Install BUGS (OpenBGUS).
   Hoist the npv function to the top.
   Remove the codaPkg=TRUE setting to obtain a result object.
   (also the "obs" and "qaly" values appear to be dead code)

   If F1000 allows additional appendix files this could be supplied as an additional plain ASCII file, to avoid scraping the text from the PDF and correcting resulting formatting.

5. I can't find the support for the n_hosp data value of 136. The tot_hosp value of 125 appears in the Donaldson BMJ paper. That paper gives 138 for the total number of confirmed deaths due to pandemic influenza.

**Competing Interests:** No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 08 Mar 2017

**Sam Watson**, University of Warwick, UK

We thank the review for their comments and detail our responses below, point by point. The referee's text is in Italics.

**Essential Revision**

------------------

1. As stated above, the reader would be well served with an additional figure similar to Figure 3, but plotting QALY threshold against mean NAI effectiveness. In my shallow experimentation with the author's model, it looks like at 20,000 GBP / QALY, NAIs cease being cost effective at around 0.94 effectiveness (relative risk, compared to 0.89).

   We have replaced figure 3 to incorporate these considerations and additional commentary.

**Commentary**

----------

The suggestions made below I think might improve the article however I think the authors can best judge whether the additional effort and added complexity would be worthwhile or would be too diverting.
I commend the authors for including the model code in the Appendix, I managed to run this code with relatively little effort.

1. The methodology of using expert opinion to mitigate potential bias in the studies estimating NAI effectiveness is a practical measure that is probably worthwhile. Some additional detail on the process would be appreciated. For example: how were assessors selected? How much time did the reviewers take in their bias estimates?

We have added additional description in the Methods section although we also refer the referee to the cited article.

1. In the discussion it could be noted that in a future pandemic with a large CAR or CFR, the proportion of severe cases receiving hospital care and the level of care are likely to be lower, simple due to hospitals being overwhelmed. The estimates of proportion of deaths occurring in hospital are from the 2009 pandemic which was very mild.

2. If NAIs have any effect in preventing further transmission, e.g if they shorten the period of viral shedding, then mass administration of antivirals may reduce the overall attack rate and consequent mortality even if NAIs are not effective for mortality reduction of severe cases. As the study’s model does not capture this, this is another way in which the study is conservative.

We will add these points to the discussion.

Minor Technical Revisions

1. The CAR and CFR parameters used in the model are for a pandemic without NAI usage. Given that NAIs were used in the 2009 pandemic, should the CAR and CFR estimates for 2009 be included alongside those of previous pandemics? If the 2009 CAR and CFR estimates are for example based on global data where NAI usage might be negligible that would be OK; but if they are based primarily on UK or US data they should possibly be excluded.

We would argue that the 2009 observed CAR and CFR are relevant data points to infer the distribution of possible CAR and CFR values. It is possible that mass NAI distribution may alter the parameters of these distributions, however, without further information it is not possible to model this. Excluding the 2009 pandemic would bias our estimates, and given the small amount of data, this data point provides a relatively large amount of information. We therefore opt to use all available data.

1. In the Appendix page 4 there is a citation [20] that isn’t given in a reference list. This has been amended.

1. In the last sentence of the 3rd paragraph, the RR derived based on the OR and 10% mortality is stated as 0.89. This is the same as the bias-corrected RR given in the next paragraph. Is this intentional? Or should it be the RR value based on the OR and 10% mortality (but without bias correction), in which case it should be 0.825 (from the formula). We believe the referee may be in error, as it is 0.83 in the third paragraph of Appendix B. However, we have resubmitted the revised appendix to ensure the correct version is available.

1. The R / BUGS code in the Appendix worked almost without alteration. I found that I had to:
Install BUGS (OpenBGUS).
Hoist the npv function to the top.
Remove the codaPkg=TRUE setting to obtain a result object.
(also the "obs" and "qaly" values appear to be dead code)

If F1000 allows additional appendix files this could be supplied as an additional plain ASCII file, to avoid scraping the text from the PDF and correcting resulting formatting.

We have provided a file for use in the program Stan to run the program as well. The models were initially run in WinBUGS before ‘upgrading’ to Stan. We have noted this in the Appendix but opt to provide both pieces of code for people using both programs.

1. I can't find the support for the n_hosp data value of 136. The tot_hosp value of 125 appears in the Donaldson BMJ paper. That paper gives 138 for the total number of confirmed deaths due to pandemic influenza.

This typo has been amended.

**Competing Interests:** As stated in the article.

---

Referee Report 18 January 2017

https://doi.org/10.5256/f1000research.10138.r17973

Pasi M. Penttinen
European Centre for Disease Prevention and Control, Stockholm, Sweden

This is a well designed, carefully executed and documented study, that provides important insights into the cost-effectiveness of national stockpiles of neuraminidase inhibitors to be used during influenza pandemics.

The analysis is relying on a number of key assumptions, such as the effectiveness of NAI antivirals against mortality due to influenza, the probability of a pandemic occurring during the shelf life of the stockpile and the proportion of pandemic influenza deaths occurring in hospital. Many of these assumptions are based on a limited or controversial evidence base, however the authors acknowledge and address most of these limitations.

The assumption that most pandemic deaths occur in hospitals, is based on the observation during the 2009 pandemic in the UK, however in many countries, already during severe influenza A(H3N2) epidemics, and during many previous pandemics, the majority of deaths are likely to occur in the community, outside of hospitals. It is confusing that the authors compare the costs of a population wide (80%) stockpile with the estimated benefits on hospital mortality only. Although this is discussed in the second paragraph of discussion, it would be helpful to see an analysis or results taking also into account outpatient and community mortality.

It is likely that such an analysis would be useful for other countries than UK. Please discuss briefly the limitations of this approach and these assumptions, when replicating the study in other settings (such as differences in societal willingness to pay per QALY).
In Box 1, the two columns are not aligned when viewing as a pop-up on MS Internet Explorer.

In Figure 1, the references to UK, and the national pandemic flu service are not helpful and distract from the more general main message of this figure.

**Competing Interests:** No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

---

**Author Response 08 Mar 2017**

**Sam Watson, University of Warwick, UK**

We thank the review for their comments and detail our responses below, point by point. The referee’s text is in italics.

*This is a well-designed, carefully executed and documented study, that provides important insights into the cost-effectiveness of national stockpiles of neuraminidase inhibitors to be used during influenza pandemics.*

*The analysis is relying on a number of key assumptions, such as the effectiveness of NAI antivirals against mortality due to influenza, the probability of a pandemic occurring during the shelf life of the stockpile and the proportion of pandemic influenza deaths occurring in hospital. Many of these assumptions are based on a limited or controversial evidence base, however the authors acknowledge and address most of these limitations.*

*The assumption that most pandemic deaths occur in hospitals, is based on the observation during the 2009 pandemic in the UK, however in many countries, already during severe influenza A(H3N2) epidemics, and during many previous pandemics, the majority of deaths are likely to occur in the community, outside of hospitals. It is confusing that the authors compare the costs of a population wide (80%) stockpile with the estimated benefits on hospital mortality only. Although this is discussed in the second paragraph of discussion, it would be helpful to see an analysis or results taking also into account outpatient and community mortality.*

We did not consider non-hospital mortality as there were no data on the effectiveness of NAIs outside of the hospital setting, where there may be differences in compliance and other factors, when the study was conducted. We note that the way we have set up the analysis is to try to be as conservative as possible: the highest stated costs with a justifiable patient pool. On this basis we note that if a decision to stockpile is supported under our assumptions then it will certainly be supported if there is any benefit outside of the hospital. Recently published analyses outside of the hospital setting suggest a potential benefit (https://doi.org/10.1093/cid/cix127), however we opt to remain conservative in our analyses.

*It is likely that such an analysis would be useful for other countries than UK. Please discuss briefly the limitations of this approach and these assumptions, when replicating the study in other settings (such as differences in societal willingness to pay per QALY).*

We have amended the discussion to reflect this.
In Box 1, the two columns are not aligned when viewing as a pop-up on MS Internet Explorer. 

This is an issue for the journal.

In Figure 1, the references to UK, and the national pandemic flu service are not helpful and distract from the more general main message of this figure.

We have removed that box from the figure.

**Competing Interests:** As stated in the article.

---

**Comments on this article**

**Version 1**

**Author Response 27 Jan 2017**

**Sam Watson**, University of Warwick, UK

We certainly acknowledge that there are studies that both confirm and refute the effectiveness of NAIs for pandemic influenza, and we will update the references in the article to include the more recent studies Rokuro Hama cites. However, the point of the article is that all of the evidence needs to be weighed and synthesized in order to arrive at a decision. For an uncertain prospect with a potentially small effectiveness on a rare endpoint we would expect to find a wide range of findings due to both natural variation and differences in study designs. Our contention is that when all is taken into account, using a framework such as Bradford-Hill’s, NAIs have an uncertain yet likely beneficial effect. Nevertheless, we provide the range of decisions in the article that should be made under different levels of effectiveness. A decision maker with a highly sceptical prior, such as Hama and Vlassov’s, should not stockpile, as we note. This article is not intended to replace decision-makers, but to provide a logical and consistent framework within which a decision can be made.

In responding to Dr Hama’s further note, we re-emphasise that we have used bias corrected relative risks obtained from work which is currently regarded by expert bodies [Ref. 9 in full text] and policy makers as the best available epidemiological evidence on mortality pertaining to the 2009 pandemic; and we have used these to illustrate how evidence synthesis and decision modelling can be used to support complex decision-making.

Finally, Dr Vlassov is quite incorrect to assert that the data upon which the current paper is based are “produced by the drug company”. In the PRIDE study the authors were clear to state that none of the source datasets which contributed to the PRIDE mortality analysis were produced by or funded by pharmaceutical companies. The authors have always readily acknowledged that the PRIDE IPD analysis was made possible by an unrestricted grant from F. Hoffman La Roche. The particulars of that contract can be viewed here: [https://www.nottingham.ac.uk/research/groups/healthprotection/projects/pride.aspx](https://www.nottingham.ac.uk/research/groups/healthprotection/projects/pride.aspx) In particular it is
stated clearly that: “The study is taking place through an unrestricted educational grant funding from F. Hoffmann-La Roche but is being undertaken fully independently of the company, which has had/will have: no input to the project design; no access to any of the data; no role in analysis or data interpretation; no preview of the study results; and no opportunity to preview or comment on any manuscripts arising from the work”.

Sam Watson, on behalf of the study authors

*Competing Interests:* As stated in the article.

Author Response 10 Oct 2016

**Sam Watson**, University of Warwick, UK

We would like to thank Rokuro Hama and Mark Jones for their comments on our article. Much of the discussion raised by Hama and Jones relate to the validity of the PRIDE mortality analysis [1] and not to the actual methodology in the current paper. Methodological disagreements about the former paper raised by members or associates of the Cochrane Respiratory Group have already been extensively discussed and replied to.[2-5]

In the present study we did not in any case take the PRIDE data at face value, but conducted a bias modelling exercise to adjust the estimates of effectiveness in order to allow for biases in the analyses of observational data. Given other evidence, such as reductions in the risk of pneumonia, reduced length of stay in hospital, and reduction in time to symptom alleviation associated with neuraminidase inhibitors (NAIs), as was also demonstrated by Hama and Jones in their cited studies,[6,7] it was considered that it was unlikely that the observed effect was driven by bias alone. Indeed, similar such considerations should be made when considering the potentially biased results of the observational studies cited by Jones.[6,7]

A key point that underlies the argument of this article is that for an effective policy decision to be made the totality of evidence has to be taken into account. How the synthesis is achieved and a decision made is a continuing debate, as evidenced by the disagreement between the FDA and CDC over the use of NAIs, as cited by Jones.[8] Nevertheless, the null hypothesis testing framework and p-values are not suitable for such a decision making framework. The probability that physiologically active compounds such as neuraminidase inhibitors have an effect size of exactly zero is negligible. The conclusions then turn on a question of *how large* the effect size is, which must be inferred from uncertain data and analyses that may exhibit both internal and external biases. This study did not use hazard ratios as implied by Hama and Jones, rather a “bias corrected” relative risk. Nevertheless, we note that the confidence intervals for the hazard ratios for mortality cited by Jones are wide and relatively uncertain, but contain the possibility of significant reductions (and increases) in the hazard of mortality. We would conjecture that there is a not insubstantial overlap in potential effect sizes between their studies and ours.

Sam Watson, on behalf of the study authors.


[2] Jones M, Del Mar C, Hama R. Statistical and methodological concerns about the beneficial effect of


**Competing Interests:** As stated in the article.